Identification of novel DNA damage response mechanisms using SLiM proteomics



A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

Ann-Marie K. Shorrocks
Department of Oncology
MRC Weatherall Institute of Molecular Medicine
St Hilda's College

Hilary Term 2022

Supervisor: Dr. Andrew Blackford

50,000 words (excluding Appendices, References and Tables)

For from him and through him and for him are all things.

To him be the glory forever! Amen.

Romans 11 v 36

Abstract

The repair of damaged DNA is essential for maintaining genome integrity. Accordingly, all known forms of life have evolved a complex repertoire of biochemical pathways to detect and repair such damage, collectively called the DNA damage response. Like all biological pathways, the DNA damage response is facilitated by a series of protein-protein interactions, but the mechanisms by which DNA repair protein complexes are assembled, disassembled and regulated by post-translational modifications are still poorly understood.

To address this issue, a combined bioinformatics and proteomics approach was developed to identify and characterize novel protein-protein interactions in DNA damage response proteins relying on Short Linear peptide Motifs (SLiMs). Using this approach, novel interactions have been identified between the DNA repair proteins: MDC1 and TOPBP1, TOPBP1 and CIP2A, AUNIP and CtIP, and between the BLM helicase complex and the single-stranded DNA binding protein RPA. Investigation of the functional relevance of the interaction between the BLM helicase complex and RPA demonstrated that RPA-binding is required for stable BLM recruitment to sites of DNA replication stress and replication fork restart.

To overcome the limitations of manually identifying SLiMs, a web-based tool was developed to predict SLiMs in DDR proteins automatically. This work forms the foundation for the development of a DNA repair interactome based on SLiMs, for a greater understanding of the diverse range of pathologies associated with mutations in DNA repair genes and the development of new therapies.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Andrew Blackford for his supervision. I have been impressed by your dedication, as well as your commitment to scientific rigor and quality.

Secondly, I would like to thank members of the Blackford Lab, past and present, for all their support and contributions made to this project. I would especially like to thank the post-docs Dr. Samuel Jones and Dr. Kaima Tsukada for their supervision and the contributions they have made to this project. Sam, especially, thank you for the hours you've given to supervising me. I do think you've succeeded in your project of toughening me up, something I am grateful for. Matthew, Robert and Julius, it's been a pleasure to do a DPhil alongside you.

At the WIMM, I would like to thank Jill Brown from the Patel lab for all your help in getting DNA fiber experiments to work. I would also like to thank Chris Lagerholm and Jana Koth from the Wolfson Imaging Centre for help with microscopy and Simon McGowan from the CBRG for your contribution to this project. At the TDI, I would like to thank Roman Fischer and Iolanda Vendrell for all your support with the proteomics aspects of this thesis.

I would like to thank the University of Oxford Medical Research Council Doctoral Training Partnership for funding this DPhil, as well as Cancer Research UK and Against Breast Cancer for funding the Blackford lab. I would also like to thank my Oxford college, St Hilda's College.

To my college advisor, Prof. Ester Hammond, thank you for all your support over this past year. I would also like to thank my undergraduate Director of Studies at the University of

Cambridge, Dr. Andrew Murray. I will forever be thankful for you taking a chance on me all those years ago, changing the trajectory of my life, and for all the ways you have supported my academic progress over the past decade.

Next, I would like to thank all my friends from StAG and St Ebbe's, I do not know whether I would have made it through my DPhil without you. From StAG, I would like to thank Maria & James Thompson, for your friendship, all the times you've helped me move house and for making me the proudest godmother to Jonny. Catriona Irvine, thank you for being such a faithful friend, always ready to point me to Jesus. From St Ebbe's, I would like to thank Barnabas Williams, Anastasia & Simon Posner, Ashley Wright, Georgie Bishop, Rachel Cross, Bethany Scott, Helen Hambling and Kara Jackman. To all the St Ebbes MA's 2021-2022, I am extremely thankful to the Lord that I've had the opportunity to serve alongside you these past couple of months. To Nuala Marshall, I do not know whether I would have made it through the past year without you.

I would like to thank my family. To my mother, the older I get, the more I can see and appreciate your gentle strength and all the sacrifices you made for us to have a better future. To my brother Jake, I adore you and could not be prouder of you. To my stepfather Colin, thank you for modelling a true husband and father to us. It always blows me away how you treat us as if we were your biological children. To Nanna and Grandpa, thank you for providing a haven for us so many times, and for all your support and encouragement over the years.

Lastly and most significantly, to my LORD and God. You are my soul's delight and crown, my rock and refuge, the one who makes the saddest heart sing for joy. As the Oxford motto declares, you are my light and my salvation. May all my days bring glory to your name.

Table of Contents

CHAPTER 1: INTRODUCTION 1

1.1 THE DNA DAMAGE RESPONSE AND ITS RELEVANCE TO HUMAN DISEASE 1

1.1.1 The DDR: an Integrated Signalling and Genome Maintenance Network 1

1.1.2 Sources of DNA Damage 2

1.1.3 Signalling in Response to DNA Damage 4

1.1.4 Pathways of DNA Repair 6

1.1.5 DNA Damage Checkpoint Signalling in Interphase and Mitosis 10

1.1.6 The DNA Damage Response and Human Disease 12

1.1.7 Harnessing the DDR to Treat Cancer 21

1.2 THE CELLULAR RESPONSE TO DNA DSBS 25

1.2.1 ATM Initiates and Sustains Chromatin-Based DDR Signalling 25

1.2.2 Pathways of DSB Break Repair 28

1.2.3 DSB Repair by Pathways Based on NHEJ 28

1.2.4 DSB Repair by Pathways Based on HDR 31

1.2.5 The Cellular Functions of the BTR Dissolvasome Complex 42

1.2.6 Establishment of DSB Repair Pathway Choice 43

1.2.7 DNA End Protection by 53BP1 45

1.2.8 Responding to DSBs in Mitosis 47

1.3 DNA REPLICATION AND THE CELLULAR RESPONSE TO REPLICATION STRESS .. 49

1.3.1 Eukaryotic DNA Replication 49

1.3.2 Controlling DNA Replication: Once and Only Once 51

1.3.3 Barriers to DNA Replication 53

1.3.4 Checkpoint Signalling in Response to DNA Replication Stress 55

1.3.5 Mechanisms of Stalled Replication Fork Restart 57

1.3.6 Resolving Collapsed Replication Forks 60

1.3.7 Human Diseases Associated with Defects in the Replication Stress Response63

1.4 SHORT LINEAR PEPTIDE MOTIFS (SLIMS) 65

1.4.1 Characterisation of SLiMs 66

1.4.2 Implications for Drug Design and Treatment Strategies 68

1.5 AIMS AND OBJECTIVES 69

CHAPTER 2: MATERIALS & METHODS 72

2.1 SEQUENCE ALIGNMENTS 72

2.2 PEPTIDE PULLDOWNS 72

2.3 SAMPLE PREPARATION FOR MASS SPECTROMETRY 75

2.4 SDS-PAGE AND WESTERN BLOTTING 76

2.5 PLASMIDS & CLONING 77

2.6 CELL LINES & CELL CULTURE CONDITIONS 78

2.6 IMMUNOPRECIPITATIONS 79

2.7 IMMUNOFLUORESCENCE AND CONFOCAL MICROSCOPY 79

2.8 GENERATION OF RMI1 KNOCKOUT RPE-1 CELLS 81

2.9 LENTIVIRAL TRANSDUCTIONS 81

2.10 RNA INTERFERENCE 82

2.11 DNA FIBRE ANALYSIS 82

CHAPTER 3: A PILOT SLIM PROTEOMICS SCREEN TO IDENTIFY PROTEIN-PROTEIN INTERACTIONS IN DNA REPAIR PROTEINS 84

3.1 AUTHORSHIP STATEMENT 84

3.2 INTRODUCTION 84

3.2.1 Experimental Identification of Protein-Protein Interactions 84

3.2.2 Using Evolutionary Conservation as a Tool to Identify Functional SLiMs in the DDR 86

3.2.3 Aims and Objectives 89

3.3 RESULTS 89

3.3.1 Pipeline to Identify Functional SLiMs in DDR Proteins 89

3.3.2 Synthesis of SLiMs as Biotinylated Peptides and Peptide Pulldowns for Mass Spectrometry 94

3.3.3 Identification of High-Confidence Interacting Proteins 98

3.3.4 Validation of Candidate SLiM-Based Interactions 115

3.3.5 MDC1 Interacts with TOPBP1 to Promote Chromosomal Stability during Mitosis 116

3.3.6 TOPBP1 Interacts with the C-terminus of CIP2A 119

3.3.7 Two SLiMs in the C-terminus of AUNIP interact with CtIP 126

3.3.8 Identification of Three Conserved RPA-binding SLiMs in the BTR Complex 130

3.4 DISCUSSION 133

CHAPTER 4: THE BLOOM SYNDROME COMPLEX SENSES RPA COATED SINGLE-STRANDED DNA TO RESTART STALLED REPLICATION FORKS 141

4.1 AUTHORSHIP STATEMENT 141

4.2 INTRODUCTION 141

4.2.1 Bloom Syndrome is Characterised by Developmental Defects and Cancer Predisposition 141

4.2.2 Aims and Objectives 143

4.3 RESULTS 144

4.3.1 Identification of Three Conserved RPA-binding SLiMs in the BTR Complex…… 144

…… 148

4.3.2 The RPA-binding SLiMs in BLM and RMI1 interact with the RPA1 N-terminal OB-fold domain …… 148

4.3.3 Generation of a Cell-Based System to Test the Physiological Relevance of BTR-RPA binding…… 151

4.3.4 RPA Binding Stabilises BLM at ssDNA Sites but is not Essential for BLM Recruitment to DSBs …… 158

4.3.5 RPA Binding is Required for the BTR Complex to Restart Stalled Replication Forks but not to Suppress SCEs …… 159

4.3.6 RPA Binding is Required for the Recruitment of BLM to Stalled Replication Forks…… 164

4.3.7 BTR-mediated Replication Fork Restart is Dependent on BLM Helicase Activity and the interaction between BLM and TOP3A/RMI1/RMI2…… 164

4.4 DISCUSSION…… 168

CHAPTER 5: TOWARDS A DNA REPAIR INTERACTOME BASED ON SLIMS…….. …… 176

5.1 AUTHORSHIP STATEMENT…… 176

5.2 INTRODUCTION…… 176

5.2.1 Computational Approaches for Predicting SLiMs…… 176

5.2.2 Prediction Performance on SLiMs: Case Study on SLiMs in the BTR Complex…… 180

5.2.3 Aims and Objectives…… 186

5.3 RESULTS…… 186

5.3.1 Generation of Automated Sequence Alignments for Every Protein Encoding Gene in the Human Genome 186

5.3.2 Automated Identification of SLiMs..... 190

5.3.3 Batch search for SLiMs in DDR proteins..... 194

5.3.4 Use of AlphaFold to validate SLiMs Predicted by the SLiMCon Tool 212

5.3 DISCUSSION 215

CHAPTER 6: GENERAL DISCUSSION..... 218

6.1 SUMMARY OF WORK PRESENTED IN THIS DPHIL 218

6.1.1. Rationale of Project 218

6.1.2. Development of a Combined Bioinformatics and Proteomics Approach to Identify SLiMs in DDR Proteins..... 218

6.1.3. The Bloom Syndrome Complex Senses RPA Coated Single-Stranded DNA to Restart Stalled Replication Forks 219

6.1.4. Development of a Web-Based Tool to Identify SLiMs..... 220

6.2 FUTURE WORK 220

6.3 FINAL REMARKS..... 221

CHAPTER 7: REFERENCES..... 223

CHAPTER 8: APPENDICES 281

8.1 MANUAL SEQUENCE ALIGNMENTS USED TO PREDICT SLIMS..... 281

8.2 SILVER STAIN GELS OF PEPTIDE PULLDOWNS FOR MASS SPECTROMETRY 282

8.3 PROTEINS ASSIGNED TO THE BACKGROUND PROTEOME 284

8.4. FULL STRING NETWORKS..... 288

8.5. REPRESENTATIVE IMAGES OF DNA FIBRES..... 302

8.6. AUTOMATED SEQUENCE ALIGNMENT OF BLM ORTHOLOGS FROM THE SLIMCON TOOL 303

8.7. LIST OF 120 DDR PROTEINS USED AS AN INPUT IN THE SLIMCON TOOL 318

List of Abbreviations

AAD	ATR activation domain
alt-EJ	Alternative end joining
APH	Aphidicolin
BPA	p-benzoyl-L-phenylalanine
BRCT	BRCA1 carboxyl terminal domains
BTR	BLM-TOP3A-RMI1-RMI2
CDKs	Cyclin-dependent kinases
CK2	Casein kinase 2
CO	Crossovers
DDR	DNA damage response
dHJ	Double Holliday junction
dNTPs	Deoxyribonucleotides
DSBs	DNA double strand breaks
ELM	Eukaryotic Linear Motif
FHA	Forkhead associated domain
HDR	Homology directed repair
HR	Homologous recombination
HU	Hydroxyurea
IP	Immunoprecipitations
IR	Ionising radiation
MEME	Multiple EM for Motif Elicitation
MMEJ	Microhomology-mediated end joining
MMR	Mismatch repair
NCO	Non crossover
NHEJ	Non-homologous end joining
OB	Oligonucleotide-binding
PARP	Poly(ADP-ribose) polymerase
PIKK	Phosphatidylinositol 3-kinase like protein kinase
rNTPs	Ribonucleotides
ROS	Reactive oxygen species
RPA	Replication protein A complex
SCE	Sister chromatid exchange
SDSA	Synthesis-dependent stand annealing
SLiMs	Short Linear peptide Motifs
SSA	Single-strand annealing
SSBs	Single stranded DNA breaks

List of Tables

TABLE 1.1. DNA LESIONS GENERATED BY ENDOGENOUS AND EXOGENOUS DNA DAMAGE 4

TABLE 1.2. HUMAN GENETIC DISEASES ASSOCIATED WITH DDR DEFECTS 14

TABLE 1.3. HUMAN GENETIC DISEASES ASSOCIATED WITH DEFECTS IN THE REPLICATION STRESS RESPONSE.....64

TABLE 1.4. INSTANCES OF SLIMS AND THEIR REGULATORY ROLES 67

TABLE 2.1. PEPTIDE SEQUENCES USED IN THIS STUDY 73

TABLE 2.2. PRIMERS USED IN THIS STUDY 78

TABLE 3.1. SLIMS IDENTIFIED AND SYNTHESISED AS BIOTINYLATED PEPTIDES 95

TABLE 3.2 SLIMS WITHOUT HIGH-CONFIDENCE CANDIDATE INTERACTING PROTEINS 101

TABLE 3.3. SLIMS WITH DDR PROTEINS IN THE TOP TEN HIGH-CONFIDENCE CANDIDATE INTERACTING PROTEINS 102

TABLE 4.1. BLM ALLELE SEQUENCING 154

TABLE 5.1. TRIAL OF FOUR EXISTING SLIM DISCOVERY ALGORITHMS IN THE PREDICTION OF SLIMS IN BLM 181

TABLE 5.2. CALCULATION OF VALUES FOR THE DEVELOPMENT OF SLIM PREDICTION PARAMETERS 192

TABLE 5.3. PREDICTED SLIMS IN 120 DDR PROTEINS USING THE SLIMCON TOOL 200

TABLE 5.4. ALPHAFOLD PREDICTIONS OF PUBLISHED AND SYNTHESISED SLIMS.....214

TABLE 8.3.1. PROTEINS ASSIGNED TO THE BACKGROUND PROTEOME.....284

TABLE 8.7.1 LIST OF 120 DDR PROTEINS USED AS AN INPUT IN THE SLIMCON TOOL.....318

List of Figures

FIG. 1.1. CELLULAR RESPONSES TO DNA DAMAGE SIGNALLING. 6

FIG. 1.2. THE WIDE VARIETY OF DNA LESIONS NECESSITATES MULTIPLE DNA-REPAIR PATHWAYS. ...7

FIG. 1.3. EXPLOITATION OF DDR PATHWAYS TO ENHANCE THERAPEUTIC RESPONSES....24

FIG. 1.4. ATM INITIATES AND SUSTAINS CHROMATIN-BASED DDR SIGNALLING. 26

FIG. 1.5. REPAIR OF DSBS BY NHEJ. 30

FIG. 1.6. REPAIR OF DSBS BY HDR....32

FIG. 1.7. BIDIRECTIONAL MODEL OF DNA END RESECTION....34

FIG. 1.8. MECHANISMS OF DOUBLE HOLLIDAY JUNCTION (DHJ) DISSOLUTION AND RESOLUTION. 43

FIG. 3.1. BIOINFORMATIC AND PROTEOMICS PIPELINE FOR THE IDENTIFICATION AND SUBSEQUENT CHARACTERIZATION OF FUNCTIONALLY IMPORTANT SLIMS. 88

FIG. 3.2. GENERATION OF MANUAL SEQUENCE ALIGNMENTS. 91

FIG. 3.3. IDENTIFICATION OF SLIMS IN MANUAL SEQUENCE ALIGNMENTS. 93

FIG. 3.4. OPTIMIZATION OF THE PEPTIDE PULLDOWN METHOD. 97

FIG. 3.5. IDENTIFICATION OF HIGH-CONFIDENCE CANDIDATE INTERACTING PROTEINS. 100

FIG. 3.6. IDENTIFICATION OF CANDIDATE SLIM-BASED INTERACTIONS BETWEEN RPA, BRCA2 AND POLQ....103

FIG. 3.7. IDENTIFICATION OF CANDIDATE SLIM-BASED INTERACTIONS BETWEEN MRN AND 53BP1, AS WELL AS PCNA, ERCC6L2 AND RNF168....104

FIG. 3.8. IDENTIFICATION OF CANDIDATE SLIM-BASED INTERACTIONS BETWEEN BARD1 AND THE SMX TRI-NUCLEASE COMPLEX/SLX4IP....106

FIG. 3.9. IDENTIFICATION OF CANDIDATE INTERACTIONS OF PUTATIVE SLIMS FROM RECQL4. 107

FIG. 3.10. IDENTIFICATION OF CANDIDATE INTERACTIONS OF PUTATIVE SLIMS FROM BRCA1. 109

FIG. 3.11. IDENTIFICATION OF CANDIDATE INTERACTIONS OF A PUTATIVE SLIM FROM BRCA2 110

FIG. 3.12. IDENTIFICATION OF CANDIDATE INTERACTIONS OF A PUTATIVE SLIM FROM FANCM....111

FIG. 3.13. IDENTIFICATION OF CANDIDATE INTERACTIONS OF PUTATIVE SLIMS FROM POLQ....112

FIG. 3.14. IDENTIFICATION OF CANDIDATE INTERACTIONS OF PUTATIVE SLIMS FROM RECQL4 AND SHLD2 113
FIG. 3.15. IDENTIFICATION OF CANDIDATE INTERACTIONS OF PUTATIVE SLIMS FROM SLX4, TOPBP1 AND XRCC4. 114
FIG 3.16. MDC1 INTERACTS WITH TOPBP1 TO PROMOTE CHROMOSOMAL STABILITY DURING MITOSIS. 118
FIG. 3.17. TOPBP1 INTERACTS WITH CIP2A. 120
FIG. 3.18. THE C-TERMINUS OF CIP2A MEDIATES THE INTERACTION BETWEEN CIP2A AND TOPBP1 124
FIG. 3.19. HIGH-RESOLUTION AIRYSCAN IMAGING OF MRUBY3-CIP2A AND GFP-TOPBP1 FOCI IN MITOSIS. 125
FIG. 3.20. A SLIM IN THE C-TERMINUS OF AUNIP INTERACTS WITH CTIP. 127
FIG. 3.21. VALIDATION OF THE CTIP-BINDING SLIM IN THE CONTEXT OF FULL-LENGTH AUNIP 129
FIG. 3.22. IDENTIFICATION OF CONSERVED RPA-BINDING SLIMS IN THE BTR COMPLEX. 132
FIG. 4.1. THE CELLULAR FUNCTIONS OF THE BTR DISSOLVASOME COMPLEX. 142
FIG. 4.2. VALIDATION OF RPA-BINDING SLIMS IN THE CONTEXT OF FULL-LENGTH RMI1. 145
FIG. 4.3. VALIDATION OF RPA-BINDING SLIMS IN THE CONTEXT OF FULL-LENGTH BLM. 146
FIG. 4.4. THE RPA-BINDING SLIMS INTERACT COOPERATIVELY WITH RPA. 148
FIG. 4.5. THE RPA-BINDING SLIMS IN BLM AND RMI1 INTERACT WITH THE RPA1 N-TERMINAL OB-FOLD. 150
FIG. 4.6. GENERATION OF BLM/RMI1-DEFICIENT RPE-1 CELLS 153
FIG. 4.7. GENERATION OF A CELL-BASED SYSTEM TO TEST THE PHYSIOLOGICAL RELEVANCE OF BTR-RPA BINDING. 158
FIG. 4.8. RPA BINDING IS NOT REQUIRED FOR THE BTR COMPLEX TO SUPPRESS SCES BUT IS REQUIRED FOR THE RESTART OF STALLED REPLICATION FORKS. 160
FIG. 4.9. BTR-MEDIATED REPLICATION FORK RESTART IS DEPENDENT ON A PHYSICAL INTERACTION BETWEEN BLM AND TOP3A/RMI1/RMI2 AND ITS HELICASE ACTIVITY. 166

FIG. 4.10. MODEL FOR THE ROLE OF RPA BINDING IN SPECIFICALLY RECRUITING THE BTR COMPLEX TO STALLED REPLICATION FORKS. 173
FIG. 5.1. PREDICTION PERFORMANCE: OUTPUT OF SLIMS IN BLM PREDICTED BY THE MEME ALGORITHM 181
FIG. 5.2. PREDICTION PERFORMANCE: OUTPUT OF SLIMS IN BLM PREDICTED BY ELM. 183
FIG. 5.3. PREDICTION PERFORMANCE: OUTPUT OF SLIMS IN BLM PREDICTED BY SCANSITE4. 185
FIG. 5.4. GENERATION OF AUTOMATED SEQUENCE ALIGNMENTS FOR EVERY PROTEIN ENCODING GENE IN THE HUMAN GENOME. 188
FIG 5.5. EXAMPLE OF A SEQUENCE ALIGNMENT GENERATED BY THE SLIMCON TOOL. 189
FIG. 5.6. AUTOMATED IDENTIFICATION OF SLIMS. 191
FIG. 5.7. BATCH SEARCH FOR SLIMS IN DDR PROTEINS. 195
FIG. 5.8. ASSESSMENT OF THE ACCURACY OF THE SLIMCON TOOL: A SLIM PREDICTED IN WRN. ..197
FIG. 5.9. ASSESSMENT OF THE ACCURACY OF THE SLIMCON TOOL: A SLIM PREDICTED IN TOP3A.198
FIG. 5.10. 64% OF SLIMS PREDICTED BY THE SLIMCON TOOL ARE PREVIOUSLY UNCHARACTERISED PUTATIVE SLIMS. 199
FIG. 5.11 ALPHAFOLD PREDICTION OF BLM. 213
FIG. 8.2.1. SILVER STAIN GELS OF PEPTIDE PULLDOWNS CONDUCTED IN CHAPTER 3. 282
283
FIG. 8.4.1. FULL STRING NETWORKS OF INTERACTING PROTEINS PEPTIDE PULLDOWNS CONDUCTED IN CHAPTER 3. 288

Chapter 1: Introduction

1.1 The DNA Damage Response and its Relevance to Human Disease

1.1.1 The DDR: an Integrated Signalling and Genome Maintenance Network

In the face of various forms of endogenous and exogeneous sources of genotoxic stress, the repair of damaged DNA is essential for maintaining genome integrity and preventing diseases associated with genome instability (Jackson and Bartek, 2009). Accordingly, all known forms of life have evolved a complex repertoire of biochemical pathways to detect and repair such damage, collectively called the DNA damage response (DDR; Blackford and Jackson, 2017). These pathways include the DNA repair processes themselves, systems that regulate them and systems that integrate the DDR with the cell cycle (Harper and Elledge, 2007). The importance of the DDR in both promoting various physiological, cellular processes and circumventing human disease is demonstrated by the broad range of pathologies associated with mutations in DNA repair genes, such as cancer, immunodeficiency, infertility, premature ageing and microcephaly (Jackson and Bartek, 2009).

Like all biological pathways, the DNA damage response is facilitated by a series of protein-protein interactions, but the mechanisms by which DNA repair protein complexes are assembled, disassembled and regulated by post-translational modifications are poorly understood. Further understanding of the DDR at the molecular level will facilitate greater understanding of the diverse range of pathologies associated with mutations in DNA repair genes and the development of new therapies.

1.1.2 Sources of DNA Damage

The human body contains 10^{13} cells and every day each cell receives tens of thousands of DNA lesions (Jackson and Bartek, 2009). Sources of DNA damage are split into two types according to their origin: endogenous or exogenous (**Table 1.1**). Endogenous DNA damage can arise from reactive oxygen species (ROS), the typical by-products of the electron transport chain during aerobic respiration, but can also result from anabolic processes, peroxisomal metabolism and catabolic oxidases (Chatterjee and Walker, 2017; Harman, 1956; Imlay and Linn, 1988). Additionally, DNA alterations can occur due to dNTP misincorporation during DNA replication, modification of DNA bases by alkylation, interconversion between DNA bases by deamination and loss of DNA bases following DNA depurination (Ciccia and Elledge, 2010). In addition to DNA replication and repair, transcription is another essential process that can result in genome instability (Datta and Jinks-Robertson, 1995). For example, RNA-DNA hybrids formed co-transcriptionally compromise genome stability (Huertas and Aguilera, 2003). As DNA transcription and replication share a common template, R-loops (a three-stranded nucleic acid structure containing a RNA-DNA hybrid and a displaced ssDNA strand), can induce transcription-replication conflicts and transcription-associated recombination. Furthermore, alcohol-derived and endogenous aldehydes are capable of forming a diverse range of DNA lesions, from base adducts to DNA-DNA or DNA-protein crosslinks (Garaycoechea et al., 2018).

DNA damage can also arise from exogenous sources, when environmental agents damage the DNA (Chatterjee and Walker, 2017). The most pervasive environmental DNA damaging agent is ultraviolet light (UV; Rastogi et al., 2010) . Although the ozone layer absorbs the most dangerous region of the UV spectrum (UV-C), residual UV-A and UV-B light can induce up to 10^5 lesions in each exposed keratinocyte each hour. Typical lesions generated by UV light are pyrimidine dimers and pyrimidine (6-4)

photoproducts, resulting from covalent linkages between two adjacent pyrimidines (Ellison and Childs, 1981). Such lesions generate distortions in the DNA helix and can block DNA replication and transcription. Moreover, collapsed replication forks at sites of DNA containing pyrimidine dimers can generate DNA double strand breaks (DSBs; Limoli et al., 2002; Wang and Smith, 1986). Another environmental genotoxic agent is ionising radiation (IR), which can oxidise DNA bases and generate single stranded DNA breaks (SSBs) and DSBs (Chatterjee and Walker, 2017; Freifelder, 1965; Kaplan, 1966; McGrath and Williams, 1966). IR is produced by a diverse range of sources including radioactive decay of terrestrial radioisotopes present in natural materials such as rocks and soil, cosmogenic radioisotopes (UNSCEAR 2000) and medical treatments employing X-rays or radiotherapy. Thirdly, chemical agents such as those used in cancer therapy can cause a variety of DNA lesions, including alkylating agents such as methyl methanesulfonate, which can attach alkyl groups to DNA bases (Brookes and Lawley, 1961). Crosslinking agents such as mitomycin C (Iyer and Szybalski, 1963, 1964) and cisplatin (Fichtinger-Schepman et al., 1985; Rosenberg et al., 1969) induce intrastand or interstrand crosslinks (Kelland, 2007), whilst the topoisomerase inhibitors camptothecin (Hsiang et al., 1985) and etoposide (Osheroff, 1989) inhibit topoisomerase I or II respectively and induce the formation of SSBs or DSBs by trapping topoisomerase-DNA cleavage complexes (Wu et al., 2011). Lastly, tobacco products are one of the most prevalent carcinogens in our society, containing approximately 50 carcinogenic chemicals and cause a wide variety of adducts and oxidative damage to DNA in the lung, oral cavity and adjacent tissues (Phillips et al., 1988). Considering DNA adducts alone, up to 6000 adducts per cell are estimated to be present in smokers lungs, oral cavities and adjacent tissues.

Table 1.1. DNA Lesions Generated by Endogenous and Exogenous DNA Damage (adapted from Ciccia and Elledge, 2010)			
Endogenous DNA Damage	**DNA Lesions Generated**	**Number Lesions/Cells/Day**	
Depurination	AP site	10000	
Cytosine deamination	Base transition	100-500	
SAM-induced methylation	3meA	600	
	7meG	4000	
	O^6meG	10-30	
Oxidation	8oxoG	400-1500	
Exogenous DNA Damage	**Dose Exposure (mSv)**	**DNA Lesions Generated**	**Estimated Number Lesions/Cell**
Sunlight (peak hr)	-	Pyrimidine dimers, (6-4) photoproducts	100,000/day
Cigarette smoke	-	Aromatic DNA adducts	45-1029
Chest X-rays	0.02	DSBs	0.0008
Dental X-rays	0.005	DSBs	0.0002
CT	7	DSBs	0.28
Mammogram	0.4	DSBs	0.016
Tumour PET scan (^{18}F)	10	DSBs	0.4
Airline travel	0.005/hr	DSBs	0.0002/hr

1.1.3 Signalling in Response to DNA Damage

In general, the DDR involves recognition of the DNA lesion, followed by the initiation of a cellular signalling cascade that promotes DNA repair (Blackford and Jackson, 2017; Ciccia and Elledge, 2010; Jackson and Bartek, 2009). The DDR is primarily mediated via proteins of the phosphatidylinositol 3-kinase like protein kinase (PIKK) family: ATM (Savitsky et al., 1995), ATR (Bentley et al., 1996; Cimprich et al., 1996) and DNA-PK (Carter et al., 1990; Jackson et al., 1990; Lees-Miller et al., 1990). ATM and DNA-PK respond principally to DSBs, whilst ATR is activated in response to a wider range of genotoxic stresses, including DSBs, base adducts, crosslinks and DNA replication stress (Cimprich and Cortez, 2008). ATR is recruited to extended tracts of ssDNA coated by the replication protein A (RPA) complex (Zou and Elledge, 2003a), explaining its wider range of substrates. RPA-coated ssDNA is generated via helicase-polymerase uncoupling at stalled replication forks (Byun et al., 2005a) or by nucleolytic processing of DNA damage (Raderschall et al., 1999).

ATM and ATR function in concert in response to DSBs. ATM is required for efficient resection of DSBs (Jazayeri et al., 2006; Myers and Cortez, 2006), a process required for ATR activation (Zou and Elledge, 2003a). Furthermore, progressive resection of DSBs promotes an ATM-to-ATR switch (Shiotani and Zou, 2009). Together, these processes allow the levels of ATM and ATR activation to be fine-tuned by DSB resection.

In addition to the PIKK family, members of the poly(ADP-ribose) polymerase (PARP) family are key regulators of the DNA damage response (Schreiber et al., 2006). Of the 16 members of the PARP family, PARP1 (Benjamin and Gill, 1980; Chambon et al., 1963; Fujimura et al., 1967; Nishizuka et al., 1967) and PARP2 (Amé et al., 1999) are DNA damage-dependent PARPs, activated by SSBs and DSBs. In response to DNA breaks, PARP1/2 use NAD^+ to catalyse the addition of ADP-ribose units onto the γ-carboxyl group of Glu residues of acceptor proteins or on PARP1 itself (Ferro and Olivera, 1982; Schreiber et al., 2006). The resulting polymer of ADP ribose can recruit various proteins involved in the cellular response to DNA damage, such as the SSB repair factor XRCC1 (El-Khamisy et al., 2003).

DNA damage signalling induces a wide variety of cellular processes including DNA repair, checkpoint activation, as well as modulation of chromatin structure and transcription (**Fig 1.1;** Blackford and Jackson, 2017). In multi-cellular organisms, cells with extensive DNA damage can permanently exit the cell cycle (cellular senescence) or undergo programmed cell death (apoptosis), inhibiting the propagation of mutated cells which may lead to cancer or other age-related pathologies.

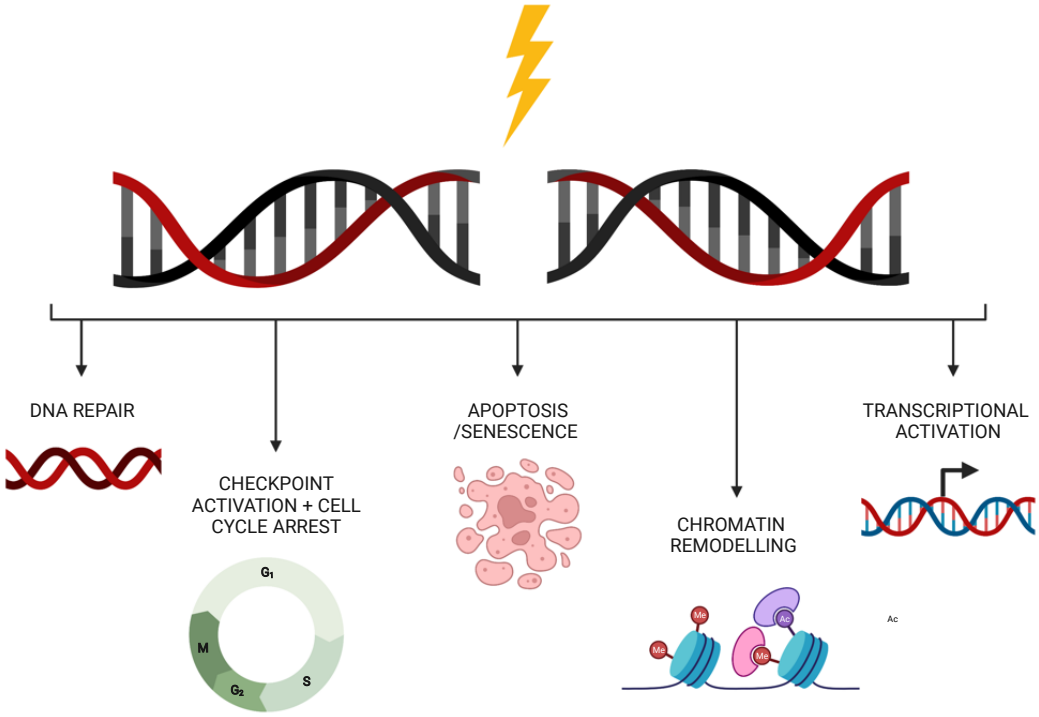


Fig. 1.1. Cellular responses to DNA damage signalling.
DNA damage signalling induces a variety of cellular processes including DNA repair, checkpoint activation, apoptosis or senescence, as well as modulation of chromatin structure and transcriptional activation. Created with BioRender.com.

1.1.4 Pathways of DNA Repair

The plethora of endogenous and exogenous assaults that exist result in a wide variety of DNA lesions (**Table 1.2**), necessitating multiple DNA-repair pathways to ensure the maintenance of genome integrity (**Fig. 1.2**; Ciccia and Elledge, 2010).



Fig. 1.2. The wide variety of DNA lesions necessitates multiple DNA-repair pathways.
The plethora of endogenous and exogenous assaults that exist result in a wide variety of DNA lesions, necessitating multiple, largely distinct DNA-repair pathways to ensure the maintenance of genome integrity. Created with BioRender.com.

Mis-paired DNA bases are replaced by correct bases via the mismatch repair pathway (Kunkel and Erie, 2015). In mismatch repair, mismatches are recognised by the MSH2-MSH6 complex (MUTSα; Gradia et al., 1999). Once bound to the mismatch, the MSH2-MSH6 complex undergoes a conformational change, encircling the DNA as a clamp (Qiu et al., 2012). The MSH2-MSH6 complex recruits the PMS2-MLH1 complex (MUTLα; Liu et al., 2016). Following mismatch recognition and strand discrimination, the PMS2-MLH1 complex incises the nascent strand in a manner dependent on PCNA and RFC (Kadyrov et al., 2006; Umar et

al., 1996). The nascent strand is degraded by EXO1 in a 5' to 3' direction (Genschel et al., 2002), or excised by the 3'-5' activity of the replicative polymerases (Longley et al., 1997; Morrison and Sugino, 1994). Other complexes can also contribute to MMR, such as the MSH2-MSH3 complex that can repair more complex insertion/deletion mutations (Palombo et al., 1996).

Small chemical DNA base alterations are repaired using the base excision repair pathway (Krokan and Bjørås, 2013). The damaged base is excised by a DNA glycosylase, leaving an AP site that is cleaved by AP endonuclease (APE-1), generating a SSB (Mol et al., 2000). Each DNA glycosylase recognises and removes specific DNA base modifications, for example the preferred substrate for uracil-DNA glycosylase (UNG) is a mis-incorporated uracil or fluorouracil (Lindahl, 1974). Gap filling and DNA ligation ensues, via short patch repair (XRCC1, LIGIIIα and POLβ) or long-patch repair (PCNA, LIGI, POLβ, POLδ and FEN1; Krokan and Bjørås, 2013).

Pyrimidine dimers and pyrimidine (6-4) photoproducts generated by UV light are repaired by nucleotide excision repair (Spivak, 2015). Nucleotide excision repair operates via two sub-pathways: transcription-coupled nucleotide excision repair (targeting lesions blocking transcription; TC-NER) and global-genome nucleotide excision repair (targeting bulky helix-distorting lesions; GG-NER). In both pathways, the first step is DNA damage recognition. In GG-NER, XPC detects bulky lesions (Sugasawa et al., 1998) and XPE/ DDB1 detects minor disruptions of the DNA helix (Reardon et al., 1993). XPC binds the ssDNA within the distortion and recruits TFIIH. TC-NER recognises and repairs DNA lesions within actively transcribed genes (Spivak, 2015). During unperturbed transcription, CSB and UVSSA-USP7 transiently interact with RNA POLII. After collision with a DNA adduct, RNA POLII stalls and strongly associates with CSB and UVSSA-USP7 (Nakazawa et al., 2012; Schwertman et al., 2012). CSA and TFIIH are then recruited. After TFIIH recruitment, the two pathways converge (Spivak, 2015). XPB and XPD, subunits of

TFIIH, unwind the DNA, generating a 20-30 bp bubble. XPA and XPD verifies the presence of a DNA adduct. 5' DNA incision (XPA/RPA directs XPF/XRCC1 endonuclease to incise the DNA) and 3' DNA incision (XPG) ensues, followed by DNA repair synthesis and DNA ligation by LIG1.

Interstrand crosslinks generated by chemotherapeutic agents or endogenous aldehydes are excised and repaired by the Fanconi anaemia (FA) pathway (Lopez-Martinez et al., 2016). The interstrand crosslink is recognised when two stalled replication forks converge on the ICL covalently linking two DNA strands (Zhang et al., 2015). The FANCM-FAAP24-MHF1/2 complex recognises the DNA lesion, recruiting the FA core complex (Yan et al., 2010). FANCL, an E3 ubiquitin ligase, is recruited as part of the multimeric core complex. This results in the mono-ubiquitination of FANCD2 and FANCI, the key regulatory step in the FA pathway (Meetei et al., 2003). Mono-ubiquitylated FANCD2-FANCI promotes the recruitment of nucleases, such as SLX4 (a scaffolding protein for the DNA endonucleases MUS81, SLX1 and XPF). The DNA contiguous to the ICL is cleaved (Klein Douwel et al., 2014), generating a DNA adduct, resolved by the nucleotide excision repair pathway, and a double-strand break, resolved by homologous recombination (Lopez-Martinez et al., 2016).

DSBs, generated by genotoxic agents such as IR, are repaired by one of two mechanisms: pathways based on non-homologous end joining (NHEJ) or homology directed repair (HDR). Due to their relevance to this thesis, DSB repair pathways will be outlined in greater depth in the next section.

1.1.5 DNA Damage Checkpoint Signalling in Interphase and Mitosis

Although some types of DNA damage can be repaired without cells halting the cell cycle (checkpoint activation), some lesions (such as DNA DSBs) are highly toxic and can induce genome instability, particularly if left unresolved when chromosome segregation occurs (Blackford and Stucki, 2020; Jackson and Bartek, 2009). Accordingly, DNA damage signalling is required to regulate cell cycle progression to facilitate timely repair of DNA damage.

Cell cycle progression in eukaryotic cells is controlled by sequential, transient activation of cyclin-dependent kinases (CDKs), which facilitate the ordered succession of the four cell cycle phases: G1, S, G2, M (Gérard and Goldbeter, 2009; Lee and Nurse, 1987; Nurse, 1975). Cyclins are expressed during specific stages of the cell cycle, generating cell cycle stage-specific cyclin-CDK complexes (Evans et al., 1983; Siddiqui et al., 2013). The cyclin subunit confers substrate specificity, resulting in cell cycle stage-dependent phosphorylation of target proteins. In G1, Cyclin D/CDK4-6 (Matsushime et al., 1992; Meyerson and Harlow, 1994) and Cyclin E/CDK2 (Dulić et al., 1992; Koff et al., 1992) promote S-phase entry, in S phase Cyclin A/CDK2 (Cardoso et al., 1993; Pagano et al., 1992) activate DNA replication, and Cyclin A/CDK1 (Pagano et al., 1992) and Cyclin B/CDK1 (Draetta et al., 1988; Pines and Hunter, 1989) promote mitosis. Another key mechanism of cyclin-CDK regulation results from the E3 ligases, the APC/C (King et al., 1995; Sudakin et al., 1995; Tugendreich et al., 1995) and the SCF complex (Bai et al., 1996; Zhang et al., 1995). E3 ligases polyubiquitinate proteins targeted to them via a substrate adaptor, which are then degraded by the proteosome (Hershko et al., 1983). The peak activity of specific CDK complexes and the E3 ligases that regulate them are temporally mutually exclusive.

Three cell cycle checkpoints halt cell cycle progression, ensuring proper cell division in eukaroytic cells in response to DNA damage: the G1/S checkpoint, the intra-S

checkpoint and the G2/M checkpoint (Kastan and Bartek, 2004). The G1/S and G2/M cell cycle checkpoints arrest cells at G1/S or G2/M boundaries. In contrast, to slow S phase progression, the intra-S phase checkpoint limits firing of DNA replication origins, increases cellular nucleotide concentrations and stabilizes stalled replication forks (Chao et al., 2017). This maintains the fidelity of DNA replication rather than halts cell cycle progression. There are differences in how these checkpoints are activated, but all share essentially the same end-point, the inhibition of the CDK activity driving cell cycle progression (Blackford and Stucki, 2020). ATM/ATR -dependent activation of checkpoint kinases inhibits CDK activity, via the inactivation of the CDC25 family of phosphatases (Falck et al., 2001; Peng et al., 1997; Sanchez et al., 1997). CDC25 phosphatases counteract the inhibitory phosphorylation of CDKs (Dunphy and Kumagai, 1991; Galaktionov and Beach, 1991; Russell and Nurse, 1986).

If effective DNA repair occurs, the DDR is inactivated, and the cell cycle resumes (Kastan and Bartek, 2004). However, if the DNA damage is too extensive or remains unresolved, cells can permanently exit the cell cycle (cellular senescence) or undergo programmed cell death (apoptosis). Both inhibit tumorigenesis. The tumour suppressor p53 is activated via the ATM-CHK2 axis in response to DSBs (Hirao et al., 2000). Activation of p53 induces cell cycle arrest, senescence or apoptosis in response to DNA damage by transcriptionally regulating (amongst others): the CDK inhibitor p21 (El-Deiry et al., 1993; Harper et al., 1993) and the pro-apoptotic factors BAX (Toshiyuki and Reed, 1995) and PUMA (Nakano and Vousden, 2001; Yu et al., 2001). In addition to apoptosis, DNA damage can induce cell death by programmed necrosis (also known as caspase-independent cell death or necroptosis; Wang et al., 2017). Necroptosis is morphologically similar to necrosis, but is regulated, strictly controlled upstream molecular signalling pathways. The initiation of necroptosis requires several different stimuli, including RIP1 and RIP3 (Cho et al., 2009; Kelliher et al., 1998).

After prophase, there is no DNA damage checkpoint until the cell enters G1 of the next cell cycle (Blackford and Stucki, 2020; Rieder and Cole, 1998). High levels of DNA damage can delay the metaphase to anaphase transition, but this is induced by defects in microtubule attachment to kinetochores and an activated spindle assembly checkpoint rather than direct DNA damage signalling (Mikhailov et al., 2002). Several mechanisms prevent CDK1 inhibition in mitosis: transcription is attenuated during mitosis (Prescott and Bender, 1962), so p21 induction and subsequent CDK inhibition is disabled; in early mitosis, WEE1 is phosphorylated by CDK1 and PLK1, resulting in SCF-βTrCP ubiquitin ligase-dependent degradation (Watanabe et al., 2004). Claspin, required for CHK1 activation (Kumagai and Dunphy, 2000), is phosphorylated by PLK1, resulting in its SCF-βTrCP ubiquitin ligase-dependent degradation (Mailand et al., 2006; Mamely et al., 2006; Peschiaroli et al., 2006). Lastly, PLK1-dependent phosphorylation of CHK2 inhibits CHK2 activation (Vugt et al., 2010). As a result, CDK1-Cyclin B activity is maximal at the end of prophase and cannot be inhibited by ATM/ATR after mitosis.

1.1.6 The DNA Damage Response and Human Disease

The importance of the DDR is demonstrated by the broad range of pathologies associated with mutations in DNA repair genes (Jackson and Bartek, 2009). Patients with hereditary syndromes associated with mutations in DNA repair genes present with neurological defects, immunodeficiency, infertility, premature aging and/or cancer predisposition (**Table 1.2**).

1.1.6.1. Neurological Defects

The nervous system relies heavily on a functional DDR, and patients with hereditary DDR syndromes present with various neurological defects (Ciccia and Elledge, 2010; Jackson and

Bartek, 2009). Neurons become terminally differentiated and post-mitotic early in development but must remain functional for decades (Aranda-Anzaldo, 2012). Neurons have high levels of mitochondrial respiration, leading to ROS accumulation (Lu et al., 2004; Weissman et al., 2007). The ensuing oxidative damage to nuclear and mitochondrial DNA is catastrophic in post-mitotic neurons, as the damaged cells are not replaced. Consistently, DDR defects often result in neuronal death and neurodegeneration (Lu et al., 2004). Several DDR syndromes affect the cerebellum, responsible for motor coordination (Allen et al., 1997; Ciccia and Elledge, 2010). Degeneration of cerebellar neurons can result in impaired motor coordination (ataxia), an eye movement defect (oculomotor apraxia) and speech disorders (dysathria). Ataxia telangiectasia is a neurological syndrome, resulting from hereditary biallelic mutations in *ATM* (Boder and Sedgewick, 1958; Savitsky et al., 1995). Ataxia telangiectasia patients develop ataxia and are confined to a wheelchair within the first ten years of life. Similar symptoms are observed in patients with the ataxia telangiectasia-like syndrome, caused by mutations in *MRE11* (Stewart et al., 1999). Ataxia telangiectasia-like syndrome patients have a later onset and slower progression of symptoms but the similarity between the syndromes is consistent for the role of MRE11 in ATM activation (Uziel et al., 2003), described in the next section.

Table 1.2. Human Genetic Diseases Associated with DDR Defects (adapted from Ciccia and Elledge, 2010)

Syndrome	Mutated Gene	DDR Defect	Phenotype				
			Neurological Defects	Immunodeficiency	Progeria	Cancer	Other
Cockayne syndrome	*CSA, CSB, XPB, XPD, XPG*	TC-NER	Microcephaly, neurodegeneration, neuronal deamination	-	Cachexia, cataracts, hearing loss, retinopathy	-	Photosensitivity, growth defects
Xeroderma pigmentosum	*XPA-G, POLH*	NER	Microcephaly, neurodegeneration	-	-	Squamous and basal cell carcinoma, melanoma	Photosensitivity
Fanconi anaemia	*FANC* genes	ICL repair	Microcephaly	Pancytopenia	Bone marrow failure	Acute myeloid leukaemia, squamous cell carcinoma (head and neck)	Abnormal skin pigmentation, infertility, limb deformities
Familial breast cancer	*BRCA1, BRCA2, FANCJ, CHK2, NBS1, PALB2, RAD50, RAD51C*	HR, chromatin response to DSBs	-	-	-	Breast, ovarian, pancreatic, prostate cancer	-
Bloom syndrome	*BLM*	HR	Microcephaly, mild learning defects	Immunoglobulin deficiency	-	All cancer types	Photosensitivity, infertility, growth defects, cafè-au-lait spots
Werner syndrome	*WRN*	HR, BER, telomere maintenance	-	-	Atherosclerosis, cataracts, grey hair, osteoporosis	Sarcomas	Type II diabetes, growth defects
Lynch syndrome	*MSH2, MSH6, MLH1, PMS2*	MMR	-	-	-	Gastrointestinal (i.e., colorectal), ovarian cancer	-
LIG4 syndrome	*LIG4*	NHEJ	Microcephaly	Hypogammaglobulinemia, lymphopenia	-	Acute lymphoblastic leukaemia, lymphomas	Growth defects

Table 1.2. Human Genetic Diseases Associated with DDR Defects (adapted from Ciccia and Elledge, 2010)							
Syndrome	**Mutated Gene**	**DDR Defect**	**Phenotype**				
			Neurological Defects	**Immunodeficiency**	**Progeria**	**Cancer**	**Other**
Cernunnos deficiency	*NHEJ1*	NHEJ	Microcephaly	Immunodeficiency	-	-	Growth defects
Radiosensitive severe combined immunodeficiency	*DCLRE1*	NHEJ	-	Immunodeficiency	-	Lymphoma	Growth defects
Microcephaly, intractable seizures, and development delay syndrome	*PNKP*	NHEJ, SSB repair	Microcephaly	-	-	-	Seizures, growth defects
Ataxia telangiectasia	*ATM*	Chromatin response to DSBs	Ataxia, cerebellar degeneration, oculomotor apraxia	Immunodeficiency	-	Lymphoma, leukaemia, breast cancer	Dilated blood vessel, infertility, metabolic defects, growth defects
Ataxia telangiectasia-like disorder	*MRE11*	Chromatin response to DSBs, HR	Ataxia, cerebellar degeneration, oculomotor apraxia	Immunodeficiency	-	-	-
Nijmegen breakage syndrome	*NBS1*	Chromatin response to DSBs, HR	Microcephaly	Immunodeficiency	-	B cell lymphoma	Facial dysmorphism, growth defects
Nijmegen breakage syndrome- like disorder	*RAD50*	Chromatin response to DSBs, HR	-	Immunodeficiency	-	-	Facial dysmorphism, growth defects
Riddle syndrome	*RNF168*	Chromatin response to DSBs, HR	Mild	Immunodeficiency	-	-	Facial dysmorphism, growth defects

The DDR is also crucial for brain development (Ciccia and Elledge, 2010; O'Driscoll and Jeggo, 2008). Patients of various DDR syndromes display a reduced head circumference (microcephaly), resulting from defective proliferation and/or apoptosis of neuroprogenitor cells during foetal development. The gross cerebral architecture in the microcephalic brain is preserved but is smaller. Microcephaly is a typical phenotype of Seckel syndrome patients (McKusick et al., 1967; Seckel., 1960). One form of Seckel syndrome is caused by biallelic hypomorphic *ATR* mutations, which results in reduced ATR protein levels due to aberrant splicing of ATR pre-mRNA (O'Driscoll et al., 2003). Syndromes with defective HDR (such as Bloom syndrome), ICL repair (FA) and disorders exhibiting defective replication fork restart after DNA damage (Schimke immunosseuos dysplasia) also lead to microcephaly (Ciccia and Elledge, 2010). These observations have led to the proposal that the DDR is crucial for the developing brain and inhibiting neurodegeneration.

1.1.6.2. Immunodeficiency

An effective adaptive immune response requires recognition of a wide variety of foreign antigens, requiring the generation of genetically diverse cells with unique major histocompatibility complexes (Gennery, 2006). In vertebrates, this is facilitated by fusion of different combinations of variable (V), diversity (D) and junction (J) DNA sequences via V(D)J recombination (Brack et al., 1978; Hozumi and Tonegawa, 1976; Seidman et al., 1978; Weigert et al., 1978). Each V(D)J gene segment is flanked by a recombination-signal sequence and recombination is initiated by the nucleases RAG1 and RAG2, which introduce site-specific DSBs at the junction between the coding exon and the recombination-signal sequence (Agrawal and Schatz, 1997; Oettinger et al., 1990; Schatz et al., 1989). These DSBs are processed by NHEJ (Lieber et al., 2004).

V(D)J variable domains are further diversified by class switch recombination (CSR) and somatic hypermutation (Davis et al., 1980; Muramatsu et al., 2000; Snapper and Paul, 1987; Stavnezer et al., 2008; Wang et al., 1970). CSR is a DNA rearrangement process whereby the constant regions of immunoglobins are switched and somatic hypermutation introduces point mutations in the V(D)J variable domains. Both processes are promoted by the activation-induced deaminase (AID) enzyme which deaminates cytosine to form uracil (Muramatsu et al., 2000; Revy et al., 2000). This results in SSB formation. In CSR, SSBs are converted into DSBs by the uracil DNA glycosylase and APE1. These DSBs are processed by NHEJ, leading to the exchange of antibody constant regions. In contrast, in somatic hypermutation, error-prone repair of the SSBs yields mutations in the variable exon. As such, patients of syndromes associated with defective NHEJ present with severe immunodeficiency caused by severe lymphopenia, such as those with Artemis (Jones et al., 1991; Moshous et al., 2001), LIG4 (Burg et al., 2006; O'Driscoll et al., 2001; Riballo et al., 1999) or XLF deficiencies (Ahnesorg et al., 2006; Buck et al., 2006; Dai et al., 2003). Furthermore, immunodeficiency has been reported in ataxia telangiectasia patients (Chun and Gatti, 2004), ataxia telangiectasia-like syndrome patients (Lähdesmäki et al., 2004), Nijmegen breakage syndrome (Seeman et al., 2004) and Riddle syndrome (Stewart et al., 2009), caused by biallelic mutations in *ATM, MRE11, NBS1* and *RNF168* respectively. These observations demonstrate the importance for the DDR in the development of an effective adaptive immune system.

1.1.6.3. Infertility

The DDR plays an important role in gamete generation during meiosis (Neale and Keeney, 2006). The meiosis-specific nuclease SPO11 generates DSBs, which are repaired by HDR (Romanienko and Camerini-Otero, 1999). This results in an exchange of genetic material

between homologous chromosomes and increased genetic diversity. The process of HDR and how it can result in exchange of genetic material will be detailed further in section 1.2. Defective DSB repair in meiosis results in infertility (Ciccia and Elledge, 2010). Patients from many DDR syndromes, such as ataxia telangiectasia, Bloom syndrome and Fanconi anaemia display infertility or sub-fertility.

1.1.6.4. Premature Aging and Stem Cell Exhaustion

Aging is characterised by the progressive deterioration of physiological integrity, resulting in impaired function and death (López-Otín et al., 2013). There are several hallmarks of aging, including genome instability, telomere attrition, cellular senescence and stem cell exhaustion.

Time-dependent accumulation of DNA damage is thought to be the general cause of aging, as somatic mutations accumulate with age in the nuclear and mitochondrial genomes (Atamna et al., 2000; Mandavilli and Rao, 1996; Szilard, 1959; Yamamoto et al., 1988). Another hallmark of aging is telomere attrition, leading to cellular senescence (Bodnar et al., 1998a; Harley et al., 1990). The ends of eukaryotic chromosome resemble DSBs, and these ends can activate ATM/ATR signalling, eliciting aberrant NHEJ-mediated chromosome fusions (Karlseder et al., 1999; Steensel et al., 1998). This is mitigated by nucleoprotein structures called telomeres, bound by the six-subunit Shelterin complex (Liu et al., 2004; Ye et al., 2004). However, replicative DNA polymerases cannot replicate the ends of telomeres, due to the so-called "DNA end-replication problem" (Olovnikov, 1973). DNA replication only occurs in a 5' to 3' direction and so the lagging strand is replicated discontinuously. For lagging strand DNA replication, short RNA primers are generated by DNA polymerase α (Tsurimoto and Stillman, 1990, 1991). These are extended by DNA polymerase δ to form Okazaki fragments (McElhinny et al., 2008). When the RNA primers are removed, the lagging strand sequence that is complementary to the extreme end of the chromosome cannot

be synthesised. In germ cells, the end-replication problem is solved by the addition of species-specific repeats (TTAGGG in vertebrates) to the end of telomeres (Greider and Blackburn, 1987), synthesised by a specialised DNA polymerase, telomerase. Telomerase contains an RNA component, providing the template for synthesising telomere repeats. However, most mammalian somatic cells do not express telomerase, leading to progressive telomere loss until cells enter replicative senescence (Bodnar et al., 1998b). Senescence is considered to be beneficial, eliminating damaged and potentially oncogenic cells (López-Otín et al., 2013). However, it requires efficient cellular replacement, involving clearance of senescent cells and proliferation of progenitors to re-establish cell numbers. In aged organs, this may be inefficient, or exhaust the regenerative capacity of stem cells, aggravating the damage and contributing to aging. Finally, the decline in the regenerative potential of tissues or stem cell exhaustion is a hallmark of aging (López-Otín et al., 2013). For example, studies in aged mice revealed old haemopoietic stem cells undergo fewer cell divisions than younger haemopoietic stem cells, correlating with levels of DNA damage (Rossi et al., 2007). Further work identified that replication stress drives functional decline in aged haemopoietic stem cells (Flach et al., 2014). Aged haemopoietic stem cells have higher levels of replication stress, cell cycle defects and chromosome gaps or breaks, caused by a decreased expression of MCM helicase components and altered DNA replication fork dynamics (Flach et al., 2014).

The link between DNA damage and ageing is supported by several DDR syndromes with premature aging (progeroid) phenotypes (Ciccia and Elledge, 2010). A key progeroid syndrome linked to the DDR is Werner Syndrome (Yu et al., 1996). Werner Syndrome is caused by a deficiency in the the RecQ helicase WRN, which promotes

HDR and telomere maintenance (Chu and Hickson, 2009). Patients develop osteoporosis, cataracts, type II diabetes, grey hair and atherosclerosis prematurely and a predisposition to cancer and cardiovascular disease. Another example of a progeroid syndrome is Cockayne syndrome (Neill and Dingwall, 1950), caused by defects in TC-NER and characterised by hearing loss, cataracts, weight loss and muscle atrophy (Garinis, 2008). Lastly, FA patients present with bone marrow failure and pancytopenia (Kee and D'Andrea, 2012). Bone marrow failure is caused by p53-mediated clearance of haemopoietic stem cells (Garaycoechea et al., 2018). Primary protection against endogenous and alcohol-derived aldehydes is conferred by aldehyde detoxification, such as that catalyzed by ALDH2. However, if aldehyde detoxification pathways are lost or overwhelmed, aldehydes damage DNA. The FA pathway is the principal mechanism to repair this damage, which is compromised in FA patients. Haemopoietic stem cells mutated by aldehydes are functionally compromised and display a myeloid bias, explaining the prevalence of acute myeloid leukaemia among Fanconi anaemia patients (Garaycoechea et al., 2018).

1.1.6.5. Cancer

Genomic instability is a hallmark of cancer (Hanahan and Weinberg, 2011). Maintenance of genome integrity by the DDR is critical in preventing tumorigenesis, indicated by the increased cancer predisposition present in many DDR syndromes (Ciccia and Elledge, 2010).

A well characterised cancer-predisposing syndrome is hereditary non-polyposis colon cancer or Lynch syndrome, in which heterozygous mutations in MMR genes, such as *MLH1* (Papadopoulos et al., 1994), *MSH2* (Fishel et al., 1993), *MSH6* (Miyaki et al., 1997), *PMS2* (Nicolaides et al., 1994) cause microsatellite instability (Negrini et al., 2010).

Xeroderma pigmentosum patients, with defects in the NER pathway, have an increased risk of skin cancer due to defective repair of UV lesions following sun exposure (Cleaver, 1968). Also, although various factors affect breast and ovarian cancer predisposition, the most powerful indicator is inherited mutations in the HDR genes *BRCA1* and *BRCA2* (Miki et al., 1994; Wooster et al., 1995). Women with heterozygous mutations of *BRCA1* or *BRCA2* have a 40-80% lifetime risk of developing breast cancer (Easton et al., 1995; Schubert et al., 1997). Mutations in various other genes, including *ATM, CHK2, MRE11, NBS1, RAD50, PALB2* and *BARD1* also increase breast cancer risk. This highlights the importance of HDR in breast cancer prevention (Frey et al., 2017). Biallelic mutations in FA genes is linked with increased cancer risk, in particular the development of acute myeloid leukaemia and head and neck squamous cell cancers (Kee and D'Andrea, 2012). The risk of AML in FA patients is approximately 800-fold higher than that of the general population, with a median age of onset of 14 years of age. These inherited cancer-predisposing disorders are associated with malignancies of specific tissue types. In contrast, Bloom syndrome patients, with biallelic mutations in the HDR gene *BLM*, are predisposed to the full spectrum of malignancies found in the general population, at roughly the same relative frequencies (German, 1993). Some studies have found that heterozygous *BLM* mutations increase cancer predisposition (Bononi et al., 2020; Goss et al., 2002; Schayek et al., 2017), although not all studies have found this (Baris et al., 2007; Cleary et al., 2003)

1.1.7 Harnessing the DDR to Treat Cancer

Cancer is the second most-frequent cause of death worldwide (Bray et al., 2018). Ironically, whilst DNA damage and its repair can drive cancer development, a defective

DDR is an Achilles' heel of tumours. Cancer cells with defective DDR mechanisms exhibit high sensitivity to DNA damaging therapeutic agents (Hoeijmakers, 2009). Other than surgery, the most prevalent cancer treatments are radiotherapy and chemotherapy. These treatments function by introducing DNA damage, which is cytotoxic to highly proliferating cancer cells. IR results in the generation of SSBs and DSBs (Freifelder, 1965; Kaplan, 1966; McGrath and Williams, 1966) and there are various chemotherapeutic agents that induce specific types of DNA damage, dependent on their mechanism of action. For example, bifunctional alkylators such as mitomycin C (Iyer and Szybaski, 1963, 1964) and cisplatin (Fichtinger-Schepman et al., 1985; Rosenberg et al., 1969) induce DNA crosslinks. Due to defects in the DDR, the inability of cancer cells to efficiently and rapidly repair high levels of DNA damage results in cancer cell death (Barnieh et al., 2021).

A key mechanism of chemotherapeutic drug resistance is the upregulation of broad spectrum ATP-dependent efflux pumps (Robey et al., 2018). Three well characterised examples are: multidrug resistance protein 1 (MDR1), encoded by *ABCB1* (Roninson et al., 1986; Ueda et al., 1986); multidrug resistance-associated protein 1 (MRP1), encoded by *ABCC1* (Cole et al., 1992)*;* and ABCG2 (also known as breast cancer resistance protein), encoded by *ABCG2* (Allikmets et al., 1998; Doyle et al., 1998). When these ATP-dependent efflux pumps are upregulated, drugs are effluxed from cancer cells, resulting in a decreased intracellular drug concentration and therefore efficiency (Robey et al., 2018).

An alternative mechanism of drug resistance is the re-activation of DDR mechanisms, allowing cancer cells to escape the potential lethal effects of these therapies (Barnieh et al., 2021). This perhaps explains why tumour resistance often arises in solid tumours after prolonged treatment and why this resistance is often accompanied by high

expression of DDR proteins (Oliver et al., 2010). For example, glioma stem cells are resistant to treatment and display high expression of DDR proteins, which could explain why glioblastoma is refractory to treatment (Bao et al., 2006). Such findings underscore the critical role of the DDR in the development and treatment of cancer.

Despite their efficacy, conventional cancer therapies will damage non-cancerous tissue (Jackson and Bartek, 2009). Despite recent advances, such as the CyberKnife technology that delivers high precision, high dose radiotherapy, IR damages normal tissue in close proximity to cancer cells (Kurup, 2010). Chemotherapeutic agents target all fast-growing cells, resulting in the common side effects experienced (such as hair loss and nausea). Accordingly, in the past twenty years, targeted cancer therapies have become the focus of anticancer drug development. Targeted cancer therapies act on specific molecular targets associated with cancer, theoretically limiting normal tissue damage.

One type of targeted cancer therapy is immunotherapy. The idea to harness the immune system to treat cancer originated in the nineteenth century (Waldman et al., 2020). Currently there are three main categories of immunotherapeutic approaches to treat cancer: immune checkpoint blockade, designed to unleash T cell responses; adoptive cellular therapies, based upon infusion of tumour-fighting immune cells; and cancer vaccines. However, these therapies are limited as they tend to invoke an autoimmune response.

The DDR has become a promising strategic target for the development of novel cancer therapies. The functional loss of DDR proteins often results in a greater reliance on residual DDR proteins to maintain cellular survival than in normal tissues, generating a "synthetic lethality" like dependency (Barnieh et al., 2021). In 2005, two groups described the synthetic lethal interaction between PARP inhibition and *BRCA1* or *BRCA2* mutations, suggesting a novel strategy for treating patients with *BRCA*-mutant tumours (Bryant et al., 2005; Farmer et al., 2005). Preclinical data suggested

that BRCA-mutant tumours were 1000x more sensitive to PARPi than *BRCA*-WT cells (Farmer et al., 2005). Two PARPi, olaparib and rucaparib, were approved for BRCA-deficient ovarian cancer and BRCA-deficient prostate cancer in 2020 (National Cancer Institute Website, 2022) and clinical trials are underway for the use of PARPi in BRCA-deficient breast and pancreatic cancers (National Cancer Institute Website, 2021). The successful development of PARPi for the treatment of BRCA-deficient cancers provided proof of concept that synthetic lethal interactions in the DDR can be exploited for cancer therapy (**Fig. 1.3**). Since, various other DDR proteins including ATR, ATM, DNA-PK, CHK1 and CHK2 have become promising targets for anti-cancer drug development with the discovery of synthetic lethal interactions between these proteins and DDR-deficiencies (Barnieh et al., 2021).

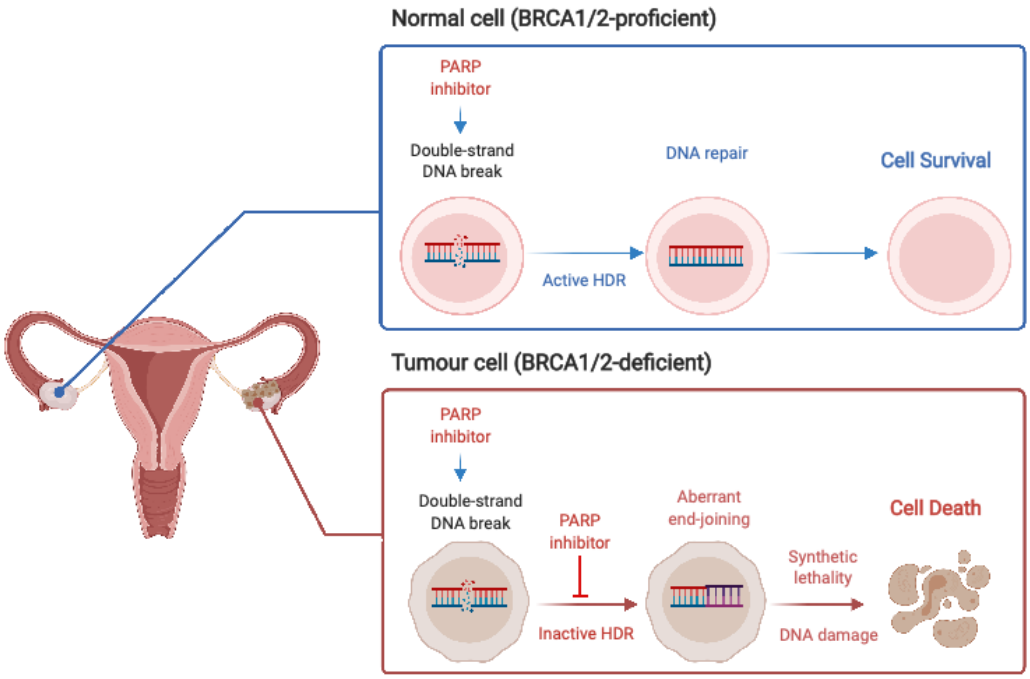


Fig. 1.3. Exploitation of DDR pathways to enhance therapeutic responses.
PARP inhibitors (PARPi) prevent the release of PARP1 from DNA, trapping PARP on DNA (Lord and Ashworth, 2017). Trapped PARP/DNA nucleoproteins complexes impair the progression of replication forks, resulting in replication fork stalling and DSB formation. HDR is the optimal DNA repair pathway for repairing and restarting replication forks stalled by PARPi. In normal (BRCA1/2-proficient) cells (top), replication-associated DSBs are repaired by HDR resulting in cellular survival. In cancerous (BRCA1/2-deficient) cells (bottom), in the absence of HDR, cells use aberrant repair processes that can potentially result in large-scale genomic rearrangements, which often leads to cancer cell death (synthetic lethality). Created with BioRender.com.

1.2 The Cellular Response to DNA DSBs

1.2.1 ATM Initiates and Sustains Chromatin-Based DDR Signalling

DNA DSBs are highly toxic lesions that pose a major threat to genome stability as their misrepair can lead to chromosomal deletions, chromosomal rearrangements, cell death and cancer (Scully and Xie, 2013). As mentioned in the previous section, defects in DSB repair are associated with neurodegeneration, infertility, immune deficiency, aging and cancer predisposition (Ciccia and Elledge, 2010; Jackson and Bartek, 2009). Interphase cells respond to DNA DSBs by initiating a chromatin-based signalling cascade, rapidly recruiting various proteins to damaged chromatin (Blackford and Jackson, 2017). Some proteins engage in DNA repair, whilst others trigger the DNA-damage checkpoint, delaying cell cycle progression.

DSBs induce an extensive response in neighbouring chromatin, which can be visualised as IR-induced nuclear foci using immunofluorescence microscopy (Scully and Xie, 2013). Generally, DSB signalling is initiated by the MRE11-RAD50-NBS1 (MRN) complex, which senses and binds DSBs (**Fig 1.4**; Kanaar and Wyman, 2008). NBS1 contains a conserved C-terminal motif that directly interacts with ATM, inducing ATM activation (Falck et al., 2005; Lee and Paull, 2004; Uziel et al., 2003). ATM phosphorylates the histone H2A variant, H2AX, on serine 139 of its C-terminus, generating γH2AX (Rogakou et al., 1998). ATM is primarily responsible for this modification, but in some contexts ATR or DNA-PK can substitute (Stiff et al., 2004). γH2AX is induced in a rapid and sensitive manner (Rogakou et al., 1998). Following induction of DSBs using IR, half maximal amounts of γH2AX are reached within a minute and maximal amounts are reached within 10 minutes (Rogakou et al., 1998).

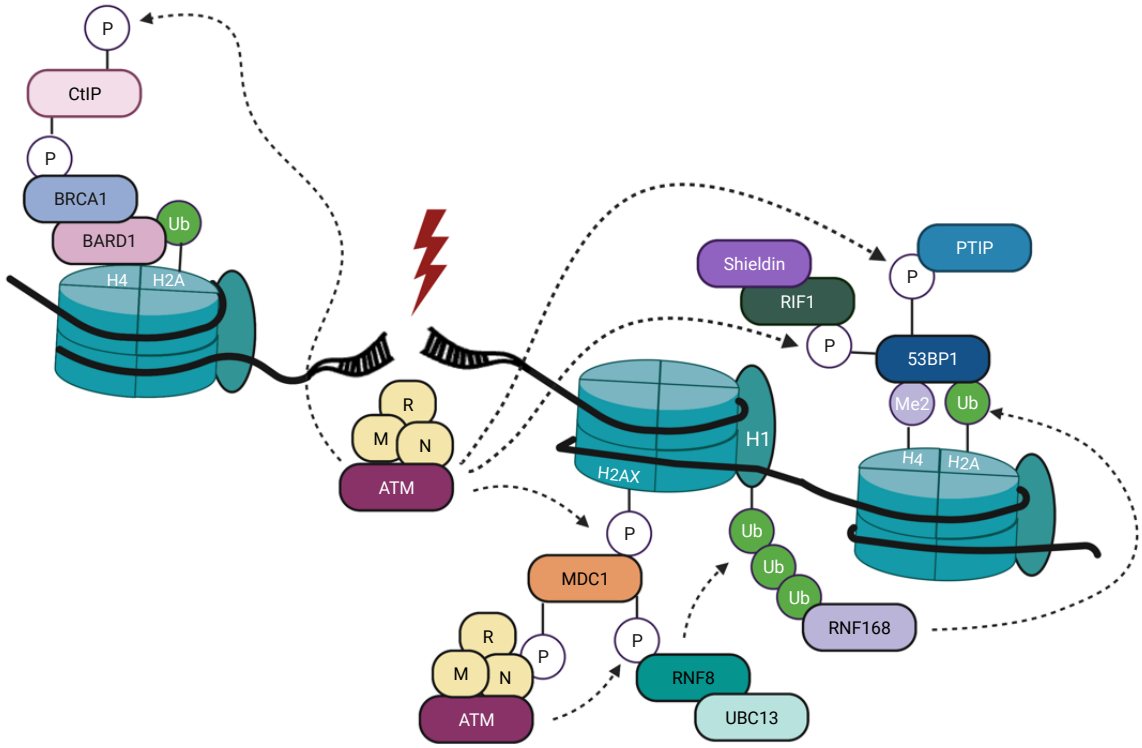


Fig. 1.4. ATM initiates and sustains chromatin-based DDR signalling.
ATM is recruited and activated by the MRN complex at DSBs. ATM phosphorylates H2AX and MDC1 to activate a ubiquitylation signalling cascade mediated by RNF8 and RNF168 that results in 53BP1 or BRCA1 recruitment. ATM phosphorylates 53BP1 to promote recruitment of its effectors, PTIP and RIF1. P, phosphorylation; Me, methylation; Ub, ubiquitylation. Adapted from (Sedelnikova:2002ck; Blackford and Jackson, 2017). Created with BioRender.com.

γH2AX induces a chromatin-based signalling cascade involving phosphorylation, ubiquitylation and other PTMs (Scully and Xie, 2013; Blackford and Jackson, 2017). The major reader of γH2AX is MDC1, via its C-terminal BRCT domains (Stucki et al., 2005). MDC1 is a large scaffold protein and recruits various DDR factors. MDC1 contains a N-terminal forkhead associated (FHA) domain, phosphorylated on T4 in an ATM-dependent manner to promote MDC1 dimerization and MDC1 retention on damaged chromatin (Jungmichel et al., 2012; Liu et al., 2012). MDC1 also contains a cluster of conserved SDTD repeats, which are

constitutively phosphorylated by casein kinase 2 (CK2) and are recognised by the FHA-BRCT domains of NBS1 (Chapman and Jackson, 2008; Melander et al., 2008; Spycher et al., 2008). This promotes the retention of MDC1-MRN on damaged chromatin. MDC1-mediated recruitment of MRN results further ATM recruitment, via a interaction between NBS1 and ATM, resulting in additional γH2AX formation and MDC1-MRN recruitment, amplifying the DDR signal (Sedelnikova et al., 2002).

Supporting the fundamental nature of the γH2AX-MDC1 interaction in DSB signalling and DNA repair, *H2afx* (H2AX) and *Mdc1* (MDC1) knockout mice display growth retardation, male infertility, immunodeficiency and chromosome instability (Celeste et al., 2002; Lou et al., 2006). However, recent work has shown the PST-repeats of MDC1 can promote H2AX-independent chromatin association of MDC1 (Salguero et al., 2019). The PST repeats are dispensable for cellular survival in response to IR in H2AX-proficent, but not H2AX-deficient backgrounds. The H2AX variant is not homogeneously distributed in the genome, with large areas being largely or completely depleted of H2AX. It has been suggested that the recruitment of MDC1 via its PST repeats facilitates the DDR in these regions, whilst the γH2AX-MDC1 cascade facilitates the DDR in regions of the genome containing the H2AX variant.

Following MDC1 recruitment, the TQXF motifs of MDC1 are phosphorylated by ATM and recognised by the FHA domain of the E3 ubiquitin ligase, RNF8 (Huen et al., 2007; Kolas et al., 2007; Mailand et al., 2007). RNF8, together with the E2 ubiquitin-conjugating enzyme, UBC13, stimulates K63-linked ubiquitylation of the linker histone H1 (Thorslund et al., 2015). Ubiquitylated H1 is recognised by the UDM1 domain of a second E3 ubiquitin ligase RNF168. RNF168 ubiquitylates K15 of H2A-type histones to form

H2AK15ub (Mattiroli et al., 2012). This histone mark promotes the recruitment of 53BP1 and BRCA1, which promote NHEJ and HDR respectively. The mechanistic basis of their recruitment and their function in controlling DSB repair pathway choice can be found in sections 1.2.6 and 1.2.7.

1.2.2 Pathways of DSB Break Repair

DSBs are repaired by pathways based on NHEJ or HDR. NHEJ re-ligates DNA ends with or without end processing (Hustedt and Durocher, 2017). As for HDR, there are various sub-pathways all initiated by DNA end resection of the 5' strand of the DSB and the generation of ssDNA. Homologous recombination (HR) is the canonical HDR pathway, using the sister chromatid as a template to mediate faithful repair. As such, HR is restricted to S and G2, after DNA replication has occurred. HR is critical for the repair of one-ended DSBs, such as those produced by broken replication forks and so HDR proteins have also been implicated in the restart of stalled replication forks and ICL repair. A non-canonical pathway of HDR exists, called single-strand annealing (SSA). SSA does not use the sister chromatid to mediate repair and is mutagenic. Lastly, alt-EJ pathways are initiated by and dependent on DNA end resection by MRN/CtIP (Truong et al., 2013). These pathways are mostly used in cells with deficiencies in components of the core NHEJ or HR pathways. Microhomology-mediated end joining (MMEJ) is the best-defined alternative end joining (alt-EJ pathway) and will be described in greater detail below.

1.2.3 DSB Repair by Pathways Based on NHEJ

NHEJ is the predominant DSB repair pathway in mammalian cells, except when DSB repair occurs in close proximity to DNA replication forks, when HDR is used (Zhao et al., 2020). As mentioned previously, NHEJ is required for V(D)J recombination and CSR.

In NHEJ, the Ku heterodimer is the first factor to bind to the DNA ends and recruits other NHEJ factors (Blackford and Jackson, 2017; **Fig. 1.5**). Ku has a high affinity for DSB ends (K_d = 6 x 10^{-10} M; Mimori and Hardin., 1986), is highly abundant (400,000 molecules per cell) and protects DSB ends from extensive resection (Chang et al., 2017). Ku70/80 recruits and activates the DNA-PK catalytic subunit (DNA-PKcs), forming the DNA-PK holoenzyme (Carter et al., 1990; Lees-Miller et al., 1990). Additional NHEJ core factors are then recruited, including XRCC4 (Li et al., 1995), DNA ligase IV (LIG4; Grawunder et al., 1997) and XLF (Ahnesorg et al., 2006; Buck et al., 2006). The NHEJ machinery is stabilized on damaged chromatin by PAXX, a second XRCC4-like NHEJ factor (Ochi et al., 2015). Some DSBs have two incompatible DNA ends, with chemical modifications or mis-matched overhangs that inhibit direct ligation and so are dependent on end processing (Chang et al., 2017). These DSB ends require accessory factors for repair, such as DNA polymerases and nucleases (Blackford and Jackson, 2017). The endonuclease Artemis is one such end processing factor, promoting the repair of a subset of DSBs in concert with DNA-PK (Ma et al., 2002). Of all IR-induced DSBs, 20-50% have been estimated to require end processing by Artemis (Chang et al., 2017). Additional end processing factors include polynucleotide kinase 3' phosphatase (PNKP; Koch et al., 2004), APTX (Clements et al., 2004; Gueven et al., 2004) and APLF (Bekker-Jensen et al., 2007; Iles et al., 2007; Macrae et al., 2008).

Single-molecule imaging has demonstrated the DNA-PK holoenzyme tethers DNA ends in a long-range synaptic complex (Graham et al., 2016). In this state, the aligned ends are not readily ligated by XRCC4-LIG4, but may provide sufficient space for end processing without allowing the ends to diffuse apart. Then DNA ends are then closely aligned, forming the short-range synaptic complex, which requires XRCC4-LIG4, XLF and DNA-PK activity (Graham et al., 2016). Promoting the formation of this complex, a

single XLF dimer bridges two XRCC4-LIG4 complexes (Graham et al., 2018). The close alignment of the ends positions them to be ligated.

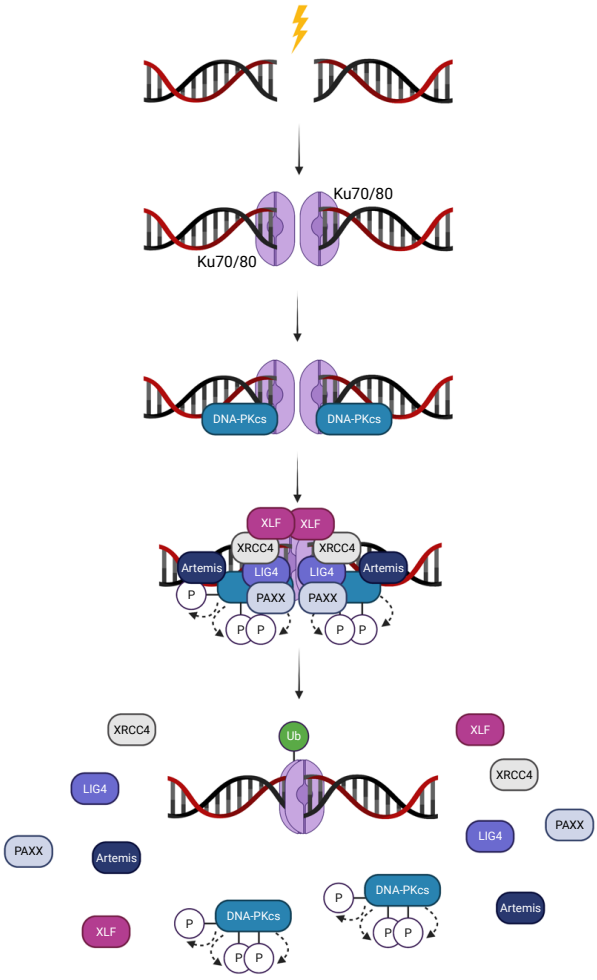


Fig. 1.5. Repair of DSBs by NHEJ.
The Ku70/80 heterodimer binds to the DSB and recruits DNA-PKcs to form the DNA-PK holoenzyme, which tethers DNA ends in a long-range synaptic complex. Additional NHEJ factors are then recruited, including XRCC4, DNA ligase IV, XLF and PAXX, which form the short-range synaptic complex. The close alignment of the ends in the short-range complex positions them to be ligated. Some DSBs are dependent on end processing prior to ligation, so require additional accessory factors, such as the endonuclease Artemis. P, phosphorylation; Ub, ubiquitylation. Adapted from Blackford and Jackson, 2017. Created with BioRender.com.

Once DNA repair has occurred, Ku70/80 needs to be extracted by structural remodelling (Postow et al., 2008). Ku80 is modified by K48-linked ubiquitin chains,

by one of three E3 ligases: SCF^{Fbxl12}, RNF8 and RNF138 (Feng and Chen, 2012; Ismail et al., 2015; Postow and Funabiki, 2013). Then, the hexameric AAA+-ATPase, p97, targets Ku80 modified by K48-linked ubiquitin chains for proteasomal degradation, extracting Ku70/80 from DNA (Boom et al., 2016).

1.2.4 DSB Repair by Pathways Based on HDR

HR is the most conservative HDR pathway, being largely error-free (Jasin and Rothstein, 2013). To avoid loss of heterozygosity, HR is restricted to the S and G2 phases of the cell cycle, where the sister chromatid is used as the template for repair (Blackford and Stucki, 2020). HR can be divided into four steps: resection of the DSB end; the presynaptic stage, involving RAD51 nucleoprotein filament assembly, the synaptic stage, characterised by strand invasion and branch migration and a postsynaptic stage, involving Holliday-junction formation and resolution (Moynahan and Jasin, 2010; **Fig. 1.6**). HR mediates the exchange of genetic information between sister or homologous chromatids, facilitating accurate repair of DSB breaks. A non-canonical pathway of HDR also exists, called single-strand annealing (SSA). SSA is dependent on extensive resection of the DSB, homology-dependent annealing of the resected ssDNA and direct ligation (Bhargava et al., 2016). This HDR pathway is donor independent and as such has the potential to induce genomic rearrangements.

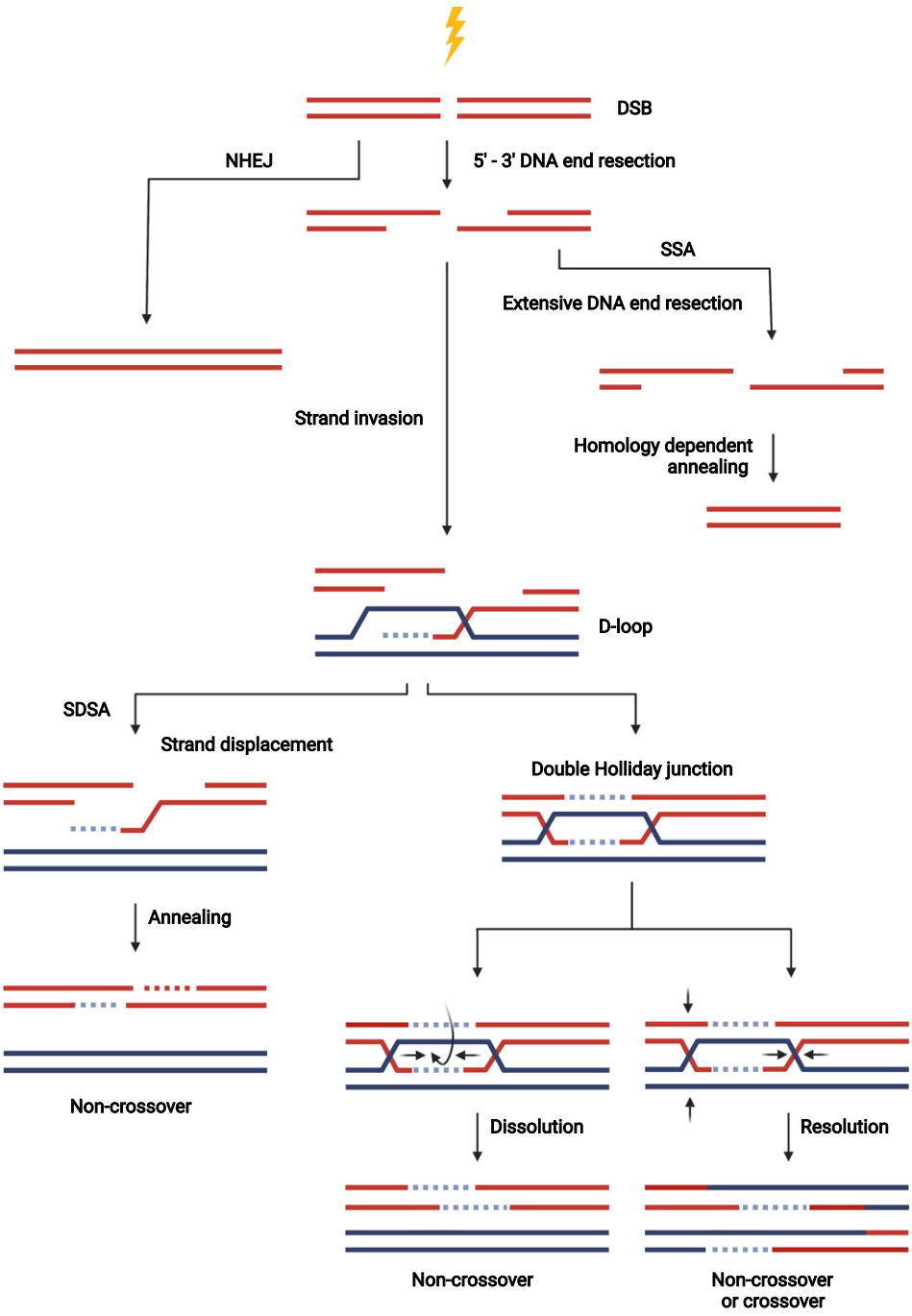


Fig. 1.6. Repair of DSBs by HDR
HDR initiates with DNA end resection, which produces a 3′ single-stranded end that can invade a homologous template to initiate repair. Alternative HR pathways can ensue from the displacement loop (D-loop) intermediate: synthesis-dependent strand annealing (SDSA) and double Holliday junction processing. In SDSA, the newly synthesized strand is displaced to anneal to the other DNA end, resulting in a non-crossover outcome with no change to the template DNA. In double Holliday junction processing, the second DNA end is captured to form a double Holliday junction, which in can result in a non-crossover or a crossover outcome. NHEJ can occur in the absence of DNA end resection. Single-strand annealing (SSA) can take place when extensive DNA end resection occurs, followed by homology-dependent annealing of the resected ssDNA and direct ligation. Adapted from Moynahan and Jasin, 2010. Created with BioRender.com.

DNA end resection occurs via two sequential mechanisms, short-range and long-range resection (**Fig. 1.7**; Cejka and Symington, 2021). The key nuclease responsible for short-range DNA end resection is MRE11. In addition to MRN, DNA end resection is promoted by CtIP (Sartori et al., 2007). MRE11 *in vitro* is a manganese-dependent exonuclease that degrades DNA with a 3'-5' polarity (Paull and Gellert, 1998). This would lead to 5' terminated DNA, disagreeing with the canonical model of DSB resection and direct observations in various organisms, which demonstrate the 5'-terminated DNA is degraded (Sun et al., 1991; White and Haber, 1990; Zhu et al., 2008). This apparent paradox can be explained by the action of co-factors of MRE11, such as CtIP, NBS1 and ATP hydrolysis by RAD50, which turn MRE11 into an endonuclease that preferentially nicks the 5' DNA end (Anand et al., 2016; Cannavo and Cejka, 2014; Deshpande et al., 2016). This led to the proposal of a bidirectional model of DNA end resection (**Fig. 1.7)**, where the endonucleolytic activity of MRE11 generates a nick adjacent to the DSB end, initiating both 3'-5' degradation of the nicked strand towards the break (catalysed by MRE11 exonuclease activity) and 5'-3' degradation by EXO1 or BLM/DNA2 (Anand et al., 2016; Cannavo and Cejka, 2014; Garcia et al., 2011). There is evidence in yeast to suggest that Mre11 and Sae2 generate multiple nicks in the resecting strand, which may enhance resection efficiency (Jazayeri et al., 2008; Zakharyevich et al., 2010). Some studies have found that CtIP possesses intrinsic endonuclease activity and could generate the initial nick (Makharashvili et al., 2014; Wang et al., 2014), but other studies have disputed this (Anand et al., 2016).

MRE11 nuclease activity has been implicated in Ku70/80 removal from DNA ends, generating ssDNA overhangs refractory to NHEJ (Langerak et al., 2011). It could be suggested that one of function of MRE11-mediated short-range resection is to counteract NHEJ and promote HDR (Chapman et al., 2012).

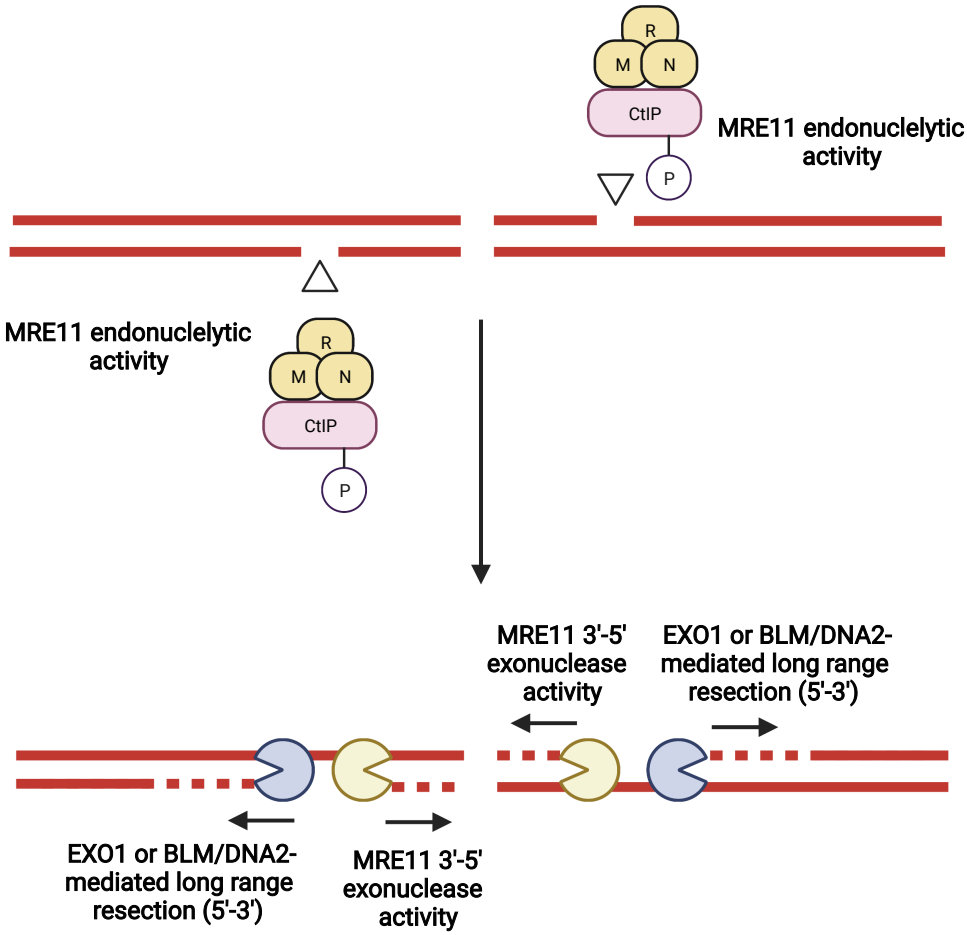


Fig. 1.7. Bidirectional model of DNA end resection
The endonucleolytic activity of MRE11 generates a nick adjacent to the DSB end, serving as initiating points for both 3'-5' degradation of the nicked strand towards the break (catalysed by MRE11 exonuclease activity) and 5'-3' degradation by EXO1 or BLM/DNA2. Adapted from Cejka and Symington, 2021. Created with BioRender.com.

BRCA1 was proposed to regulate DNA end resection via a direct interaction with CtIP, mediated by the CDK-dependent phosphorylation of CtIP on S327 (Yu and Chen, 2004), however various reports have disputed this (Escribano-Díaz et al., 2013; Nakamura et al., 2010; Ochs et al., 2016; Reczek et al., 2013; Schmidt et al., 2015).

Extensive, long-range 5'-3' resection, is mediated by one of two parallel pathways: via the exonuclease EXO1 or the combined helicase/nuclease activities of BLM/DNA2 (Nimonkar et al., 2011). BLM functions within cells with topoisomerase III alpha (TOP3A)

and RecQ-mediated genome instability proteins 1 and 2 (RMI1 and RMI2, respectively), forming the BLM dissolvasome (BTR) complex (Bizard and Hickson, 2014). The helicase activity of BLM is stimulated by TOP3A, RMI1 and RMI2 to promote DNA resection by DNA2 (Kowalczykowski, 2015).

The ssDNA generated by resection is rapidly bound by RPA, protecting the ssDNA from nucleolytic degradation and inhibiting the formation of secondary structures that can interfere with repair (Jasin and Rothstein, 2013). RPA has an extremely high affinity for ssDNA, with a K_d of 10^{-10} M (Kim et al., 1994). RPA is heterotrimer consisting of three subunits, RPA70, RPA32 and RPA14 (Bhat and Cortez, 2018). Between them, the three RPA subunits have six oligonucleotide binding (OB) fold domains. Four OB domains (70A, 70B, 70C, 32D) bind to ssDNA. Two OB domains (32D, 70C) and RPA14 assemble the RPA heterotrimer, whilst 70N functions as a protein-protein interaction domain. RPA32 has a winged-helix C-terminal domain that mediates protein-protein interactions.

RPA binding to ssDNA is regulated in a mechanochemical fashion (Chen et al., 2015). DNA polymerases are able to generate forces of up to 30 pN (Wuite et al., 2000), so ssDNA produced during DNA replication is likely subjected to force. This mediates a switch of RPA bound ssDNA from non-productive amorphous aggregation to a regular extended conformation (Chen et al., 2015).

RPA coating the ssDNA is displaced, generating a nucleoprotein filament (the pre-synaptic complex) containing RAD51 that is essential for ATP-dependent homologous base pairing and strand exchange (Kowalczykowski, 2015). Each monomer of RAD51 interacts with three nucleotides of ssDNA, forming a right-handed helical RAD51-ssDNA filament comprising of 6 RAD51 monomers per turn (Stasiak et al., 1994). RAD51 is a DNA-dependent ATPase, and is loaded onto DNA in its ATP-bound form

(Kowalczykowski, 2015). ATP hydrolysis permits conversion to the inactive ADP-bound form, which is less stable and promotes disassembly.

The pre-synaptic complex searches for DNA sequence homology and mediates the pairing and exchange of DNA strands (Kowalczykowski, 2015). Although capable of forming filaments on naked ssDNA *in vitro*, RAD51 has a lower affinity for ssDNA than RPA (K_d of 10^{-6} M), so requires mediator proteins to displace RPA (Bhat and Cortez, 2018). In mammalian cells, the primary mediator of the formation of the pre-synaptic complex is BRCA2 (Jensen et al., 2010; Liu et al., 2010; Thorslund et al., 2010). BRCA2 contains eight conserved sequence motifs (the BRC repeats), which interact with at least six monomers of RAD51 (Wong et al., 1997). BRCA1 recruits BRCA2 to DSBs via PALB2 (Prakash et al., 2015). The coiled-coil domain of BRCA1 recognizes the coiled-coil domain of PALB2, which in turn binds BRCA2 via its WD40 domain. In addition, RNF168 can act redundantly with BRCA1 to load PALB2 directly onto damaged chromatin (Zong et al., 2019). This may compensate for BRCA1 loss in BRCA1-deficient cells.

BRCA2 interacts with ssDNA via C-terminal OB folds, as well as a second ssDNA binding region in the N-terminus (Yang et al., 2002). BRCA2 has been proposed to promote RAD51 filament formation by using its ssDNA-binding activity to target RAD51 to ssDNA, enabling RAD51 to displace RPA from ssDNA (Kowalczykowski, 2015). The BRC repeats mimic a motif in RAD51 that promotes oligomerization between RAD51 monomers, enabling BRCA2 to promote RAD51 nucleoprotein filament assembly on ssDNA (Pellegrini et al., 2002). Furthermore, BRC repeats 1-4 inhibit ATP hydrolysis of RAD51, stabilizing RAD51-ssDNA filaments by blocking ATP hydrolysis (Jensen et al., 2010; Liu et al., 2010; Thorslund et al., 2010). Accordingly, BRCA2 functions to initiate assembly as well as promoting the stability of the presynaptic complex.

In addition to BRCA2, other RAD51 mediators exist in mammalian cells (Prakash et al., 2015). RAD51 paralog complexes interact with RAD51 and may promote/stabilise RAD51 filaments. The RAD51 paralogs are XRCC2, XRCC3, RAD51B, RAD51C, RAD51D (Takata et al., 2001) They show 20%-30% amino acid similarity to RAD51 and possess ATP-binding/hydrolysis activity (Prakash et al., 2015). The phenotypes of cells deficient in XRCC2, XRCC3, RAD51B, RAD51C, RAD51D are partially suppressed by RAD51 overpression, supporting the hypothesis that these paralogs mediate formation of RAD51 filaments (Takata et al., 2001). At least four distinct heterodimeric and heterotetrameric complexes exist: RAD51B-RAD51C, RAD51D-XRCC2, RAD51C-XRCC3, and RAD51B-RAD51C-RAD51D-XRCC2 (Kowalczykowski, 2015). Furthermore, SWS1 and its interacting partner SWSAP1 can also interact with RAD51 and all RAD51 paralogs except XRCC2, leading to a multitude of possible co-complexes (Liu et al., 2011).

The next step of HR is the synaptic stage, characterised by strand invasion and branch migration. The RAD51 nucleoprotein filament performs a homology search and strand invasion into the homologous sequence to form the synaptic complex (Baumann et al., 1996). In addition to ssDNA binding activity, RAD51 also possesses dsDNA binding activity, which can promote lateral contact with dsDNA (Wright et al., 2018). RAD51-dsDNA binding must be tightly regulated, as RAD51 binding to undamaged chromatin can deplete RAD51 cellular pools, interfering with RAD51-dependent recombination (Mason et al., 2015; Solinger et al., 2002). To inhibit this, in human cells, RAD54L and RAD54B disrupt RAD51-dsDNA binding in an ATPase-dependent fashion (Mason et al., 2015).

One strand of the duplex DNA is destabilised to open the DNA helix, allowing the bases within the duplex DNA to be sampled for complementarity to those within the filament (base flipping; Wright et al., 2018). Additional co-factors have been proposed to stimulate RAD51-mediated strand invasion by bridging RAD51 nucleoprotein filaments and the duplex DNA, including RAD51AP1, and PALB2 (Dray et al., 2010; Modesti et al., 2007; Wiese et al., 2007). For an efficient homology search, dsDNA must be transiently bound and rapidly turned over if the sequence is not complementary to the nucleoprotein filament (Wright et al., 2018). Single-molecule imaging demonstrated that RAD51 uses a length-based recognition mechanism to identify regions with a high probability of being a homologous target (Qi et al., 2015) Microhomologies of as few as eight nucleotides promote extended lifetimes of the RAD51-ssDNA-dsDNA complexes. Upon identification of homology in the donor dsDNA, a synaptic complex forms, consisting of the presynaptic filament and the homologous duplex DNA (Wright et al., 2018).

After synaptic complex formation, the 3' end of the invading stand must pair with its complementary site in the donor, forming a primer-template junction competent for DNA synthesis (Wright et al., 2018). This key intermediate is known as the displacement or D-loop. One strand of the double helix is displaced by the filament, forming a single stranded loop. The D-loop contains a Holliday junction and the 3' end of the D-loop can be used to prime DNA synthesis. DNA repair synthesis uses the undamaged homologous sequence and is dependent on DNA polymerase δ in mammalian cells (Sebesta et al., 2013; Sneeden et al., 2013). DNA repair synthesis extends the D-loop.

There are two pathways for the resolution of the extended D-loop, the double Holliday junction (dHJ) pathway and synthesis-dependent stand annealing (SDSA). SDSA is the preferred HDR pathway in mitotic cells (Bzymek et al., 2010). In dHJ pathway, the second 3' end pairs with the displaced strand of the extended D-loop, providing a template for

further repair synthesis (Wright et al., 2018). Second end capture could occur via one of two mechanisms: a second independent invasion or annealing to the displaced strand of the joint molecule, both of which result in dHJ formation (Nimonkar and Kowalczykowski, 2009). dHJs must be removed prior to chromosome segregation, so two mechanisms of dHJ processing exist in mammalian cells.

The primary mechanism for dHJ processing involves the BTR complex, which catalyses dHJ dissolution (Bizard and Hickson, 2014). dHJ dissolution is a two-step process, starting with the BTR complex promoting convergent dHJ branch migration, generating a hemicatenane intermediate (**Fig. 1.8**). Then TOP3A, with RMI1 and RMI2, decatenates this intermediate, generating exclusively non crossover (NCO) products (Singh et al., 2008; Wu and Hickson, 2003; Wu et al., 2006). Two types of HR repair products exist, crossovers (CO) and NCOs, depending on whether the arms of sister chromatids (or homologs) are exchanged or not respectively (Kowalczykowski, 2015). The generation of CO or NCO products is an important distinction as crossovers at non-allelic sites can induce translocations and crossovers between homologs can lead to loss of heterozygosity (Moynahan and Jasin, 1997), both of which are associated with cancer progression (Ali et al., 1987; Benedict et al., 1983; Cavenee et al., 1983; Lasko et al., 1991). Accordingly, in somatic cells, NCOs are the preferred outcome, avoiding potential loss of heterozygosity and chromosome rearrangements.

An alternative mechanism of dHJ processing exists, mediated by structure-specific endonucleases that catalyse dHJ resolution (**Fig. 1.8**). Three structure-specific endonucleases exist in human cells, SLX1-SLX4, MUS81-EME1 and GEN1 (Fekairi et al., 2009; Ip et al., 2008; Taylor and McGowan, 2008; Wyatt et al., 2013). These complexes define two pathways of HJ resolution, one mediated by GEN1, whereas SLX1-SLX4 and MUS81-EME1 mediate a second and genetically distinct pathway (Wyatt et al., 2013). Dependent

on the orientation of dHJ cleavage, dHJ resolution results in a mixture of CO and NCO products (Wechsler et al., 2011). As dHJ resolution can result in COs (unlike dHJ dissolution), it is under strict temporal control and resolvase activities are restrained until late in the cell cycle (Hustedt and Durocher, 2017). CDK1-dependent phosphorylation of SLX4 and EME1 at the onset of mitosis mediates the interaction of SLX1-SLX4 and MUS81-EME1 (Wyatt et al., 2013). GEN1 is regulated by nuclear exclusion, driven by a nuclear export signal, restricting GEN1 activity to mitosis after nuclear membrane breakdown (Chan and West, 2014). This suggests that HJ resolvases provide a safeguard mechanism, ensuring undissolved dHJs are removed prior to chromosome segregation (Hustedt and Durocher, 2017). This is confirmed by the observation that in BLM-deficient cells, HJ resolvases become the predominant pathway of dHJ resolution (Wechsler et al., 2011). The characteristic feature of BLM-deficient cells is a ten-fold increase in sister chromatid exchanges (SCEs), which serves as a diagnostic test for Bloom syndrome (Flanagan and Cunniff, 2019). The resultant increase in CO events between sister chromatids in BLM-deficient cells is proposed to explain the increase in SCE events.

Alternatively to second end capture, the extended D-loop can be disrupted after the DNA synthesis from the first processed DNA end and this displaced end can anneal to the other resected DSB end, followed by gap-filling, DNA synthesis and ligation (Nassif et al., 1994; Nimonkar and Kowalczykowski, 2009). This process is called synthesis-dependent stand annealing (SDSA). In human cells, various proteins including BLM, RECQL1/5, WRN, RTEL1, SMARCAL1, ZRANB3 and FANCM have all been implicated in D-loop dissolution, promoting SDSA (Heyer, 2015; Kowalczykowski, 2015). Unlike the dHJ pathway, SDSA always results in NCO outcomes and so is the preferred HDR pathway in somatic cells (Bzymek et al., 2010).

A non-canonical pathway of HDR of DSBs also exists, called single-strand annealing (SSA). SSA is dependent on extensive resection of the DSB and if homologous sequences are revealed, annealing of the resected ssDNA ensues, mediated by RAD52 (Bennardo et al., 2008). This is followed by endonucleolytic cleavage of the non-homologous 3'ssDNA tails, mediated by ERCC1/XPF and ligation (Motycka et al., 2004). It has been suggested 66-69% of the human genome consists of repetitive DNA elements, such as long interspersed elements and short interspersed elements (Koning et al., 2011). Extensive DSB end resection can reveal divergent repeats and accordingly SSA can result in a deletion between these distant repeats, resulting in loss of the intervening sequences (Bhargava et al., 2016).

In mammalian cells, SSA is promoted by RAD52 and DNA end resection factors, such as CtIP, BLM, EXO1 and DNA2 (Mendez-Dorantes et al., 2018). Consistent with SSA serving as a backup to other canonical, faithful DSB repair pathways, SSA is suppressed by KU70 and RAD51, which promote NHEJ and HR respectively (Mendez-Dorantes et al., 2018). Furthermore, when divergent repeats are annealed, the dsDNA duplex often contains mismatched bases, forming a heteroduplex. This intermediate is prone to reversal by heteroduplex rejection, mediated by various factors, including the mismatch repair factor MSH2 and BLM helicase (Mendez-Dorantes et al., 2018; 2020). These factors have been found to substantially suppress rearrangements between sequences that diverge by 1-3%.

Lastly, alt-EJ pathways for the repair of DSBs exist (Zhao et al., 2020). These pathways are mostly used in cells with NHEJ/HDR-deficiencies but in certain contexts alt-EJ operates when NHEJ and HDR are proficient (Brambati et al., 2020). For example, alt-EJ is essential for DSB repair in *C. elegans* germ cells, despite being NHEJ/HDR-proficient (van Schendel et al., 2015). Microhomology-mediated end joining (MMEJ) is the best-defined alt-EJ pathway and like HDR is initiated by and dependent on DNA end resection by MRN/CtIP (Truong et al., 2013). MMEJ requires 2-20 bp of

microhomology (Chang et al., 2017) and PARylation by PARP1 promotes the synapsis of resected ends (Audebert et al., 2004). Annealing of resected DSB ends is driven by internal microhomology, leading to the formation of 3' ssDNA flaps processed by FEN1 prior to POLθ-mediated fill-in synthesis (Mengwasser et al., 2019). This results in small deletions (of around 7-13 bp) at the DSB locus (Schendel et al., 2015). POLθ-mediated fill-in synthesis fills in the flanking ssDNA regions, preventing long-range resection that would promote HDR and generates a more stable intermediate that can be sealed by either DNA ligase I or DNA ligase III (Liang et al., 2008; Brambati et al., 2020). Although NHEJ is often described as error prone and mutagenic, it is efficient and mostly accurate (Bétermier et al., 2014). Instead, it is alt-EJ pathways, such as MMEJ, that are proposed to introduce extensive mutations (Schendel et al., 2015).

1.2.5 The Cellular Functions of the BTR Dissolvasome Complex

Loss-of-function mutations in the RecQ helicase BLM result in a rare inherited chromosomal disorder called Bloom syndrome (Ellis et al., 1995). As described, BLM functions in cells with TOP3A, RMI1 and RMI2. Accordingly, mutations in TOP3A, RMI1 and RMI2 can cause a Bloom syndrome-like disorder (Hudson et al., 2016; Martin et al., 2018).

The best characterised function of the BTR dissolvasome is its role in dHJ dissolution (**Fig. 1.7**). BLM-catalyzed dHJ dissolution has been proposed to suppress tumorigenesis by preventing loss-of-heterozygosity via the suppression of crossing over between homologous chromosomes during mitotic recombination by promoting dHJ dissolution (Groden et al., 1990; Luo et al., 2000). However, recent data revealed that loss-of-heterozygosity events in BLM-deficient cells is rare, suggesting that other cellular functions

of BLM may contribute to the increased cancer predisposition in Bloom syndrome patients (Wietmarschen et al., 2018), such as DNA end resection, the resolution of ultrafine DNA bridges between sister chromatids in anaphase and promoting stalled replication fork restart (Chan et al., 2007; Davies et al., 2007; Gravel et al., 2008).

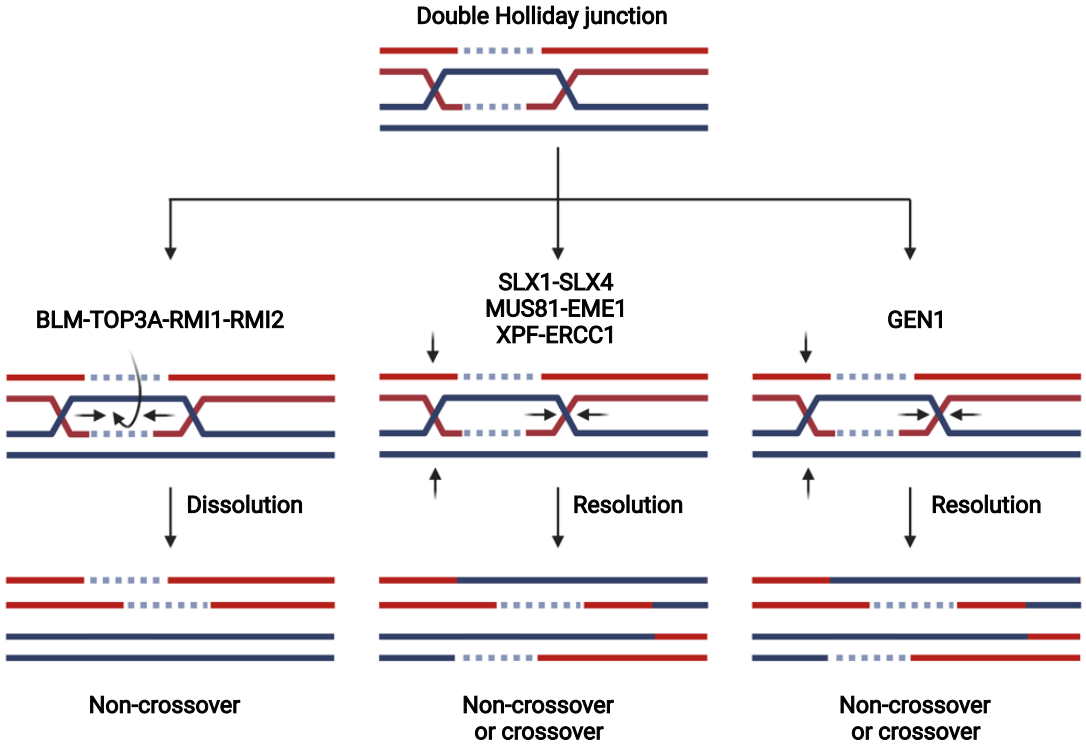


Fig. 1.8. Mechanisms of double Holliday junction (dHJ) dissolution and resolution.
Created with BioRender.com.

1.2.6 Establishment of DSB Repair Pathway Choice

The decision on how to repair a DSB involves whether to utilise NHEJ or one of the pathways dependent on DNA end resection (Morales et al., 2018). These pathways include HR, alt-EJ and SSA. DSB pathway choice is tightly regulated at the level of the cell cycle, which controls DNA end resection progression. CDK-dependent phosphorylation of

T847 in CtIP is critical for effective DNA end resection (Huertas and Jackson, 2009). Cohesin, which holds sister chromatids together after replication, promotes RAD51 nucleofilament invasion of the sister chromatid (Sjogren and Nasmyth, 2001). This counteracts ectopic recombination, which could lead to genomic rearrangements. Defects in the regulation of DSB pathway choice can induce genome instability and cancer (Blackford and Stucki, 2020).

Primary mediators of DSB repair pathway choice are BRCA1 and 53BP1, with the antagonism between these two proteins forming the basis of DSB pathway choice (Hustedt and Durocher, 2017). This antagonism between 53BP1 and BRCA1 is demonstrated by mouse models, where embryonic lethality, tumour predisposition and HDR defects present in *Brca1*-deficient mice is rescued by deleting 53BP1 (Bunting et al., 2010).

For many years, the chromatin-associated factor 53BP1 has been known to be a reader of H2AK15ub, channelling DSB repair towards NHEJ and opposing HDR (Fradet-Turcotte et al., 2013). RNF168-dependent chromatin modifications had been shown to orchestrate BRCA1 accumulation to DSBs, promoting HDR (Doil et al., 2009; Stewart et al., 2009) but how RNF168 promoted BRCA1 recruitment and function remained unclear. Recent work has shown that BARD1, the binding partner of BRCA1, also recognises H2AK15ub via its BRCT-domain-associated, ubiquitin-dependent recruitment motif (BUDR), recruiting BRCA1 to DSBs in a manner dependent on DNA damage (Becker et al., 2021; Hu et al., 2021). In addition to H2AK15ub, 53BP1 and BRCA1 complexes also bind to histone H4 via recognition of distinct K20-methylation states that oscillate during the cell cycle (Botuyan et al., 2006; Nakamura et al., 2019). In G1, nucleosomes are predominantly di-methylated on H4K20 (H4K20me2; Saredi et al., 2016). In S phase, unmethylated histone H4 (H4K20me0) is incorporated into newly synthesised DNA and mixed 1:1 with parental, methylated nucleosomes. H4K20me0 persists until G2/M

when the methyltransferases, SET domain-containing protein 8 (SET8) and suppressor of variegation 4-20 homologue (SUV4-20h1/2) catalyse mono-methylation (H4K20me1) and demethylation of H4K20 (H4K20me2) respectively (Beck et al., 2012). Accordingly, H4K20me2 and H4K20me0 mark the pre- and post-replicative state of chromatin. 53BP1 recognises H4K20me2 present on old histones throughout the cell cycle via its tandem tudor domains (Botuyan et al., 2006). 53BP1 accumulation at DSBs is reduced in post-replicative chromatin, correlating with the replication-dependent dilution of H4K20me2 (Pellegrino et al., 2017; Simonetta et al., 2018). Conversely, BARD1 recognises H4K20me0 present on new histones in post-replicative chromatin, routing only DSBs within sister chromatids towards the HDR pathway (Nakamura et al., 2019). Together, these two histone PTM states – regulated by the cell cycle (H4K20me0 vs. H4K20me2) and the presence of DNA damage (H2AK15 vs. H2AK15ub) – specify bivalent interactions with the domains of 53BP1 and BARD1, regulating the equilibrium between opposing DSB repair pathways (Becker et al., 2021). In summary, the shared affinity of 53BP1 and BARD1 for H2AK15ub-modified nucleosomes, but inverse affinities for H4K20me explains the respective presence of these proteins for pre- and post-replicated regions of the genome and the establishment of DSB repair pathway choice.

1.2.7 DNA End Protection by 53BP1

DNA end resection is the key determinant of DSB repair pathway choice and is tightly regulated (Chapman et al., 2012). In G1, DSB ends are “protected” by a 53BP1-dependent pathway, limiting DNA end resection, favouring repair by NHEJ. Further regulation of DNA end resection occurs via CDK-dependent phosphorylation of factors promoting resection. 53BP1-dependent DSB end protection requires various interacting partners, demonstrating it is not sufficient to prevent DNA end processing

(Setiaputra and Durocher, 2019). 53BP1 is phosphorylated in an ATM-dependent manner on 28 conserved S/TQ sites in its N-terminus (Jowsey et al., 2007), which is crucial for recruitment of binding partners. PTIP, associates with the N-terminus of 53BP1, in a manner dependent on ATM-dependent phosphorylation of 53BP1 on S25 and PTIP BRCT domains (Munoz et al., 2007). A second factor, RIF1, interacts with 53BP1 independently of PTIP but also dependent on ATM-mediated phosphorylation of the N-terminus of 53BP1 (Chapman et al., 2013; Escribano-Díaz et al., 2013; Virgilio et al., 2013; Zimmermann et al., 2013). Shieldin, a four-subunit protein complex with single-stranded DNA binding activity, was identified as an effector of 53BP1-dependent end protection downstream of RIF1 (Setiaputra and Durocher, 2019). Shieldin consists of REV7/MAD2L2, already identified as a 53BP1-pathway component (Boersma et al., 2015; Xu et al., 2015) and three previously uncharacterised proteins: C20orf196 (SHLD1), FAM35A (SHLD2) and CTC-534A2.2 (SHLD3; Dev et al., 2018; Findlay et al., 2018; Gao et al., 2018; Ghezraoui et al., 2018; Gupta et al., 2018; Noordermeer et al., 2018; Tomida et al., 2018).

53BP1 and RIF1 also stabilizes the three-dimensional nuclear architecture at sites of DSBs independently of Shieldin (Ochs et al., 2019). Super-resolution imaging demonstrated 53BP1 accumulates at regions of compact chromatin co-localising with topologically associating domain sequences, preceding RIF1 recruitment to boundaries between these domains. Depletion of either 53BP1 or RIF1 leads to decompaction of DSB-flanking chromatin, reduction in interchromatin space, aberrant spreading of DNA repair proteins (including BRCA1) and hyper-resection. Similar topological distortions were triggered by depletion of cohesin. In this mechanism, RIF1 functions to stabilize the three-dimensional nuclear architecture independent of its role in recruiting Shieldin to promote DNA end protection, as SHLD2 depletion did not lead to decompaction of DSB-flanking chromatin. This is consistent with Shieldin complex

components being phylogenetically younger than upstream components of the pathway (Gupta et al., 2018), suggesting that the 53BP1-RIF1 interaction could have primarily evolved to stabilize the three-dimensional nuclear architecture post DSB formation.

1.2.8 Responding to DSBs in Mitosis

When cells encounter DSBs in interphase, the cell cycle can be halted to promote DNA repair before the cell enters mitosis (Blackford and Stucki, 2020). However, when DSBs occur in mitosis, cells no longer arrest, and DSB repair is largely attenuated. In mitotic cells, a "primary" DDR is activated in response to mitotic DSBs, including ATM activation, γH2AX generation and MDC1 recruitment (Giunta et al., 2010). The signalling cascade downstream of MDC1 does not form in mitotic cells, as RNF8, RNF168, 53BP1 and BRCA1 are not recruited in mitotic cells (Giunta et al., 2010; Nakamura et al., 2014; Nelson et al., 2009; Vugt et al., 2010). CDK1-mediated inhibition of the RNF8-MDC1 interaction inhibits RNF8 recruitment to DSBs (Orthwein et al., 2014). Secondly, 53BP1 is phosphorylated on T1609 and S1618 by mitotic kinases (PLK1 and CDK1), inhibiting its interaction with ubiquitylated H2A and recruitment to DSBs (Lee et al., 2014; Orthwein et al., 2015). These residues are not dephosphorylated until G1 entry, by protein phosphatase 4 (PP4C; Lee et al., 2014). Without 53BP1 recruitment, hyper-resection could theoretically occur in mitosis. However, this is not observed (Lee et al., 2014; Peterson et al., 2011), possibly because chromosomes are too highly condensed for long-range resection to occur. Furthermore, key HDR factors such as BRCA1 are not recruited to mitotic DSBs (Giunta et al., 2010). It is unclear why NHEJ does not occur

in mitotic cells (Blackford and Stucki, 2020). It was suggested that DSB repair is universally inhibited during mitosis (Rieder and Cole, 1998; Zirkle and Bloom, 1953), however the absence of NHEJ in mitosis has never been experimentally proven.

In the absence of efficient DSB repair in mitosis, broken chromosomes represent a major threat for mitotic cells (Blackford and Stucki, 2020). If DSBs are not tethered during mitosis, the chromosome breaks into two pieces: a chromatid fragment containing a centromere (centric fragment) and a chromatid fragment lacking a centromere (acentric fragment). The centric fragment can attach to the mitotic spindle and segregate into a daughter cell, but its free DNA end could fuse to another centric fragment or deprotected telomere from another chromosome in the subsequent G1, resulting in a dicentric chromosome (Bailey et al., 2004; Latre et al., 2003). Dicentric chromosomes are sources of breakage-fusion-breakage cycles (McClintock, 1941), a major source of chromosome instability in cancer (Hanahan and Weinberg, 2011). Acentric fragments cannot interact with the mitotic spindle and are unable to segregate without compensatory mechanisms and can be converted into micronuclei (Countryman and Heddle, 1976). If the micronucleus is reincorporated into the nucleus in the subsequent cell cycle, it may undergo chromothripsis (Kato and Sandberg, 1968; Stephens et al., 2011). Chromothripsis is present in 2-3% of all cancers and 25% of bone cancers (Stephens et al., 2011). One mechanism avoiding these outcomes would be the tethering broken chromosome ends until they can be repaired in the following G1 phase and DDR re-establishment (Royou et al., 2010). This has been described in *Drosophila melanogaster,* but it is unknown whether an equivalent mechanism exists in human cells.

1.3 DNA Replication and the Cellular Response to Replication Stress

1.3.1 Eukaryotic DNA Replication

DNA replication is a universal process, and its accurate execution is essential for the propagation of all forms of life (Wyman and Botchan, 1995). It is strongly conserved throughout evolution, although the complexity of DNA replication machineries and their regulation increases with genome complexity. Genome duplication is carried out by a multi-protein complex called the replisome, which is assembled at origins of replication initiation.

DNA replication initiation in eukaryotic cells occurs via a two-step mechanism (Deegan and Diffley, 2016). Firstly, an inactive form of the replicative helicase (the pre-replication complex) is assembled at origins of replication in late mitosis or G_1 (Bochman et al., 2008; Costa et al., 2011; Davey et al., 2003; Li et al., 2015; Remus et al., 2009). This is called origin licensing. The core of the replicative helicase in eukaryotes is a hetero-hexameric ring-shaped AAA+ ATPase called MCM, in which six related subunits (MCM2-7) interact. The loading of MCM onto DNA requires three additional proteins: the hexameric AAA+ ATPase Origin Recognition Complex (ORC), CDC6 and CDT1 (Evrin et al., 2009; Gambus et al., 2011; Gillespie et al., 2001; Remus et al., 2009). The second step, helicase activation, occurs upon progression into S-phase, when the combined activities of CDK (Heller et al., 2011; Muramatsu et al., 2010; Tanaka et al., 2007; Yeeles et al., 2015; Zegerman and Diffley, 2007) and Dbf4-dependent kinase (DDK; (Heller et al., 2011; Tanaka et al., 2011; Yabuuchi et al., 2006; Yeeles et al., 2015) remodel the inactive MCM double hexamer into two active CMG (CDC45-MCM-GINS) helicases. Completion of origin firing results in the melting of the origin DNA, MCM ring

opening, the separation of the double hexamer and bidirectional fork progression (Yeeles et al., 2015).

The MCM2-7 hexamer and CDC45 form the active replicative DNA helicase, unwinding the parental DNA by translocating along the leading strand in an ATP-dependent fashion with a 3'-5' directionality (Burgers and Kunkel, 2016). The ssDNA generated is coated by RPA. The eukaryotic replisome contains three DNA polymerases, the polymerase α-primase (Byrnes et al., 1973) and the replicative polymerases ε (Kunkel et al., 1987) and δ (Byrnes et al., 1976) The replisome contains the clamp loader replication factor C (RFC), which recruits the sliding clamp, PCNA (Podust et al., 1995; Tsurimoto and Stillman, 1989). PCNA confers processivity to DNA polymerases δ (Podust and Hübscher, 1993; Waga and Stillman, 1994) and ε (Lee et al., 1991). Due to the antiparallel structure of DNA and the 3' end extension specificity of the DNA polymerases, DNA replication proceeds semi-continuously. On the leading strand, RFC loads PCNA and DNA polymerase ε and the leading strand is replicated continuously in the direction of fork progression (Miyabe et al., 2011; Pursell et al., 2007). In contrast, the lagging strand is replicated discontinuously via short Okazaki fragments (Okazaki et al., 1967, 1968; Sugimoto et al., 1968). The lagging strand is primed by DNA polymerase α, which synthesises a short RNA primer that is extended by DNA polymerase α to produce ~20-30 nucleotides of DNA (Nethanel et al., 1988). RFC displaces DNA polymerase α, initiating the switch from the priming to elongation mode (Tsurimoto and Stillman, 1990, 1991). DNA polymerase δ extends the fragment, generating an Okazaki fragment (McElhinny et al., 2008) . When the 3' end of one Okazaki fragment reaches the 5' end of the downstream fragment, DNA polymerase δ performs strand displacement synthesis, generating a short 5' flap. The endonuclease FEN1 recognises and cleaves the displaced flap, with DNA ligase I sealing the resulting nick

between the two Okazaki fragments (short flap pathway; Goulian et al., 1990; Harrington and Lieber, 1995; Ishimi et al., 1988; Turchi et al., 1994). In some instances, Pif1 elongates the flap (Rossi et al., 2008). These flaps are coated by RPA, which inhibits FEN1 endonuclease activity (Bae et al., 2001). Dna2 displaces RPA and cleaves the flap, leaving a terminal product ~5-6 nucleotides in length (Bae et al., 2003, 2001). This fragment can be processed by FEN1 (long flap pathway). PARP1 senses unligated Okasaki fragments, and facilitates their repair by recruiting the SSBR protein XRCC1, forming a PARP1 and XRCC1-dependent "backup" pathway (Hanzlikova et al., 2018).

When converging replisomes meet, DNA synthesis is completed, nascent strands are ligated, replisomes are disassembled, and daughter molecules are decatenated ("replication termination"; Dewar and Walter, 2017). Compared with replication initiation and elongation, replication termination is poorly understood. Replisome disassembly during replication termination involves polyubiquitylation of the MCM7 subunit of the CMG helicase by $CRL2^{Lrr1}$, followed by the removal of CMG from the chromatin by the p97 ATPase (Maric et al., 2014; Moreno et al., 2014). CMG ubiquitylation and removal results in the unloading of various CMG interacting partners, dismantling the replisome (Dewar et al., 2017). During elongation, a DNA structure associated with elongating replisomes suppresses CMG ubiquitylation (Low et al., 2020). This suppression is relieved when CMGs converge, leading to CMG unloading. If the CMG fails to be ubiquitylated by $CRL2^{Lrr1}$ in interphase, it can be ubiquitylated and unloaded in mitosis by a back-up pathway that depends on the E3 ubiquitin ligase TRAIP (Deng et al., 2019).

1.3.2 Controlling DNA Replication: Once and Only Once

In order for genetic information to be faithfully propagated during cell division, parental genomes must be copied once and only once per generation (Bertolin et al., 2020;

Deegan and Diffley, 2016). Re-replication can generate multiple copies of a single genomic region within a single cell, promoting gene amplification and genome instability (Green et al., 2010). This phenomenon is observed in many human cancers. To avoid this, the precise replication of the genome is achieved through strict regulation of the replicative helicase (Deegan and Diffley, 2016). High APC/C activity and low S-CDK activity (in late M/G1) promote inactive replication complex assembly but inhibit origin firing. This inhibits premature origin activation, which could result in re-licensing and re-replication. Conversely, high S-CDK and low APC/C activity facilitates progression into S phase and origin firing but inhibits re-licensing and thereby prevents re-replication. Re-replication is detected by the S-phase checkpoint, inducing cell cycle arrest or cell death (Liu et al., 2007). Consistently, re-replication tends to be detected in cells with severe dysregulation of the licensing system or compromised checkpoint pathways, such as cancer cells (Tatsumi et al., 2006; Vaziri et al., 2003).

In contrast, the likelihood of under-replication is much higher (Bertolin et al., 2020). Even a small region of under-replicated DNA poses a challenge at mitosis because unreplicated DNA persisting into anaphase results in a bridge linking the sister chromatids (McClintock, 1941). These bridges either create a lagging chromosome pair (that fail to be properly incorporated into the daughter nuclei, creating aneuploidy) or will rupture, resulting in unstable, broken-ended DNA that can lead to genome rearrangements that are a hallmark of many human cancers (Gisselsson et al., 2000; Hanahan and Weinberg, 2011).

Cells have various strategies to restart or complete replication, which will be discussed in subsequent sections. If under-replicated DNA persists until mitosis, mitotic DNA synthesis (MiDAS) can occur (Bhowmick et al., 2016). MiDAS is RAD52 dependent, which is required for the recruitment of MUS81-EME1 endonuclease and

POLD3. If under-replicated DNA from the previous cell cycle is present in the following G1, it is sequestered into specific nuclear sub-compartments called 53BP1 nuclear bodies (53BP1-NBs; Lukas et al., 2011). 53BP1-NBs persist into late S phase, when 53BP1/RIF1/Shieldin channels the under-replicated DNA from the previous S phase into replication-coupled RAD52-mediated repair (Spies et al., 2019). In this way, cells are provided a "second chance" to complete the replication of under replicated DNA loci, halting generation of chromosomal instability.

1.3.3 Barriers to DNA Replication

Replication stress, defined as the slowing or stalling of the DNA replication fork, is a complex phenomenon with serious implications for genome stability, cell survival and human disease (Zeman and Cimprich, 2014). If unresolved, replication fork stalling and collapse will interfere with the proper progression of DNA replication, threating genome integrity.

Replication forks are vulnerable to stalling or collapse as they encounter obstacles on the DNA template, which can be unrepaired DNA damage, DNA-bound proteins or DNA secondary structures (Petermann and Helleday, 2010). For example, if nicks and gaps caused by DNA damage or intermediates of DNA repair are encountered by the replication machinery, they can be converted to DSBs (Zeman and Cimprich, 2014). Misincorporation of ribonucleotides (rNTPs) can cause replication stress (Dalgaard, 2012). DNA polymerases δ and ε are less stringent in discriminating deoxyribonucleotides (dNTPs) from rNTPs than base-pairing during DNA replication. rNTPs stall replicative polymerases and can be processed into non-ligatable nicks by topoisomerase I (McElhinny et al., 2010). DNA lesions such as interstrand crosslinks and DNA-protein crosslinks form physical barriers to replication fork progression (Kim and D'Andrea, 2012; Shoulkamy et al.,

2012). Chemical agents such as hydroxyurea (Skoog and Nordenskjold, 1971; Slater, 1973) and aphidicolin (Ikegami et al., 1978) can inhibit replication elongation and lead to replication fork stalling. Trinucleotide repeats can form secondary DNA structures (hairpins etc.) that can block replication fork progression and/or promote replication fork slippage resulting in the expansion or contraction of repeat sequences and subsequent gene dysfunction (McMurray, 2010). G-quadruplexes (G4) form in GC-rich DNA sequences and chemical stabilisation of these structures or loss of the helicases that resolve them (BLM, WRN, PIF1, FANCJ) can slow DNA replication, result in the formation of DNA DSBs, and cause deletions (Bochman et al., 2012; Teng et al., 2021). This could be due to aberrant processing of forks stalled by these structures, or inaccurate replication of a template in which these structures were not properly unfolded.

Fork stalling can occur anywhere in the genome, but certain DNA loci have a higher propensity for DNA replication failure (Bertolin et al., 2020). These DNA loci are characterised by intrinsically hard-to-replicate repetitive regions (such as centromeres, telomeres and fragile sites), replication in late S phase, or paucity of active and dormant origins. As replication and transcription complexes both operate on DNA, these two processes can interfere with each other (Helmrich et al., 2013). Early replicating fragile sites are found in highly transcribed regions replicated early in S-phase. Furthermore, the nascent transcript may inappropriately re-hybridise with the DNA behind the transcription complex, forming an R-loop, which may interfere with DNA replication and cause genome instability (Datta and Jinks-Robertson, 1995).

Lastly, replication requires various factors, that can slow replication fork speed and induce DNA replication stress when availability is limited (Zeman and Cimprich, 2014). These factors include nucleotides and replication machinery components, as well as

histones and histone chaperones that package replicated DNA. Constitutive activation of oncogenes such as HRAS, MYC and cyclin E can induce DNA replication stress by promoting increased origin firing, depleting nucleotide pools and/or increasing collisions with transcription complexes (Bartkova et al., 2005, 2006; Denko et al., 1994; Micco et al., 2006).

1.3.4 Checkpoint Signalling in Response to DNA Replication Stress

ATR is the apical DNA replication stress response kinase (Blackford and Jackson, 2017). ATR is recruited to extended tracts of RPA-coated ssDNA (Zou and Elledge, 2003b), which is generated by helicase-polymerase uncoupling at stalled replication forks (Byun et al., 2005b). ATR recruitment to RPA-coated ssDNA does not result in optimal activation of ATR, which is dependent on activator proteins and ssDNA/dsDNA junctions. The best characterised ATR activator is the DDR scaffold protein, TOPBP1 (Kumagai et al., 2006). TOPBP1 contains an ATR-activation domain, which stimulates ATR kinase activity via an interaction with ATRIP and the C-terminal PRD in ATR (Mordes et al., 2008). TOPBP1-mediated ATR activation is dependent on the TOPBP1-RAD9 interaction, a subunit of the RAD9-RAD1-HUS1 (9-1-1) clamp complex (Delacroix et al., 2007; Lee et al., 2007). The 9-1-1 complex is loaded onto ds/ssDNA junctions by the RAD17/RFC2-5 clamp loader (Ellison and Stillman, 2003; Zou et al., 2002, 2003). The TOPBP1-RAD9 interaction is vital for ATR activation, but not for TOPBP1 recruitment to stalled replication forks (Lee and Dunphy, 2010; Yan and Michael, 2009).

How TOPBP1 is recruited to stalled replication forks is not clear, but various other factors (such as BLM and 53BP1) are involved in TOPBP1 recruitment to DNA damage (Blackford et al., 2015; Cescutti et al., 2010). Recently, a second ATR-activator protein, ETAA1 was identified (Bass et al., 2016; Feng et al., 2016; Haahr et al., 2016). Like TOPBP1, ETAA1 contains an ATR-activation domain but unlike TOPBP1, ETAA1 is

directly recruited to RPA-coated ssDNA by RPA. TOPBP1 and ETAA1 are likely to be recruited to distinct aberrant DNA structures, stimulating ATR activity towards different substrates.

A key function of ATR is to phosphorylate and activate the protein kinase CHK1 (Guo et al., 2000; Hekmat-Nejad et al., 2000; Liu et al., 2000). When activated, CHK1 promotes the proteasomal degradation of CDC25, inhibiting cyclin A/E-CDK2 and cyclin B-CDK1 (Peng et al., 1997; Sanchez et al., 1997; Bartek et al., 2004). This slows cell cycle progression through S-phase and prevents mitotic entry, allowing more time to restart or repair stalled replication forks or, if the damage is too extensive, activating cell senescence or apoptosis.

ATR phosphorylates helicases such as SMARCAL1, which can remodel stalled replication forks into configurations that can be recognised and cleaved by SLX4/MUS81 (Couch et al., 2013). This prevents "excessive" fork reversal, which can lead to fork breakage. Finally, ATR regulates dNTP availability by promoting the upregulation of the ribonucleotide reductase (RNR) subunit RRM2 in response to DNA damage, at both the transcriptional and post-translational level (D'Angiolella et al., 2012). RNR catalyses the conversion of rNTPs to dNTPs, which are necessarily for DNA synthesis. During G2, RRM2 is degraded via $SCF^{cyclin\ F}$ to maintain balanced dNTP pools. After DNA damage, cyclin F is downregulated in an ATR-dependent manner to allow RRM2 accumulation.

ATR has also been shown to prevent replication catastrophe by preventing global RPA exhaustion (Toledo et al., 2013) Via CHK1, ATR suppresses the firing of dormant origins, limiting the number of stalled forks to those that were active at the onset of replication stress and ensures that RPA remains in excess. If ATR signalling fails, dormant origins fire, which can severely deplete nuclear RPA. When all nuclear RPA is

sequestered, every active replisome generates unprotected ssDNA that is rapidly converted to DSBs. RPA exhaustion may facilitate chromothripsis (Stephens et al., 2011; Toledo et al., 2013). Uncontrolled replication in oncogene-transformed cells is expected to reduce the nuclear RPA pool in these cells, explaining why ATR inhibitors selectively kill oncogene-transformed cells compared to normal cells (Toledo et al., 2011). Linked with this, there is a correlation between high levels of RPA and therapeutic resistance in cancer (Dahai et al., 2013; Givalos et al., 2007), suggesting RPA overexpression could be selected for during oncogenesis.

1.3.5 Mechanisms of Stalled Replication Fork Restart

In most cases replication stress only transiently pauses the replisome (Cortez, 2015). Most lagging-strand template DNA damage will not hinder replication fork progression as a new Okazaki fragment will bypass the lesion. Leading strand damage is more likely to stall the replisome and induce helicase-polymerase uncoupling (Byun et al., 2005b). A stalled replication fork is arrested but capable of resuming replication once the inhibition has been removed (Petermann and Helleday, 2010). In contrast, a collapsed fork has been inactivated via the dissociation of the replication machinery or DSB generation (Cobb et al., 2003, 2005; Lopes et al., 2001; Tercero et al., 2003). Failure to stabilise stalled replication forks induce fork collapse, as will persistent replication inhibition.

At least three pathways overcome leading strand lesions: translesion synthesis, PrimPol-dependent repriming, and replication fork reversal (Cortez, 2019). Replication fork reversal is the conversion of a typical replication fork (a three-way junction) into a four-way junction or “chicken foot” by the coordinated annealing of the two newly synthesised strands

and the re-annealing of the parental strands, forming a fourth "regressed" arm at the fork elongation point (Neelsen and Lopes, 2015). Fork reversal has been proposed to promote replication fork restart via various mechanisms (Bhat and Cortez, 2018). Firstly, post lesion repair, the reversed fork can be converted back into a standard replication fork and replication elongation may continue. Secondly, fork reversal could stabilize the stalled fork until a converging fork from a nearby origin passively replicates the region. Replication origins are present in large excess within eukaryotic cells and so activation of dormant origins provides an efficient mechanism to promote completion of DNA replication upon replication stress. Thirdly, as the nascent DNA contains homology to the rewound parental DNA, RAD51 bound to the reversed fork could promote strand invasion to generate a D-loop. This could restart DNA replication by regenerating a replication fork. Finally, the reversed fork can be cleaved by endonucleases, generating a DSB that can be repaired by homologous recombination-mediated repair and promote replication fork restart.

Replication fork reversal is thought to promote genome stability for least three reasons (Cortez, 2015). Firstly, it may function to limit ssDNA at the fork, promoting replication fork stabilisation. Secondly, by placing the template DNA lesion back into the context of duplex DNA, it may facilitate BER. Finally, fork reversal could promote template switching, using the newly synthesised DNA strand as an undamaged template.

A number of proteins, mostly belonging to the RECQ helicase or SWI/SNF protein families are able to reverse replication forks *in vitro* (Cortez, 2019). The SWI/SNF helicase-like proteins SMARCAL1 and ZRANB3 can re-anneal RPA-coated DNA strands (Yusufzai and Kadonaga, 2008; 2010) and reverse synthetic replication forks *in vitro* (Bétous

et al., 2012; Ciccia et al., 2012). A third SWI/SNF helicase-like protein, HLTF, can reverse synthetic replication forks *in vitro* (Blastyak et al., 2010). In addition, the RECQ helicases BLM and WRN can catalyze the regression of a model replication fork *in vitro* (Machwe et al., 2006). Lastly, the translocase FANCM possesses DNA junction-specific helicase activity and dsDNA translocase activity, both of which are required for replication fork reversal *in vitro* (Gari et al., 2008a; 2008b).

Many questions remain regarding fork reversal, such as why so many enzymes with proposed fork reversal activity exist and how they work together (Cortez, 2019). Although they can catalyse similar biochemical reactions, they do not act redundantly as individual knockdown of any of these proteins causes replication-associated DNA damage. Furthermore, their inactivation yields different phenotypic outcomes, which may be due to distinct substrate specificities and their regulation by accessory domains that bind DNA, RPA and ubiquitylated PCNA. For example, SMARCAL1 physically interacts with RPA2, which is necessary for the recruitment of SMARCAL1 to stalled replication forks (Ciccia et al., 2009). The SMARCAL1-RPA2 interaction regulates SMARCAL1 activity (Bétous et al., 2013). RPA can stimulate SMARCAL1-mediated fork reversal activity when bound to the leading strand, and inhibit iSMARCAL1 when bound to the lagging strand. RPA, bound to a leading strand gap mimicking a stalled fork, blocks ZRANB3-mediated fork reversal (Bétous et al., 2013). RAD51 is also required to generate reversed forks but it is not clear why (Zellweger et al., 2015). Several models have been proposed (Cortez, 2019). Firstly, RAD51 could bind to ssDNA on the template strand and catalyze an annealing reaction in cooperation with the fork-reversal enzymes. Alternatively, RAD51 could capture the

nascent DNA as it is formed by the reversal enzymes, shifting the equilibrium towards fork reversal.

As well as catalyzing fork regression, SMARCAL1 can promote fork restoration (Bétous et al., 2013; Cortez, 2015). Fork restoration regenerates a replication fork that can resume DNA synthesis. This reaction is also regulated by RPA, as RPA inhibits SMARCAL1-catalysed fork restoration when bound to a longer lagging nascent-strand and activates SMARCAL1-catalysed fork restoration if bound to a leading nascent-strand. Furthermore, the RECQ helicases, WRN (in concert with DNA2) and RECQL1 have been implicated in replication fork restoration (Berti et al., 2013; Thangavel et al., 2015).

1.3.6 Resolving Collapsed Replication Forks

Reversed forks can be deleterious if they persist, due to the four-way junction structure (chicken foot) structure with an exposed DNA end (Bhat and Cortez, 2018). The junction itself can cleaved by structure-specific endonucleases such as MUS81-EME2 (Hanada et al., 2007; Pepe and West, 2014). Cleavage produces an intact chromatid and a single ended, partially replicated chromatid (replication fork collapse). This end can be repaired via break-induced replication (BIR). Unlike IR-associated DSBs, replication-associated DSBs are typically one ended (Cortez, 2019). HDR is the preferred repair mechanism and EJ is inhibited to avoid genome rearrangements. Although HDR is frequently studied as a mechanism to repair two-ended breaks, the most frequent requirement for HDR repair proteins seems to be in the repair of one-ended DSBs formed at collapsed forks. This explains why PARP inhibitors that increase the frequency of fork breakage induce synthetic

lethality in HDR-deficient cancers (Bryant et al., 2005; Farmer et al., 2005). Although HDR may be beneficial for resuming replication, repair can also lead to chromosomal rearrangements and instability through inappropriate recombination and MMEJ events.

As with two-ended DSBs, replication-associated DSBs are resected such that the 3' single-stranded end can engage in recombination (Epum and Haber, 2021). There are at least two distinct BIR pathways in yeast, one that is Rad51-dependent and one that is Rad51-independent (Le et al., 1999; Lundblad and Blackburn, 1993). Unlike normal DNA replication, BIR leads to conservative replication in which all the newly synthesised DNA is inherited at the repaired end (Donnianni and Symington, 2013; Saini et al., 2013). The two strands of DNA are not simultaneously copied, instead there is a long first-strand synthesis (leading strand-like) that remains single-stranded until the second (lagging strand-like) strand is copied (Saini et al., 2013). First strand synthesis depends on DNA polymerase δ, which normally only performs Okasaki fragment synthesis (Liu et al., 2021). The ability of DNA polymerase δ to processively replicate extensive regions is dependent on several factors that are not required for normal DNA replication: the non-essential Pol32 subunit of the DNA polymerase δ complex and the 5' to 3' helicase, Pif1 (Lydeard et al., 2007; Wilson et al., 2013). Studies of BIR in mammalian cells confirm the conservation of the core recombination/replication complex, including the POLD3 subunit of DNA polymerase δ and PIF1 helicase (Costantino et al., 2014; Li et al., 2021).

BIR synthesis can be restricted by replication fork convergence with a nearby fork and the ability of nucleases to process the D-loop (Cortez, 2019). This maintains genome stability as BIR is error prone. DNA repair by BIR in human cells is accompanied by a >100-fold increased rate of mutagenesis in comparison to normal DNA replication (Deem et al., 2011). The persistence of the single-stranded first new strand of DNA makes it a target for cytosine deamination by enzymes, such as those of the APOBEC family, and chemical damage (Elango et al., 2019).

Due to frequent dissociation of DNA polymerase δ from the template, BIR exhibits strand slippage and improper template switching, generating chromosomal deletions, translocations and truncations (Anand et al., 2014). These jumps readily occur between disparate homologous sequences such a long terminal repeat regions, interstitial telomere-like sequences and telomeres.

In addition to replication fork collapse, the end of the reversed arm resembles a DSB and susceptible to nucleolytic degradation by nucleases such as MRE11 (Bhat and Cortez, 2018). MRE11 can remove end-binding proteins to promote fork restoration; however, excessive nucleolytic degradation of the reversed arm can lead to genome instability. RAD51 binding to the reversed arm of the chicken foot has been proposed to protect the reversed arm from excessive MRE11-mediated nascent strand degradation (Hashimoto et al., 2010; Schlacher et al., 2011). This process, called fork protection, is dependent on BRCA2-mediated stabilisation of RAD51-ssDNA filaments. A BRCA2 C-terminal region that binds and stabilises RAD51 filaments is required for fork protection but dispensable for HDR, demonstrating fork protection and HDR can be functionally separated. Many additional HR proteins, including BRCA1, PALB2 and the RAD51 paralogs, as well as various Fanconi anaemia proteins prevent replication fork degradation (Bhat and Cortez, 2018). However, the relevance of fork protection in promoting cellular viability and its impact on cancer development has been questioned. Cells with a mutation at the BRCA2-S3291 residue are severely disrupted in fork protection but do not have a survival defect in either unperturbed contexts or in the presence of PARPi/platinum treatment (Feng and Jasin, 2017).

1.3.7 Human Diseases Associated with Defects in the Replication Stress Response

There are several diseases associated with defects in the resolution of replication stress, demonstrating its importance in human physiology and preventing human disease (**Table 1.3**). Individuals or animals with a hypomorphic allele of *ATR*, containing a splicing mutation that results in very low ATR expression (O'Driscoll et al., 2003), or mutations in ATR's obligate binding partner *ATRIP* (Ogi et al., 2012), develop Seckel syndrome, which is characterised by dwarfism, microcephaly and mental retardation (McKusick et al., 1967; Seckel., 1960).

Deficiencies in proteins that recognise or repair lesions that cause replication stress cause a number of human diseases (Zeman and Cimprich, 2014). For example, defects in the recognition and repair of DNA inter-strand crosslinks causes Fanconi Anaemia (Kee and D'Andrea, 2012). Replication-associated expansions in repetitive DNA repeats underlie ~20 severe neuromuscular and neurodegenerative disorders, including Huntington's disease (Zeman and Cimprich, 2014).

Table 1.3. Human Genetic Diseases Associated with Defects in the Replication Stress Response (adapted from Zeman and Cimprich, 2014)			
Syndrome	**Mutated Gene**	**Affected pathway**	**Phenotype**
Aicardi-Goutieres syndrome	*RNase H2, TREX1, SAMHD1*	Removal of ribonucleotides, RNA-DNA hybrids	Neurological dysfunction
Amyotrophic lateral sclerosis 4	*Senataxin*	Resolution of RNA-DNA hybrids, transcription termination	Childhood- or adolescent-onset degeneration of motor control
Ataxia-occular apraxia 2	*Senataxin*	Resolution of RNA-DNA hybrids, transcription termination	Adolescent-onset cerebellar ataxia
Bloom syndrome	*BLM*	DNA remodelling, replication fork structure resolution, HR	Growth retardation, photosensitivity, infertility, growth defects, cafè-au-lait spots, predisposition to all cancer types
Fanconi Anaemia	*FANC* family of genes	ICL repair (FANC family of proteins), Replication fork remodelling (FANCM)	Bone marrow failure, AML, squamous cell carcinoma (head and neck), abnormal skin pigmentation, infertility, limb deformities
Huntington's disease	*HTT*	Trinucleotide repeat expansion	Neurodegeneration (ataxia, loss of coordination, mental abilities)
Rothmund-Thomson syndrome	*RECQL4*	DNA remodelling, replication fork structure resolution	Premature ageing, growth retardation, cancer predisposition
Schimke immune-osseous dysplasia	*SMARCAL1*	Replication fork stabilisation and reversal, DNA re-annealing	Dwarfism, skeletal abnormalities, renal failure, mental retardation
Seckel syndrome	*ATR, ATRIP, CENPJ, CEP152, PCNT*	ATR signalling	Growth retardation, dwarfism, microcephaly, mental retardation
Werner syndrome	*WRN*	DNA remodelling, replication fork structure resolution	Premature ageing, growth retardation, predisposition to sarcomas

Furthermore, mutations in various helicases implicated in the replication stress response are associated with disease. The RecQ helicases remodel and stabilize stalled replication forks and at least three family members, BLM, WRN and RECQL4, are associated with disease: Bloom syndrome, Werner syndrome and Rothmund-Thomson syndrome respectively (Ellis et al., 1995; Kitao et al., 1999; Yu et al., 1996). *RECQL4* mutations are also responsible for two other disorders, RAPADILINO syndrome and Baller-Gerold syndrome (Siitonen et al., 2003; Van Maldergem et al., 2006). Patients of all five syndromes have increased cancer predisposition, however RecQ helicases, especially BLM helicase, have various additional cellular functions and so it is difficult to discern which cellular function suppresses human disease (Chu and Hickson, 2009). Finally, the annealing helicase SMARCAL1 may reverse and help restart forks in an ATR-dependent manner (Couch et al., 2013). Loss of SMARCAL1 leads to Schimke immuno-osseous dysplasia, characterised by kidney and skeletal abnormalities, as well as immunodeficiency (Boerkoel et al., 2002; Schimke et al., 1974; Spranger et al., 1991). Together, these syndromes demonstrate the importance of resolving replication stress in preventing genome instability and human disease.

1.4 Short Linear peptide Motifs (SLiMs)

Protein-protein interactions play essential roles in many biological processes (Rual et al., 2005). Identification of protein-protein interactions is useful for the identification of multimeric protein complexes, identifying biological pathways, and predicting protein function. Historically, most research has investigated interactions between structured domains, once seen as the sole mediators of protein-protein interactions (Davey et al., 2012b). However, accumulating evidence has revealed a continuous spectrum of binding interfaces with varying degrees of flexibility, ranging from rigid globular domains to disordered regions that natively

lack structure. It has become clear that a large proportion of functional protein-protein interactions are mediated by compact interaction interfaces within intrinsically disordered regions called Short Linear peptide Motifs (SLiMs; Tompa et al., 2014).

1.4.1 Characterisation of SLiMs

SLiMs are recognised and/or post translationally modified by interacting partners, which may function to promote or inhibit an interaction with a binding partner (Tompa et al., 2014). The known instances of SLiMs mediate a wide range of important cellular tasks, including cellular signalling, providing specificity for post-translational modifications, targeting proteins to subcellular compartments, acting as recognition sites for proteolytic cleavage, membrane binding, protein folding and cell adhesion (Bhowmick et al., 2015; **Table 1.4**).

Table 1.4. Instances of SLiMs and their regulatory roles (adapted from Tompa et al., 2014)

Type	Description	Examples of Binding Partner	SLiMs	Function
Binding SLiMs (mediate interactions with globular domains)				
Classical binding SLiMs	Act as binding sites, often facilitating protein complex assembly	SH2 domain family members	pYXXϕ	Phospho-dependent SH2 domain binding SLiM
Trafficking SLiMs	Mediate binding to protein necessary for localisation to a specific subcellular region	Exportin-1	ϕX{2,3}ϕX{2,3}ϕXϕ	Nuclear export signal
Targeting SLiMs	Retain protein at a specific subcellular localisation by binding to an already localised protein	PCNA	QXXϕXX[FHT][FHY]	Replication fork localisation SLiM
Docking SLiMs	Recruit enzyme to protein and promote modification at a site distinct from the docking site	Cyclin family members	[RK]XLX{0,1}ϕ	CDK/cyclin kinase complex docking SLiM
Degron SLiMs	Control the stability of a protein by recruiting ubiquitin ligases, resulting in polyubiquitination and subsequent proteasomal degradation	Cdc20 and Cdh1 APC/C subunits	RXXLXXϕ	APC/C-specific degradation SLiM
Post-translational modification sites (recognised and altered by modifying enzymes)				
Moiety addition sites	Recognised by modifying enzymes for the attachment or removal of biochemical chemical functional groups	SUMO-conjugating enzyme UBC9	ϕ(K)X[DE]	Site SUMOylated by SUMO ligase
		PIKK	p[S/T]Q	Site phosphorylated by PIKKs
Cleavage sites	Recognised by proteases for proteolytic cleavage	Caspase-3	[DE]XXD|[GSAN]	Site of cleavage by caspase protease
Structural modification sites	Recognised by enzymes that catalyse the structural conversion of a region	Peptidyl-prolyl cis/trans-isomerase Pin1	p[ST](P)	Site isomerised by Pin1

Φ, bulky hydrophobic residues; X, any amino acid; p, phosphorylation; ( ), a site of modification; |, site of cleavage; { }, how many times the preceding residue can be used; [ ], use one of the residues within the brackets to create a SLiM

SLiMs are ubiquitous in eukaryotic proteomes. Based on the known instances of SLiMs, several defining traits of SLiM have been revealed, differentiating them from globular domains (Davey et al., 2011). Firstly, SLiMs have the capacity to encode a functional interaction interface within a short sequence (typically up to 30 amino acids). Binding motifs are on average 6-7 amino acids in length, with only 3-4 core positions conferring much of the interaction specificity (Tompa et al., 2014). The limited binding surface area results in low-affinity, transient interactions, generally within the 1-150 μM range (Diella et al., 2008). Secondly, whilst specific positions within the SLiM are evolutionarily conserved (marking the specificity and affinity determinants of the SLiM), a lower degree of sequence specificity is often observed within the SLiM flanks. Due to this, SLiMs exhibit high evolutionary plasticity and can arise de novo by a few mutations (Tompa et al., 2014). A subset of SLiMs involved in critical processes are conserved over large evolutionary distances. Thirdly, SLiMs tend to be enriched in intrinsically disordered regions of proteins, and possess the ability to function independently of their tertiary structural context. More than 30% of the human proteome is at least partially intrinsically disordered and current estimates suggest that these regions contain more than 100,000 SLiMs (Mészáros et al., 2009). The bias for disordered regions likely reflects the accessibility of SLiMs, allowing recognition of their binding partner and the flexibility of SLiMs, enabling them to adopt a specific conformation upon binding (Tompa et al., 2014).

1.4.2 Implications for Drug Design and Treatment Strategies

Given that SLiMs mediate protein-protein interactions that promote various cellular processes, including DNA replication, cellular differentiation and apoptosis, mutations in

regions of proteins containing SLiMs can contribute to disease, including cancer (Uyar et al., 2014). Oncogenic processes can be driven by deregulated expression of intrinsically disordered proteins containing SLiMs. For example, over-expression of the E3 ubiquitin ligase MDM2 suppresses the pro-apoptotic activity of p53, as binding of MDM2 to a SLiM in p53 (FxxxWxxL) promotes the proteasomal degradation of p53 (Haupt et al., 1997).

It has been suggested that the evolutionary plasticity of SLiMs can easily add functionality to proteins and rewire pathways, which is crucial for the evolution of eukaryotic regulation (Neduva and Russell, 2005). However, this has generated an Achilles' heel by facilitating mimicry of SLiMs by pathogens such as bacteria, fungi and viruses. They can utilise SLiMs to invade host organisms and generate environments that facilitate pathogen proliferation (Uyar et al., 2014). For example, SARS-CoV-2, the coronavirus responsible for the COVID-19 pandemic, uses SLiM mimicry on the spike protein to interact with host cell integrin receptors, facilitating viral entry into host cells (Mészáros et al., 2021). Deregulated expression of intrinsically disordered proteins containing SLiMs and pathogenic mimicry of the SLiMs of the host cell are just two examples of the therapeutic relevance of SLiMs. Due to their smaller interaction interface, interactions based on SLiMs are inherently druggable and molecular compounds have shown promise for targeted treatment strategies (Uyar et al., 2014). Harnessing this knowledge could aid the development of treatment strategies.

1.5 Aims and Objectives

Proteins consist of at least two discrete classes of functional modules: globular domains and SLiMs (Tompa et al., 2014). Although resources exist for predicting protein domains from sequence alone (e.g. Pfam; El-Gebali et al., 2018), it is more challenging to do this for SLiMs

due to their low sequence complexity. Furthermore, the protein-protein interactions they mediate often depend on other factors, such as post-translational modifications. As such, it is experimentally challenging to identify functional SLiMs and their study lags behind that of protein domains. Due to their relatively small size, SLiMs are swiftly gained and lost in evolution, but a subset are highly conserved throughout vertebrates, other metazoans and even down to yeast. Such conserved motifs are likely to be involved in critical processes.

There are a diverse range of DNA-lesions that can occur, and accordingly multiple, largely distinct DNA repair pathways are required to repair the damage and maintain genome stability. Whilst the majority of proteins directly involved within the DDR have likely been identified, the recent discovery the role of CIP2A in promoting genome integrity in mitotic cells (Adam et al., 2021a; Laine et al., 2021a) suggests that not all DNA-repair associated proteins have been found as-of-yet. We lack a comprehensive view of how proteins associated with the DDR interact with one another, how complexes are assembled in response to DNA damage and disassembled once repair is complete, and crucially, how these processes are regulated by post-translational modifications. Further understanding of the DDR at the molecular level will facilitate greater understanding of the diverse range of pathologies associated with mutations in DNA repair genes and the development of new and existing therapies.

Protein-protein interactions mediated by SLiMs play essential roles in many biological processes, including the DDR. The aim of this DPhil has been to use SLiM proteomics to identify novel DNA repair mechanisms. This required developing a combined bioinformatics and proteomics pipeline to reliably identify and characterize conserved SLiMs in the DDR. These methods have previously been used to identify and characterize novel protein-protein interactions for specific DNA repair proteins (Blackford et al., 2015; Ochi et al., 2015; Polo et al., 2012; Stucki et al., 2005). These examples provide

proof-of-principle that functional interactions can be identified in this way, as well as provide insight into regulatory mechanisms in the DDR. However, these methods have never been used to identify SLiMs within the DDR on a larger scale. I aimed to apply this bioinformatics and proteomics approach to identify and characterize SLiMs in DDR proteins, seeking to improve our limited understanding of how proteins in the DDR interact.

In Chapter 3, I will describe the development of a combined bioinformatics and proteomics approach to identify SLiMs in various different DSB proteins and their subsequent validation. In Chapter 4, I will describe recently published work that came from investigating three hits from this screen: three SLiMs in the Bloom helicase (BTR) complex and their requirement for replication fork restart (Shorrocks et al., 2021). Lastly, in Chapter 5, I will describe the development of a web-based tool capable of identifying SLiMs in any protein in an automated fashion.

Chapter 2: Materials & Methods

2.1 Sequence Alignments

Representative orthologue sequences for each protein of interest were acquired from NCBI. Sequences were aligned using T-Coffee (Notredame et al., 2000), accessible at (https://www.ebi.ac.uk/Tools/msa/tcoffee/), with conserved and similar residues highlighted using Boxshade. During the writing of this thesis, Boxshade was discontinued. The alternative tool Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html) is available.

2.2 Peptide Pulldowns

Biotinylated peptides (GenScript) were bound to streptavidin-coupled Dynabeads M-280 (Thermo Fisher Scientific). HeLa nuclear extracts (Ipracell) were diluted 1:1 with dilution buffer (200 mM NaCl, 10 mM NaF, 0.2 mM EDTA, 0.2% Igepal CA-630, 20 mM HEPES-KOH, pH 7.4), supplemented with cOmplete protease inhibitor cocktail (Roche), unless specified, and cleared by centrifugation. Dynabead-conjugated peptides were incubated with clarified extracts with end-to-end mixing at 4 °C for 2 hr, unless specified. Beads were washed five times with peptide pulldown buffer (100 mM NaCl, 50 mM KCl, 5 mM NaF, 0.2 mM EDTA, 0.2% Igepal CA-630, 10% glycerol, 20 mM HEPES-KOH, pH 7.4), supplemented with cOmplete protease inhibitor cocktail (Roche), before elution in 2X SDS sample buffer for mass spectrometry or SDS-PAGE. The sequences of the peptides used in this study are listed in table 2.1, as well as how they are dissolved and their concentration.

Table 2.1. Peptide sequences used in this study

Peptide	Residues	Amino Acid Sequence	C-terminus	Dissolved in	Concentration (mg/ml)
53BP1	301-313	Biotin-SGSG-EPEVLSTQEDLFD	NH2	DMSO	5
53BP1	301-313	Biotin-SGSG-EPEVL(pS)(pT)QEDLFD	NH2	DMSO	5
AUNIP	340-357	Biotin-SGSG-DLLFTQDSEGNQVIRHQF	-	DMSO	5
AUNIP	340-357	Biotin-SGSG-DLLF(pT)QDSEGNQVIRHQF	-	DMSO	5
AUNIP	340-357	Biotin-SGSG-DLLFTQD(pS)EGNQVIRHQF	-	DMSO	5
BARD1	140-160	Biotin-SGSG-KNSIKMWFSPRSKKVRYVVSK	NH2	WATER	5
BARD1	140-160	Biotin-SGSG-KNSIKMWF(pS)PRSKKVRYVVSK	NH2	WATER	5
BLM	137-163	Biotin-SGSG-KKLEFSSSPDSLSTINDWDDMDDFDTS	NH2	DMSO	5
BLM	290-309	Biotin-SGSG-EFDDDDYDTDFVPPSPEEII	NH2	DMSO	5
BLM	290-309	Biotin-SGSG-EFDDDDYDTDFVPP(pS)PEEII	NH2	WATER	1
BLM	552-568	Biotin-SGSG-DIDNFDIDDFDDDDDWE	NH2	DMSO	5
BLM	1392-1417	Biotin-SGSG-ANSKLGIMAPPKPINRPFLKPSYAFS	-	WATER	5
BRCA1	455-467	Biotin-SGSG-NIEDKIFGKTYRK	NH2	WATER	5
BRCA1	501-527	Biotin-SGSG-KLKRKRRPTSGLHPEDFIKKADLAVQK	NH2	WATER	5
BRCA2	1100-1132	Biotin-SGSG-NHNLTPSQKAEITELSTILEESGSQFEFTQFRK	NH2	WATER	1
BRCA2	2401-2414	Biotin-SGSG-RPTKVFVPPFKTKS	NH2	WATER	5
BRCA2	3340-3360	Biotin-SGSG-IADEELALINTQALLSGSTGE	NH2	DMSO	5
BRCA2	3340-3360	Biotin-SGSG-IADEELALIN(pT)QALLSGSTGE	NH2	WATER	1
CtIP	651-676	Biotin-SGSG-ENIQWSIDPGADLSQYKMDVTVIDTK	NH2	DMSO	5
CtIP	651-676	Biotin-SGSG-ENIQWSIDPGADL(pS)QYKMDVTVIDTK	NH2	WATER	2
ERCC6L2	796-812	Biotin-SGSG-PGQLTLLQCGFSKLLET	NH2	DMSO	5
FANCM	1126-1143	Biotin-SGSG-LSTDQDESLLLFEDVNTE	NH2	DMSO	5
FANCM	1126-1143	Biotin-SGSG-LSTDQDE(pS)LLLFEDVNTE	NH2	DMSO	5
MDC1	184-204	Biotin-GGSG-SRTTSSSVIVPESDEEGHSPV	NH2	WATER	5
MDC1	184-204	Biotin-GGSG-SRTTSSSVIVPE[pS]DEEGHSPV	NH2	WATER	5
POLQ	1317-1334	Biotin-SGSG-DFEDSFYLDTQSEKIIQQ	NH2	WATER	1
POLQ	1317-1334	Biotin-SGSG-DFEDSFYLD(pT)QSEKIIQQ	NH2	DMSO	2
POLQ	1434-1454	Biotin-SGSG-EVSVTDSQLNSFLQGYQTQET	NH2	DMSO	5
POLQ	1434-1454	Biotin-SGSG-EVSVTD(pS)QLNSFLQGYQTQET	NH2	WATER	5
POLQ	1477-1497	Biotin-SGSG-PVPETSLNMSDSLLFDSFSDD	NH2	DMSO	5
POLQ	1631-1652	Biotin-SGSG-WSGASFDLSPGLQRILDKVSSP	NH2	WATER	2
RAD17	632-651	Biotin-SGSG-LPLSQNSASELPASQPQPFS	NH2	WATER	5
RAD17	632-651	Biotin-SGSG-LPL(pS)QNSASELPA(pS)QPQPFS	NH2	WATER	5

RECQL4	68-87	Biotin-SGSG-AAEEAPEPRCWGPHLNRAAT	NH2	WATER	5
RECQL4	101-124	Biotin-SGSG-SVPDYGQRLKANLKGTLQAGPALGRR	NH2	WATER	5
RECQL4	171-197	Biotin-SGSG-RLQHLQASLSQRLGSLDPGWLQRCHSE	NH2	WATER	5
RMI1	240-260	Biotin-SGSG-LGPSDEELLASLDENDELTAN	NH2	DMSO	5
RMI1	308-329	Biotin-SGSG-DGELDDFSLEEALLLEETVQKE	NH2	WATER	1
RNF168	223-234	Biotin-SGSG-GDIQKYLTPKSQ	NH2	WATER	2
RNF168	223-234	Biotin-SGSG-GDIQKYL(pT)PKSQ	NH2	WATER	2
RNF168	272-284	Biotin-SGSG-EDMPTLSPQISLG	NH2	DMSO	5
RNF168	558-571	Biotin-SGSG-SQKSVFQMFQRCTK	-	WATER	2
SHLD2	265-283	Biotin-SGSG-ISTDTEFLSIITSSQVAFL	NH2	DMSO	5
SHLD2	333-345	Biotin-SGSG-QNQAYSLELFSPV	NH2	WATER	1
SLX4	603-628	Biotin-SGSG-SASQREHQALQDLVDLAREGLSASPW	NH2	WATER	2
TOPBP1	486-503	Biotin-SGSG-ADEDLLSQYENGSSTVVE	NH2	DMSO	2
TOPBP1	486-503	Biotin-SGSG-ADEDLL(pS)QYENGSSTVVE	NH2	DMSO	2
TOPBP1	774-798	Biotin-SGSG-RKTVVTPLDMNRFQSKAFRAVVSQH	NH2	WATER	5
TOPBP1	774-798	Biotin-SGSG-RKTVV(pT)PLDMNRFQSKAFRAVVSQH	NH2	WATER	5
XRCC4	255-278	Biotin-SGSG-DSIISSLDVTDIAPSRKRRQRMQR	NH2	WATER	5
XRCC4	318-334	Biotin-SGSG-SLETLRNSSPEDLFDEI	-	WATER	5

For pulldowns with UV-crosslinking, biotinylated peptides containing BPA (Genosphere Biotechnologies) were incubated with clarified HeLa nuclear extracts as above. Where indicated, samples were exposed to light from a Stratalinker UV Crosslinker 2400 (Stratagene) for 1 hr on ice, followed by washing with peptide pulldown buffer at 4 °C or denaturing wash buffer (137 mM NaCl, 3 mM KCl, 0.1% Tween 20, 1% SDS, 25 mM Tris-HCl, pH 7.5) at room temperature. Samples were eluted in 2× SDS sample buffer for SDS-PAGE. BPA-modified peptides were biotinylated on the N-terminus, with the following sequences: motif 1 (SGSG- KKLEFSSSPDSLSTIND-[BPA]-DDMDDFDTS), motif 3 (SGSG-DIDNFDIDD-[BPA]-DDDDDWE) and motif 5 (SGSG-LGPSDEELLAS-[BPA]-DENDELTAN).

For pulldowns with in vitro translated heterotrimeric RPA (incorporating either $RPA1^{WT}$ or $RPA1^{R41E/Y42F}$), RPA was first transcribed and translated using the TNT Quick

Coupled Transcription/Translation System (Promega) according to the manufacturer's instructions. In vitro translated complexes were diluted in modified (50 mM NaCl instead of 150 mM) peptide pulldown buffer supplemented with 1 mM $MgCl_2$, cOmplete protease inhibitor cocktail and 25 U/mL Benzonase (Novagen). After nuclease digestion, NaCl and EDTA concentrations were adjusted to 200 mM and 2 mM, respectively, and samples were cleared by centrifugation before being used for pulldowns with biotinylated peptides as above.

2.3 Sample Preparation for Mass Spectrometry

Prior to mass spectrometry, material in SDS sample buffer was reduced using dithiothreitol (DTT) to break disulphide bonds between cysteines and alkylated using iodoacetamide (IAA). IAA reacts with the free sulfhydryl groups of cysteine residues to form S-carboxyamidomethyl-cysteine, which cannot be reoxidised to form disulfide bonds (Gundry et al., 2009), ensuring the trypsin has maximum access to cleavage sites within the protein. The reduced and alkylated sample was subsequently subjected to two rounds of chloroform-methanol precipitation to remove the SDS from the sample buffer. SDS is required to elute the sample from the streptavidin-coupled Dynabeads, but SDS is not compatible with mass spectrometry and so its removal is required to achieve a higher peptide/protein coverage in mass-spectrometric analysis. The precipitated material was resuspended and digested by trypsin, before being desalted and analysed by nano-flow liquid chromatography tandem mass spectrometry (LC-MS/MS; LC-MS/MS was carried out by the Target Discovery Institute at the University of Oxford).

2.4 SDS-PAGE and Western Blotting

SDS-PAGE and western blotting were performed using 7% Tris-Bicine gels with the SE400 and TE42 systems from Hoefer, or Bolt 4-12% Bis-Tris Plus gels from Thermo Fisher Scientific. Biotinylated peptides were resolved using Novex 10-20% Tricine gels from Thermo Fisher Scientific. Gels of peptide pulldowns were stained with the 2D Silver Stain Kit II (Daiichi) or the Colloidal Blue Staining Kit (Thermo Fisher Scientific) following the manufacturer's instructions.

Membranes were blocked in 5% milk in TBST for 30 min, before incubation in primary antibody diluted in in 5% milk in TBST for 2 hr. The following primary antibodies were used at the indicated dilutions: 53BP1 (NB100-305, Novus Biologicals, 1/500), AUNIP (bs-15019R, Stratech Scientific, 1/500), Biotin (ab53494, Abcam, 1/100), BLM (A300-110A, Bethyl Laboratories, 1/2000), CIP2A (1/250, ab99518, Abcam), CtIP (1/500, ab155988, Abcam), FLAG (F1804, Sigma-Aldrich, 1/2000), GFP (11814460001, Roche, 1/5000), H2AX (NB100-383, Novus, 1/5000), γH2AX (05-636, Millipore, 1/1000), HA (sc-57592, Santa Cruz Biotechnology, 1/1000), MDC1 (ab11171, Abcam, 1/5000), MRE11 (PC388, Merck, 1/500), NBS1 (ab32074, Abcam, 1/2000), RMI1 (NB100-1720, Novus Biologicals, 1/1000), RMI2 (NBP1-89962, Novus Biologicals, 1/4000), RPA1 (NA13, Merck, 1/200), RPA2 (ab10359, Abcam, 1/10,000), RPA3 (HPA005708, Sigma-Aldrich, 1/1000), SPOT (1/2000, etb-250, ChromoTek), TOP3A (14525-1-1AP, Proteintech, 1/1000), TOPBP1 (A300-111A, Bethyl Laboratories, 1/2000), XPA (ab65963, Abcam, 1/2000). Membranes were washed six times in TBST, before incubation in secondary antibody diluted in 5% milk in TBST for 1 hr. The following secondary antibodies were used: goat anti-rabbit IgG HRP (P044801-2, Agilent Technologies), goat anti-mouse IgG HRP (P044701-2, Agilent Technologies) and goat anti-llama IgG HRP (ab112786, Abcam). ECL (VWR International) was used as the

chemiluminescent substrate for HRP, and the signal was exposed on X-ray film (Scientific Laboratory Supplies).

2.5 Plasmids & Cloning

All plasmids were maxi-prepped using GeneElute HP Endotoxin-Free Plasmid Maxiprep Kits (Sigma-Aldrich), according to the manufacturer's instructions, and verified by Sanger sequencing (Source BioScience).

Plasmids expressing GFP-TOPBP1 were mutagenized to produce constructs with T779A and Δ774-798 using the Quik-Change II Site-Directed Mutagenesis kit (Agilent Technologies). GFP-TOPBP1 Δ774-849 was generated by mutagenizing GFP-TOPBP1 using In-Fusion Cloning (Takara Bio). GFP-TOPBP1 6A was generated by subcloning a 6A insert synthesised by GeneArt Gene Synthesis (Thermo Fisher Scientific) into GFP-TOPBP1 using In-Fusion Cloning (Takara Bio).

Spot-mRuby3-CIP2A constructs (WT, N230E, 1-571, 561-904) and GFP-AUNIP constructs (WT, Δ12, 5A) were synthesised by GeneArt Gene Synthesis (Thermo Fisher Scientific). GFP-AUNIP was mutated to produce constructs with Δ285-300, Δ340-357, S292A and S347A using the Quik-Change II Site-Directed Mutagenesis kit (Agilent Technologies).

Plasmids expressing GFP-RMI1 were constructed by subcloning human RMI1 cDNA (a gift from Andrew Deans) into mClover2-C1 (a gift from Michael Davidson; Addgene plasmid # 54577). Vectors expressing WT or R41E/Y42F RPA1 were a gift from Marc Wold. Primers used in this study are listed in Table 2.2.

Table 2.2. Primers used in this study

Primer	Sequence (5'-3')
TOPBP1_T779A_F	GGTTCATATCTAAAGGTGCAACGACGGTTTTTCTGTGAG
TOPBP1_T779A_R	CTCACAGAAAAACCGTCGTTGCACCTTTAGATATGAACC
TOPBP1_Δ774-798_F	GCCTGCAAACTCACGCCAGACAGGTCGC
TOPBP1_Δ774-798_R	GCGACCTGTCTGGCGTGAGTTTGCAGGC
TOPBP1_Δ774-849_F	CCACTCTCAGAAGTTATTG
TOPBP1_Δ774-849_R	AACGACGGTTTTTCTGTG
TOPBP1_6A_Insert_F	GGCACACGCCTGCAAACTCACAGAAAAACCGTCGTTGC
TOPBP1_6A_Insert_R	TTCTTCCTTCTCTGACTGGGCCTCTTTCAGTTGAG
TOPBP1_6A_Backbone_F	TCAGAGAAGGAAGAAGCCCCAAAGC
TOPBP1_6A_Backbone_R	TTGCAGGCGTGTGCCAGG
AUNIP_Δ285-300_F	GAAAAGGAGCTCTAGTGTTCCAGGAGTCCTTGTCATTC
AUNIP_Δ285-300_R	GAATGACAAGGACTCCTGGAACACTAGAGCTCCTTTTC
AUNIP_Δ340-357_F	TCAAAATCTGAAGCCTTAAGGTACCGCGGGCC
AUNIP_Δ340-357_R	GGCCCGCGGTACCTTAAGGCTTCAGATTTTGA
AUNIP_S292A_F	CGCTGGCCTTGAGCATCTTCAGTGAAAAGTTGACT
AUNIP_S292A_R	AGTCAACTTTTCACTGAAGATGCTCAAGGCCAGCG
AUNIP_S347A_F	GATAACTTGATTACCTTCAGCGTCCTGGGTAAAGAGCAAAT
AUNIP_S347A_R	ATTTGCTCTTTACCCAGGACGCTGAAGGTAATCAAGTTATC

2.6 Cell Lines & Cell Culture Conditions

All cells were grown in humidified incubators supplied with 5% CO_2 and maintained at 37 °C. Cells were regularly tested for mycoplasma contamination to confirm mycoplasma-free status using a LookOut Mycoplasma PCR Detection Kit (Sigma-Aldrich). RPE-1 FRT/TR cells were purchased from Ximbio and were cultured in Dulbecco's modified Eagle's medium (DMEM; Thermo Fischer Scientific), supplemented with 10% fetal calf serum (FCS; Thermo Fisher Scientific) and 100 µ/ml penicillin/100 µg/ml streptomycin (Lonza). 293FT cells were obtained from Thermo Fisher Scientific and cultured in DMEM supplemented with 10% FCS, 2 mM glutamine (Lonza), 1% MEM non-essential amino acids (Thermo Fisher Scientific) and 500 µg/ml Geneticin (G418; Thermo Fisher Scientific).

2.6 Immunoprecipitations

Plasmids were transfected into 293FT cells using Lipofectamine 2000 (Thermo Fisher Scientific) according to manufacturer's instructions. For preparation of lysates for immunoprecipitations (IPs), cells were washed in phosphate-buffered saline (PBS), and lysed in IP buffer (100 mM NaCl, 0.2% Igepal CA-630, 1 mM $MgCl_2$, 10% glycerol, 5 mM NaF, 50 mM Tris-HCl, pH 7.5), supplemented with cOmplete EDTA-free protease inhibitor cocktail and 25 U/ml Benzonase (Novagen). After nuclease digestion, NaCl and EDTA concentrations were adjusted to 200 mM and 2 mM, respectively, and lysates were cleared by centrifugation. Lysates were then incubated with 50 µl of EZview Red Anti-HA Affinity Gel (Sigma-Aldrich), 15 µl of GFP-Trap magnetic agarose beads (ChromoTek) or 15 µl of SPOT-Trap magnetic agarose beads (ChromoTek) for 2 hr with end-to-end mixing at 4 °C. Immunoglobulin- antigen complexes were washed five times with IP buffer before elution in 2X SDS sample buffer for SDS-PAGE.

2.7 Immunofluorescence and Confocal Microscopy

For analyses of MDC1 and GFP-TOPBP1 foci, cells were grown in 35-mm #1.5H glass bottom dishes (IB-81158, Thistle Scientific). For analyses of GFP-TOPBP1/SPOT-mCherry-CIP2A foci, BLM foci and RAD51 foci, cells were grown on 18 × 18 mm #1.5H glass coverslips (Marienfeld Superior). Indicated SPOT-mCherry-CIP2A constructs were transfected into U2OS cells stably expressing GFP-TOPBP1 using Lipofectamine 3000 (Thermo Fisher Scientific) according to manufacturer's instructions.

Cells were washed once in PBS before fixing for 10 min in 4% paraformaldehyde (PFA) in PBS at room temperature. PBS was added simultaneously during PFA removal to

prevent samples from drying due to PFA evaporation. To visualize BLM recruitment, a mild pre-extraction step (ice-cold 0.2% Triton X-100 in PBS for 1 min followed by a wash in PBS) was carried out between the first PBS wash and the fixation step. After fixation, cells were permeabilized in 0.2% Triton X-100 in PBS for 5 min and then washed once more in PBS before incubation with primary antibodies diluted in antibody buffer (DMEM supplemented with 10% FCS and 0.05% sodium azide; filtered) for 1 h. Where indicated, EdU was labelled with Alexa Fluor 647 using a Click-iT EdU Cell Proliferation Kit for Imaging (Thermo Fisher Scientific) according to the manufacturer's instructions prior to antibody incubations. The following primary antibodies were used at the indicated concentrations: Cyclin A (611268, BD Biosciences, 1/200), GFP (11814460001, Roche, 1/500), GFP (PABG1, ChromoTek, 1/500), γH2AX (05-636, Merck, 1/500), γH2AX (2577, Cell Signaling Technology, 1/500), MDC1 (ab11171, Abcam, 1/5000), RAD51 (70-002, BioAcademia, 1/1000), RPA2 (ab2175, Abcam, 1/300). Cells were washed three times in 0.2% Tween 20 in PBS, before incubation with secondary antibodies diluted in antibody buffer containing 0.5 μg/mL DAPI in the dark for 30 min, followed by five washes in 0.2% Tween 20 in PBS. The following secondary antibodies were from Thermo Fisher Scientific and diluted 1/1000: goat anti-rabbit IgG Alexa Fluor 488 (A11034), donkey anti-mouse IgG Alexa Fluor 568 (A10037), donkey anti-rabbit IgG Alexa Fluor 568 (A10042) and goat anti-mouse IgG1 Alexa Fluor 647 (A21240). Cells in glass bottom dishes were imaged directly in PBS, whereas cells on coverslips were washed once in water and mounted in VECTASHIELD (unless imaging samples with Alexa Fluor 647, in which case SlowFade Diamond Antifade Mountant (Thermo Fisher Scientific) was used to avoid fluorescence quenching).

All images were acquired using an LSM 880 Airyscan inverted microscope (ZEISS) equipped with a 63x/1.4 NA Plan-Apochromat objective and an Airyscan 32-pinhole detector unit. Foci were quantified using Fiji (Schindelin et al., 2012). For Airyscan imaging, raw data

were processed using Airyscan processing with Wiener Filter strength 6 using ZEN Black software version 2.1, yielding 16-bit images with approximately 180-nm lateral resolution. Display of images was adjusted for intensity for optimal display of structures of interest, and false-coloured in magenta or cyan where indicated.

2.8 Generation of RMI1 knockout RPE-1 cells

Small guide RNAs (sgRNAs) were designed to target exon 3 of *RMI1* using the CRISPR Design Tool (https://crispr.mit.edu), targeting the following sequences: sgRNA-RMI1 (anti-sense), AATGAATAAACAAGTGTTTG; sgRNA-RMI1 (sense), GAGCATCCTCTTTTACCCGA. sgRNA oligonucleotides (Sigma-Aldrich) were cloned into AIO-GFP (a gift from Steve Jackson) to produce a vector expressing $Cas9^{D10A}$ nickase, the two sgRNAs, and GFP for sorting successfully transfected cells by FACS. Plasmids were transfected into RPE-1 FRT/TR cells using Lipofectamine 3000 (Thermo Fisher Scientific), according to the manufacturer's instructions. Cells were allowed to recover for 3 days before sorting GFP-positive cells. Colonies were grown from single cells and screened for BLM or RMI1 expression by western blotting.

2.9 Lentiviral Transductions

To assemble lentiviruses, 293FT cells were transfected with pLenti PGK Neo DEST plasmids encoding WT or mutant GFP-BLM and FLAG-RMI1 proteins or GFP alone, along with pHDM-tat1b, pHDM-G, pRC/CMV-rev1b, and pHDM-Hgpm2 lentiviral assembly plasmids (all gifts from Ross Chapman). Viral supernatants were harvested at 24 and 48 h, pooled, passed through a 0.45-μm filter, and stored at −80 °C.

For lentiviral infections, RPE-1 cells were incubated in viral media for 48 h, before selection in fresh medium containing 500 μg/ml G418 for up to 7 days. GFP-positive cells were sorted by FACS to generate stable lines for further study, followed by verification by western blotting.

2.10 RNA Interference

Small interfering RNAs (siRNAs) were transfected using Lipofectamine RNAiMAX (Life Technologies) according to the manufacturer's instructions, with the following sequences: siCtrl (targeting firefly luciferase), 5′-CGUACGCGGAAUACUUCGA- 3′; siRMI1, 5′-AGCCUUCACGAAUGUUGAU-3′.

2.11 DNA Fibre Analysis

Cells were treated with 25 μM 5-iodo-2'-deoxyuridine (IdU) for 30 min, washed three times in PBS, before addition of 2 mM HU for 2 h. Cells were then washed three times in PBS before addition of 250 μM 5-chloro-2'-deoxyuridine (CldU) for 30 min. Cells were washed in PBS and harvested using Accutase (Thermo Fisher Scientific), before being embedded in low melting point agarose (Bio-Rad). Agarose plugs were solidified at 4 °C for 30 min, then incubated overnight at 50 °C in 20 mM NaCl, 100 mM EDTA, 0.5% SDS, 10 mM Tris-HCl pH 8 solution containing 1 mg/ml proteinase K (Sigma-Aldrich). Digested proteins were removed by two washes in 10 mM Tris-HCl, 1 mM EDTA (pH 8). Plugs were used immediately or stored at 4 °C in 500 μM EDTA (pH 8). Plugs were melted at 67 °C for 30 min and cooled to 42 °C before addition of β-Agarase (New England Biolabs) according to the manufacturer's instructions. Plugs were left to digest overnight and the following day, MES buffer (pH 5.4) was added. DNA solution was heated to 65 °C for 30 min, and samples were

left to cool to room temperature before transfer to the Teflon reservoir of a FiberComb Molecular Combing System (Genomic Vision). Silanized coverslips were dipped into the reservoir for 15 min and pulled out at a constant speed of 300 μm/s. Coverslips were baked overnight at 60 °C, and then incubated in 2.5 M HCl for 45 min with gentle shaking to denature the DNA. Coverslips were washed three times in PBS, before incubation in blocking buffer (1% bovine serum albumin, 0.2% Tween 20 in PBS) for 15 min. Samples were then incubated in the following primary antibodies diluted in blocking buffer for 1 h: anti-BrdU (347580, BD Biosciences, 1/25) to recognize IdU, and anti-BrdU (ab6326, Abcam, 1/400) to recognize CldU. Samples were washed three times in 0.2% Tween 20 in PBS, before incubation in the following secondary antibodies from Thermo Fisher Scientific diluted 1/500 in blocking buffer for 45 min: goat anti-rat IgG Alexa Fluor 488 (A11006) and donkey anti-mouse IgG Alexa Fluor 568 (A10037). Samples were washed three times in 0.2% Tween 20 in PBS and once in PBS alone, before mounting in ProLong Gold Antifade Mountant (Thermo Fisher Scientific). At least 100 DNA fibres were analyzed using Fiji (Schindelin et al., 2012).

Chapter 3: A pilot SLiM proteomics screen to identify protein-protein interactions in DNA repair proteins

3.1 Authorship Statement

Dr. Iolanda Vendrell and Dr. Roman Fischer from the Target Discovery Institute performed all of the liquid chromatography tandem mass spectrometry (LC-MS/MS) and assisted with the mass spectrometry data analysis. The peptide pulldowns for mass spectrometry in **Fig. 3.22D** and **Fig. 3.22E** were conducted by Dr. Andrew Blackford.

3.2 Introduction

3.2.1 Experimental Identification of Protein-Protein Interactions

Protein-protein interactions play essential roles in many biological processes, including the DDR (Ding and Kihara, 2019). Identification of protein-protein interactions is useful for identifying multimeric protein complexes and biological pathways, as well as predicting protein function. Protein-protein interactions also form important targets for the development of new therapies to treat human disease.

For many years, the gold standard for the identification of protein-protein interactions was affinity purification of epitope-tagged proteins of interest, coupled to mass spectrometry (Ramberger et al., 2021). One drawback of this method is that it requires genetic introduction of an affinity tag, which can disrupt protein function. Interactions of lower affinity and/or transient interactions can be lost during the purification process. In addition, protein-protein interactions that are dependent on modifications that only occur in specific cell cycle stages (e.g., CDK-dependent phosphorylation events) or in response to stimuli such as DNA damage

(e.g., PIKK-dependent phosphorylation events) may be present in very low amounts or only transiently and so may remain undetected. Methods based on proximity labelling, such as BioID or APEX are becoming increasing popular (Rhee et al., 2013; Sears et al., 2019). These methods result in the biotinylation of interacting and proximal proteins to the protein of interest within live cells, which can be identified by mass spectrometry. These methods allow for the detection of weak and transient interactions in live cells but are dependent on time-consuming cloning and optimisation steps, target cell/tissue delivery and do not provide information about which regions of the protein are involved or mechanisms of regulation by post-translational modifications (Ramberger et al., 2021). Furthermore, it is impossible to distinguish between proteins that are direct interactors of the protein of interest, or merely proximal.

Historically, most research has investigated protein-proteins interactions between structured domains, but it has become clear that a large proportion of functional protein-protein interactions are mediated by compact interaction interfaces within intrinsically disorder regions (Tompa et al., 2014). These short interaction interfaces, SLiMs, are typically up to thirty amino acids in length. The known instances of SLiMs mediate a wide range of important cellular tasks, as described in the introduction.

SLiM-based protein-protein interactions can be recapitulated *in vitro* using synthetic peptides (Ramberger et al., 2021). Amino acids with post-translational modifications (PTMs) can be incorporated, aiding the detection of PTM-specific binding events. Capture of interacting proteins is facilitated by the immobilising synthetic peptides on beads through a linker group, such as biotin and incubating the peptides with cell lysate prior to an affinity purification step. Interacting partners can be eluted and identified using mass spectrometry. Nano-flow liquid chromatography tandem mass spectrometry (LC-MS/MS) is extremely sensitive and so is suitable for detection of weak protein-protein interactions (Kulyyassov et al., 2021). The high local density of the peptides (conjugated to beads) may

overcome the increased dissociation of low affinity, transient interactors by providing alternative binding sites in the direct vicinity, after dissociation of the weak interactor leading to a higher avidity in general (Ramberger et al., 2021). Furthermore, this method has the potential to map these interactions to the specific region of the protein the peptide was derived from and determine if the interaction is dependent on post-translational modifications that may normally be present in very low amounts or transiently within cells. Given these advantages, this approach was used for the screen.

A variation of this method, proteomic peptide-phage display, has been extensively applied to identify SLiM specificity determinants (Teyra et al., 2020), as well as novel SLiM instances in the human proteome (Benz et al., 2022). In proteomic peptide-phage display, a phage-encoded peptide library is designed to display the disordered regions of a target proteome. These peptide libraries are incubated with pre-defined SLiM-binding domains and SLiM-based interactions are identified. These methods form a powerful pipeline for proteome-wide discovery of SLiM-based interactions but only identify interactions for pre-defined SLiM-binding domains. Therefore, in this screen, a method was developed that was not limited to identifying between SLiMs and pre-defined binding domains.

3.2.2 Using Evolutionary Conservation as a Tool to Identify Functional SLiMs in the DDR

Due to the fundamental importance of maintaining genome integrity in the face of DNA damage, all known forms of life have evolved a complex repertoire of biochemical pathways to detect and repair such damage (Jackson and Bartek, 2009). Around 70-80% of DDR proteins have orthologous counterparts in distant eukaryotes, meaning the orthologues of many DDR proteins are present in most eukaryotic organisms (Arcas et al., 2014). To this end, frequent examples of the use of model organisms, such as mouse, chicken, zebrafish and yeast to analyse

the function of DDR proteins exist in literature. Although these orthologues have diverged, they display significant similarity in both structure and function to the human orthologues. Furthermore, cross-species complementation has been performed, complementing human cells with orthologues or vice versa. A key example of this is the seminal work conducted by Lee & Nurse in 1987 (Lee and Nurse, 1987). Lee & Nurse used a temperature-sensitive mutation in the *S. pombe cdc2* gene, which was unable to proceed beyond the G1 and G2 checkpoints. Transformation of the *cdc2* temperature-sensitive mutant with the human *CDC2* gene rescued the checkpoint defect. The observation that a human protein could complement this defect demonstrated that elements of cell cycle control are conserved between yeast and humans.

Based on the known instances of SLiMs, the general features of SLiMs are becoming clear (Tompa et al., 2014). On average, SLiMs are up to 30 amino acids in length, with 3-4 core residues conferring most of the interaction specificity. Whilst resources exist for predicting protein domains from sequence alone (e.g. Pfam; El-Gebali et al., 2018), it is more challenging to do this for SLiMs due to their low sequence complexity. Furthermore, the protein-protein interactions they mediate often depend on other factors, such as PTMs. As such, it is experimentally challenging to identify functional SLiMs based on sequence alone. Due to their relatively small size, SLiMs are swiftly lost and gained in evolution but a subset are conserved throughout vertebrates, other metazoans and sometimes to yeast. These SLiMs are likely to be involved in critical processes.

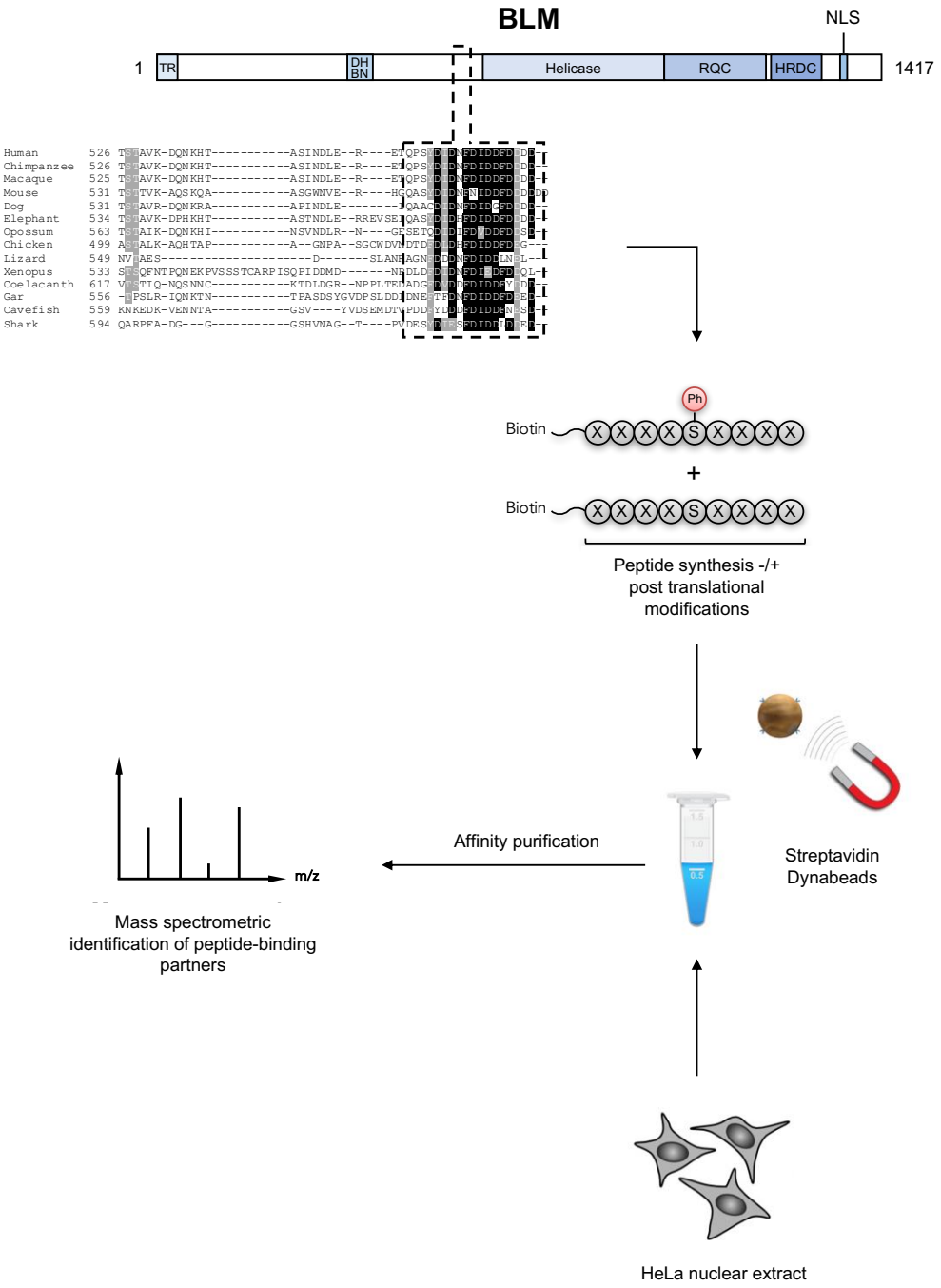


Fig. 3.1. Bioinformatic and proteomics pipeline for the identification and subsequent characterization of functionally important SLiMs.
Multiple sequence alignments are generated based on primary amino acid conservation, such as the one shown for BLM helicase. Within these sequence alignments SLiMs are identified as an "island of conservation" in intrinsically disordered regions. Once a potential SLiM has been identified, it is synthesised as a biotinylated peptide. These peptides were dissolved and conjugated to streptavidin coupled Dynabeads and used as baits to identify physiological binding partners from HeLa nuclear extract by mass spectrometry. Grey spheres, amino acid residues; Ph, phosphorylation.

3.2.3 Aims and Objectives

Firstly, I aimed to develop a combined bioinformatics and proteomics approach to identify protein-protein interactions in the DDR based on SLiMs, using evolutionary conservation as a tool to assist in their detection (outlined in **Fig. 3.1**). Secondly, I aimed to use this pipeline to identify and subsequently characterize conserved SLiMs in DNA repair proteins in order to test whether these methods can successfully identify SLiMs at scale.

3.3 Results

3.3.1 Pipeline to Identify Functional SLiMs in DDR Proteins

To identify SLiMs in a protein of interest based on evolutionary conservation, a multiple sequence alignment was required. A multiple sequence alignment facilitates the study of evolutionary relationships between sequences (Chatzou et al., 2016). For each protein of interest, such as BLM helicase (used as a representative example in **Fig. 3.2**), amino acid sequence alignments from representative vertebrate orthologues were generated. Despite the high level of conservation of DDR in distant eukaryotes, various key DDR proteins are absent in fungi and invertebrates. Systematic analysis of 118 DDR proteins in the human proteome and their orthologues in 46 species revealed that 50 proteins present in humans could not be found in model fungi and invertebrates (Arcas et al., 2014). For example, key HDR components *BRCA1, BARD1* and *BRCA2* were lost in model fungi (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe)* and model invertebrates (*Caenorhabditis elegans* and *Drosophila melanogaster*).

Due to these absences and a primary interest in function of the human proteins, in this screen sequence alignments were restricted to vertebrate orthologues. However, it is important

to note that there are instances where DDR components are missing in vertebrate species. For example, the zebrafish proteome has no known BRCA1 orthologue (Howe et al., 2013).

Manual sequence alignments were generated for the following DSB repair proteins, given their key role in DNA repair: 53BP1, AUNIP, BARD1, BLM, BRCA1, BRCA2, CtIP, ERCC6L2, MDC1, POLQ, RECQL4, RNF168, SHLD2, SLX4, TOPBP1, and XRCC4 (Section 8.1). These proteins were selected from a curated list of proteins associated with DSB repair, assembled by experts (Pearl et al., 2015).

Manual sequences alignments are labour intensive but guarantee the inclusion of sequences spanning an appropriate evolutionary distance, as well as the ability to exclude partial sequences. To generate multiple sequence alignments, representative protein sequences across vertebrate evolution were obtained from NCBI (https://www.ncbi.nlm.nih.gov). HomoloGene (https://www.ncbi.nlm.nih.gov/homologene/) allows for the automatic detection of orthologues (similarity that is attributable to descent from a common ancestor) among the annotated genes of several eukaryotic genomes. Starting with a protein of interest (such as BLM, **Fig. 3.2A**), HomoloGene automatically collects orthologue sequences using BlastP (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) and the NCBI orthologues database (**Fig. 3.2B**) stores all orthologues for a given protein query. Sequences are matched up and placed into groups, using a taxonomic tree based on sequence similarity (**Fig. 3.2C**). For the human sequence, the largest isoform was selected, to avoid the use of partial sequences and smaller splice variants (**Fig. 3.2D**). Orthologous protein sequences were chosen from species representing distinct taxonomic groups (**Fig. 3.2E**), such as: mouse (*Mus musculus*), elephant (*Loxodonta africana*), opossum (*Monodelphis domestica*), chicken (*Gallus gallus*), lizard (*Anolis carolinensis*), frog (*Xenopus laevis*), coelacanth *(Latimeria chalumunae*), spotted gar (*Lepisosteus oculatus*), zebrafish (*Danio rerio*) and shark (*Callorhinchus milii*).

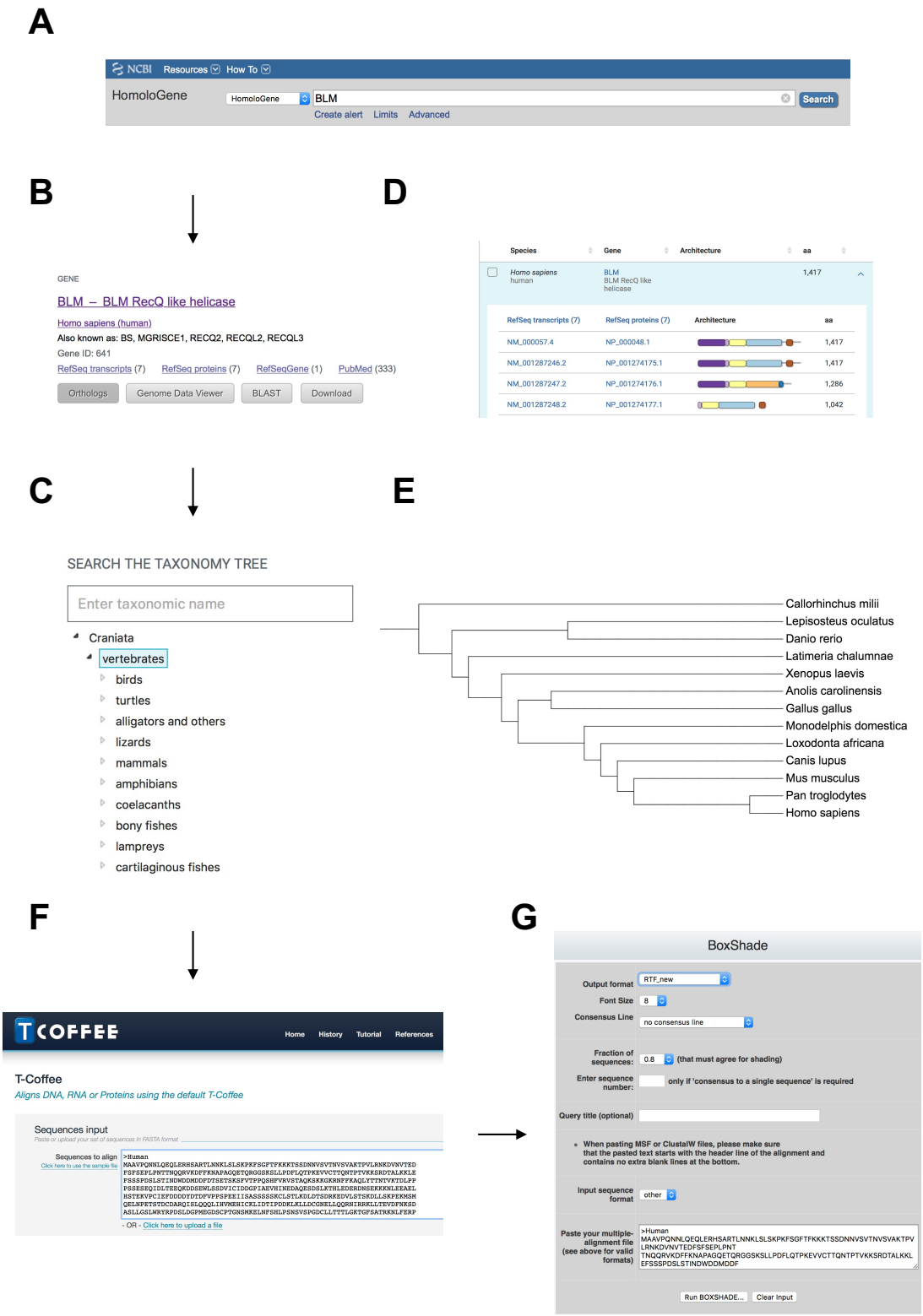


Fig. 3.2. Generation of manual sequence alignments.
A) Starting with a protein of interest (such as BLM), HomoloGene collects all orthologue sequences. B) NCBI orthologues database within HomoloGene stores all orthologues for a given protein query. C) Within the NCBI orthologues database, sequences are matched up and placed into groups, using a taxonomic tree based on sequence similarity. D) For the human sequence, the largest isoform was acquired. E) Orthologous protein sequences were acquired from species representing distinct taxonomic groups, with increasing evolutionary distance from humans. F) Once representative protein sequences across vertebrate evolution were obtained, they were aligned based on primary amino acid conservation, using T-Coffee. G) Within the multiple sequence alignment, similar residues were highlighted using Boxshade.

Once representative protein sequences across vertebrate evolution were obtained, they were aligned based on primary amino acid conservation, using T-Coffee (Notredame et al., 2000), accessible at (https://www.ebi.ac.uk/Tools/msa/tcoffee/). T-Coffee generates multiple sequence alignments by generating a library of pairwise alignments. Each pair of aligned residues is assigned a weight, which is used to guide the development of the multiple sequence

alignment. Then, similar residues were highlighted using Boxshade (**Fig. 3.2G**). During the writing of this thesis, Boxshade was discontinued. The alternative tool Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html) is available. In the program output, identical and conservative residues are shaded black or grey respectively (**Fig. 3.3A**). Divergent amino acids are shaded white. For this thesis, 80% of sequences were required to contain a similar/identical amino acid at each corresponding residue to be shaded. An amino acid substitution can be classified as either radical or conservative depending on whether it induces a change in a certain physicochemical property (e.g., charge, hydrophobicity or size) of an amino acid (Zhang, 2000). Shading similar residues grey (such as R or K) considers that conservative substitutions are likely to be neutral in relation to protein function.

In the output (**Fig. 3.3A**), structured domains typically manifest as large regions of shaded residues (such as the bottom row of the sequence alignment in **Fig. 3.3A**). In contrast, unconserved (and typically intrinsically disordered) regions are often represented by larger regions without shaded residues (such as the second row of the sequence alignment in **Fig. 3.2A**). This is because in general, disordered regions are less conserved compared to structured regions, due to the lack of structural constraints in intrinsically disordered regions (Erdős et al., 2021).

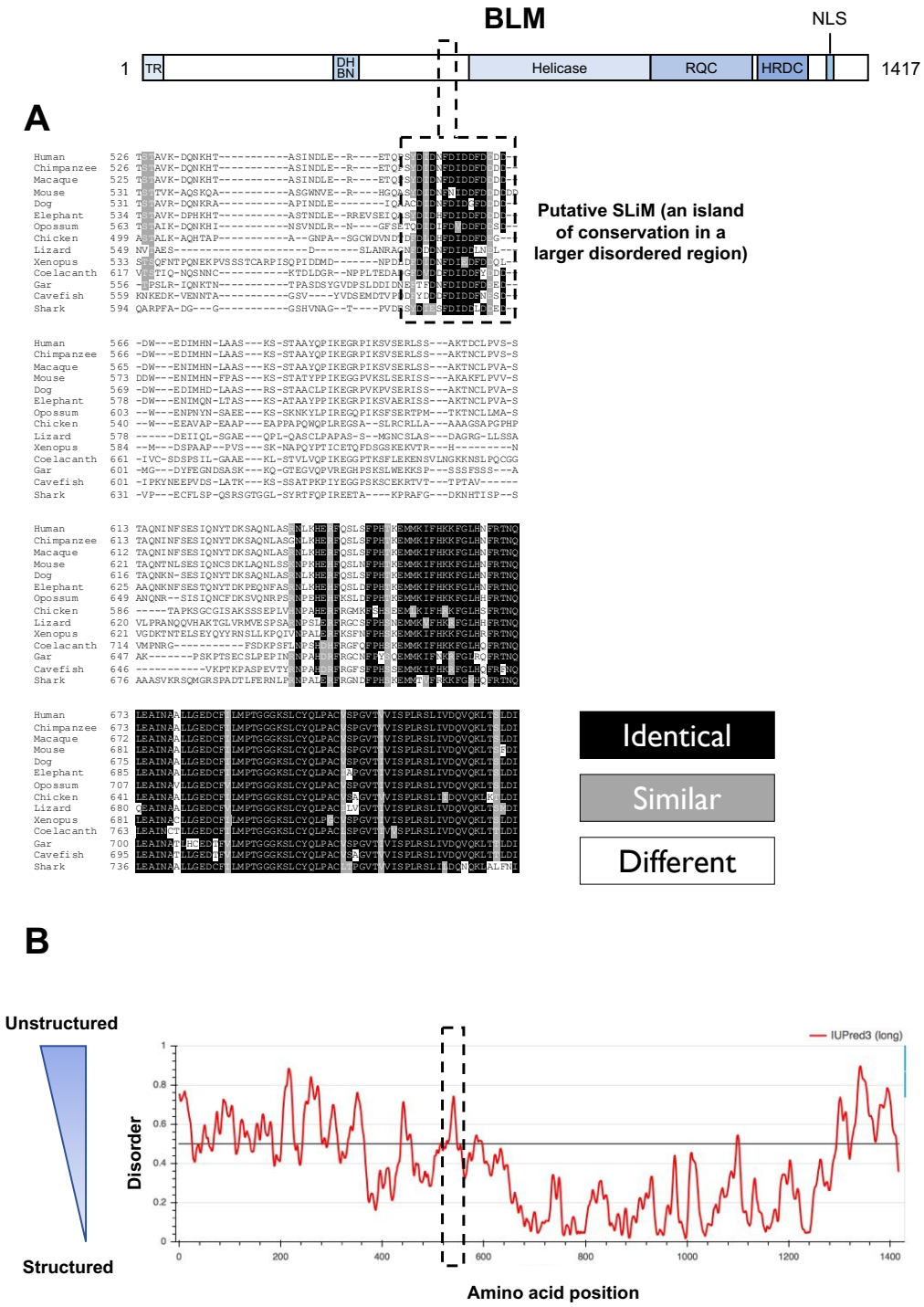


Fig. 3.3. Identification of SLiMs in manual sequence alignments.
A) A representative sequence alignment based on primary amino acid conservation for BLM. Identical and conservative residues are shaded black or grey respectively, divergent amino acids are shaded white. Structured domains typically are represented by large regions of shaded residues (such as the bottom row of the sequence alignment). Unconserved (and typically intrinsically disordered) regions are often represented by larger regions without any shaded residues (such as the second row of the sequence alignment). Putative SLiMs typically appear as islands of conservation within larger unconserved regions in multiple sequence alignments, such as the one shown. B) IUPred3 was used to verify that the predicted SLiMs were located within intrinsically disordered regions of proteins. A higher IUPred score represents a higher level of intrinsic disorder, whilst a lower IUPred score indicates the region is structured. The boxed region of the IUPred profile corresponds to the SLiM highlighted in A.

Known attributes of SLiMs were used to aid the prediction of functional SLiMs in these sequence alignments: a short sequence (typically up to 30 amino acids), conserved over a large evolutionary distance and located in intrinsically disordered regions of proteins (Tompa et al., 2014). Putative SLiMs typically appear as islands of conservation within larger unconserved regions in multiple sequence alignments (**Fig. 3.3A**). To verify that predicted SLiMs were located within intrinsically disordered regions of proteins, the IUPred3 tool was used (Erdős et al., 2021 **Fig 3.3B**). The IUPred web server (https://iupred3.elte.hu) takes an amino acid sequence as an input and predicts the tendency for each amino acid to be in a disordered region of the protein. A higher IUPred score represents a higher level of intrinsic disorder, whilst a lower IUPred score indicates the region is structured.

3.3.2 Synthesis of SLiMs as Biotinylated Peptides and Peptide Pulldowns for Mass Spectrometry

Using the sequence alignments generated in section 3.3.1, 36 potential SLiMs were identified and synthesised as biotinylated peptides to assess their potential function (**Table 3.1**). 14 putative SLiMs had a CDK (S/TP), a PIKK (S/TQ) or a CK2 (S/TxxD/E) consensus sequence. For each of these SLiMs a corresponding phospho-peptide was generated, resulting in 51 peptides in total. These PTM consensus sequences were chosen due to their relevance in the DNA damage response, although other modifications could be assessed using this method.

Table 3.1. SLiMs identified and synthesised as biotinylated peptides

Peptide	Residues	Amino Acid Sequence
53BP1	301-313	Biotin-SGSG-EPEVLSTQEDLFD
53BP1	301-313	Biotin-SGSG-EPEVL(pS)(pT)QEDLFD
AUNIP	340-357	Biotin-SGSG-DLLFTQDSEGNQVIRHQF
AUNIP	340-357	Biotin-SGSG-DLLF(pT)QDSEGNQVIRHQF
AUNIP	340-357	Biotin-SGSG-DLLFTQD(pS)EGNQVIRHQF
BARD1	140-160	Biotin-SGSG-KNSIKMWFSPRSKKVRYVVSK
BARD1	140-160	Biotin-SGSG-KNSIKMWF(pS)PRSKKVRYVVSK
BLM	137-163	Biotin-SGSG-KKLEFSSSPDSLSTINDWDDMDDFDTS
BLM	290-309	Biotin-SGSG-EFDDDDYDTDFVPPSPEEII
BLM	290-309	Biotin-SGSG-EFDDDDYDTDFVPP(pS)PEEII
BLM	552-568	Biotin-SGSG-DIDNFDIDDFDDDDDWE
BLM	1392-1417	Biotin-SGSG-ANSKLGIMAPPKPINRPFLKPSYAFS
BRCA1	455-467	Biotin-SGSG-NIEDKIFGKTYRK
BRCA1	501-527	Biotin-SGSG-KLKRKRRPTSGLHPEDFIKKADLAVQK
BRCA2	1100-1132	Biotin-SGSG-NHNLTPSQKAEITELSTILEESGSQFEFTQFRK
BRCA2	2401-2414	Biotin-SGSG-RPTKVFVPPFKTKS
BRCA2	3340-3360	Biotin-SGSG-IADEELALINTQALLSGSTGE
BRCA2	3340-3360	Biotin-SGSG-IADEELALIN(pT)QALLSGSTGE
CtIP	651-676	Biotin-SGSG-ENIQWSIDPGADLSQYKMDVTVIDTK
CtIP	651-676	Biotin-SGSG-ENIQWSIDPGADL(pS)QYKMDVTVIDTK
ERCC6L2	796-812	Biotin-SGSG-PGQLTLLQCGFSKLLET
FANCM	1126-1143	Biotin-SGSG-LSTDQDESLLLFEDVNTE
FANCM	1126-1143	Biotin-SGSG-LSTDQDE(pS)LLLFEDVNTE
MDC1	184-204	Biotin-GGSG-SRTTSSSVIVPESDEEGHSPV
MDC1	184-204	Biotin-GGSG-SRTTSSSVIVPE[pS]DEEGHSPV
POLQ	1317-1334	Biotin-SGSG-DFEDSFYLDTQSEKIIQQ
POLQ	1317-1334	Biotin-SGSG-DFEDSFYLD(pT)QSEKIIQQ
POLQ	1434-1454	Biotin-SGSG-EVSVTDSQLNSFLQGYQTQET
POLQ	1434-1454	Biotin-SGSG-EVSVTD(pS)QLNSFLQGYQTQET
POLQ	1477-1497	Biotin-SGSG-PVPETSLNMSDSLLFDSFSDD
POLQ	1631-1652	Biotin-SGSG-WSGASFDLSPGLQRILDKVSSP
RAD17	632-651	Biotin-SGSG-LPLSQNSASELPASQPQPFS
RAD17	632-651	Biotin-SGSG-LPL(pS)QNSASELPA(pS)QPQPFS
RECQL4	68-87	Biotin-SGSG-AAEEAPEPRCWGPHLNRAAT
RECQL4	101-124	Biotin-SGSG-SVPDYGQRLKANLKGTLQAGPALGRR

RECQL4	171-197	Biotin-SGSG-RLQHLQASLSQRLGSLDPGWLQRCHSE
RMI1	240-260	Biotin-SGSG-LGPSDEELLASLDENDELTAN
RMI1	308-329	Biotin-SGSG-DGELDDFSLEEALLLEETVQKE
RNF168	223-234	Biotin-SGSG-GDIQKYLTPKSQ
RNF168	223-234	Biotin-SGSG-GDIQKYL(pT)PKSQ
RNF168	272-284	Biotin-SGSG-EDMPTLSPQISLG
RNF168	558-571	Biotin-SGSG-SQKSVFQMFQRCTK
SHLD2	265-283	Biotin-SGSG-ISTDTEFLSIITSSQVAFL
SHLD2	333-345	Biotin-SGSG-QNQAYSLELFSPV
SLX4	603-628	Biotin-SGSG-SASQREHQALQDLVDLAREGLSASPW
TOPBP1	486-503	Biotin-SGSG-ADEDLLSQYENGSSTVVE
TOPBP1	486-503	Biotin-SGSG-ADEDLL(pS)QYENGSSTVVE
TOPBP1	774-798	Biotin-SGSG-RKTVVTPLDMNRFQSKAFRAVVSQH
TOPBP1	774-798	Biotin-SGSG-RKTVV(pT)PLDMNRFQSKAFRAVVSQH
XRCC4	255-278	Biotin-SGSG-DSIISSLDVTDIAPSRKRRQRMQR
XRCC4	318-334	Biotin-SGSG-SLETLRNSSPEDLFDEI

Once synthesised, these peptides were dissolved in water or DMSO (**Table 2.1**) and the peptide pulldown protocol used was adapted from Ochi et al., 2015. Before proceeding with peptide pulldowns for mass spectrometry, it was necessary to ensure the quantity of biotinylated peptide was sufficient to saturate the streptavidin-coupled Dynabeads. Theoretically, the capacity of 25 µl of Dynabeads is 0.1 µg of peptide, but to ensure the beads were saturated, a peptide titration was performed. A biotin-specific antibody was identified and validated (**Fig. 3.4A**). 2.5 µg of peptide was found to saturate the streptavidin-coupled Dynabeads (**Fig. 3.4B**). To ensure maximum saturation, 10 µg of peptide per 25 µl of streptavidin-coupled Dynabeads was used in subsequent experiments. To identify physiological binding partners of biotinylated peptides, the interacting partners must be eluted from the Dynabead-bound peptides. Using a γH2AX peptide known to interact with MDC1 (Stucki et al., 2005), four serial elutions were performed to determine the optimal elution conditions. When compared to the first elution only a small amount of MDC1 was present in the second elution (**Fig 3.4C**) but to ensure all interacting proteins are stripped from the

Dynabead-bound peptides, two serial elutions were performed in all subsequent peptide pulldown experiments.

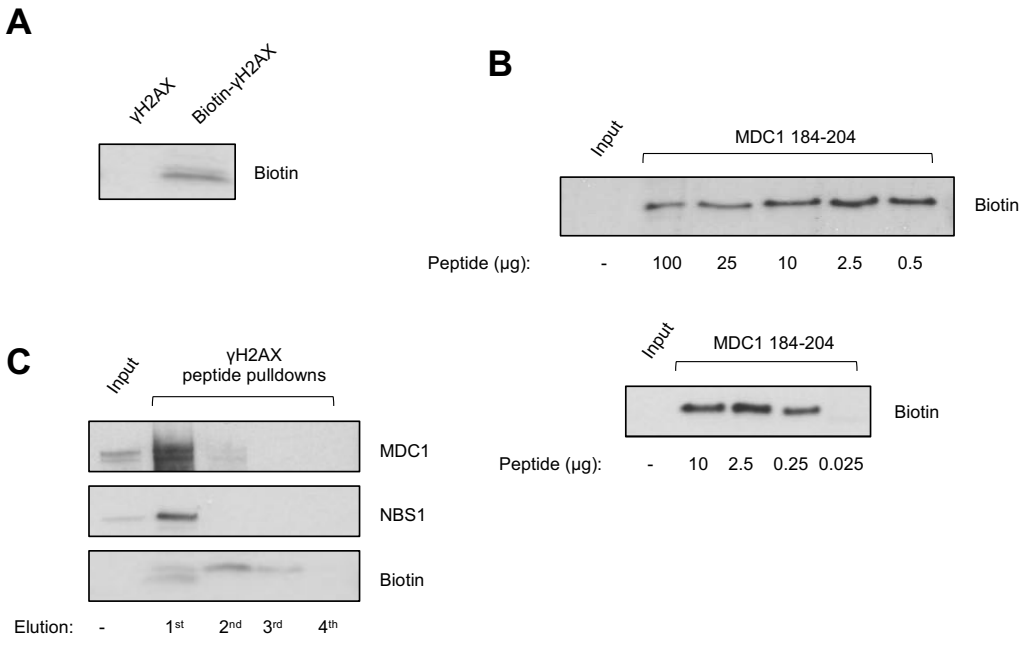


Fig. 3.4. Optimization of the peptide pulldown method.
Validation of a biotin-specific antibody, using a biotinylated and non-biotinylated peptide derived from H2AX-pS139 (γH2AX). B) Peptide pulldowns demonstrating that 2.5 μg of peptide saturated the streptavidin-coupled Dynabeads. C) Peptide pulldowns demonstrating that two elutions are sufficient to strip the Dynabeads of the interactors.

All peptides synthesised above (**Table 3.1**) were conjugated to streptavidin-coupled Dynabeads and used as baits to identify physiological binding partners from HeLa nuclear extract by LC-MS/MS (**Fig 3.1**). Affinity purification is widely used in proteomics research to enrich for target proteins and to reduce sample complexity. This facilitates mass spectrometric identification as most non-binding proteins are washed away after washing. Consistent with their biological role, the vast majority of DDR proteins are nuclear, justifying the use of a nuclear extract, as opposed to whole cell extract (Pearl et al., 2015). 10% of the sample was analysed by silver stain and the remaining 90% was retained for mass spectrometry. Silver

staining was used to verify the pulldown had worked (**Fig. 8.2.1**), with only samples with identical banding patterns in two independent repeats being analysed for mass spectrometry.

3.3.3 Identification of High-Confidence Interacting Proteins

Mass spectrometric analysis of the samples revealed that their complexity varied widely. Although some samples had 10-20 putative interacting proteins, more complex samples had up to 1500 putative interacting proteins. Due to the nature of SLiM based protein-protein interactions, each SLiM is only likely to have one direct interacting partner. The sample may contain indirect binding partners that copurify with the interactor, however some proteins in the sample are likely to be non-physiological interactors (or background contaminants/proteome) that were copurified and identified by mass spectrometry (Mellacheruvu et al., 2013). Accordingly, there was need to develop a strategy to distinguish between biologically relevant interactors of each peptide and the background proteome.

Each sample was filtered to identify high-confidence candidate interacting proteins (**Fig. 3.5A**). For a protein to be classified as a high-confidence candidate interacting protein, it had to meet the following criteria: (1) ≥2 unique peptides present corresponding to the interacting protein and (2) not considered to be part of the background proteome (Sowa et al., 2009). To identify the background proteome, a list of all proteins identified in all of the peptide pulldowns was generated. Then, the number of peptide pulldowns that each protein appeared in was calculated and this number was expressed as a proportion of the total number of peptide pulldowns. If the protein that the biotinylated peptide was derived from appeared amongst the putative interacting proteins for that peptide pulldown, it was not counted. This was to take into account the likelihood that the biotinylated peptide used as the bait was detected by the mass spectrometer. The lower the percentage of peptide pulldowns a protein is in, the lower the likelihood that protein is part of the background proteome. For example, the keratin KRT9

was present in 90% of all peptide pulldowns and therefore likely to be part of the background proteome. In contrast, the protein ARID1A was present in less than 1% of all peptide pulldowns and therefore is unlikely to be part of the background proteome. A threshold of 25% was used, meaning if a protein was present in greater than 25% of all peptide pulldowns, it was considered to be part of the background proteome. This resulted in a list of 245 proteins. Above this threshold, the list largely consisted of proteins that have been annotated to be common contaminants in other affinity purification experiments, such as keratins, ribosomal proteins and histones (Mellacheruvu et al., 2013). Below this threshold, many DDR proteins were observed. In addition to this manually generated background proteome, an online database of common contaminants in affinity purification mass spectrometry experiments was used (Mellacheruvu et al., 2013). This repository is available at http://www.crapome.org/. The resource contains all proteins in the human proteome, as well as the number of experiments in the database that the protein has been detected in. This repository was acquired. If a protein was found to be in greater than 25% of all experiments in this database, it was considered to be part of the background proteome. This resulted in a list of 951 proteins. This threshold was set for the same reasons as above. 58 proteins were unique to the manually generated background proteome, 756 proteins were unique to the published background proteome and 187 proteins were found in both lists. Commonly identified contaminants from both lists were then excluded from the dataset (**Table 8.3.1**). Once the background proteome was excluded from a peptide pulldown corresponding BLM 290-309, the number of putative interacting proteins was reduced from 650 (**Fig. 3.5B**) to 225 (**Fig. 3.5C**). Furthermore, the biologically relevant interacting protein, TOPBP1 (validated in section 3.3.8) was the highest candidate interacting protein once the background proteome was excluded.

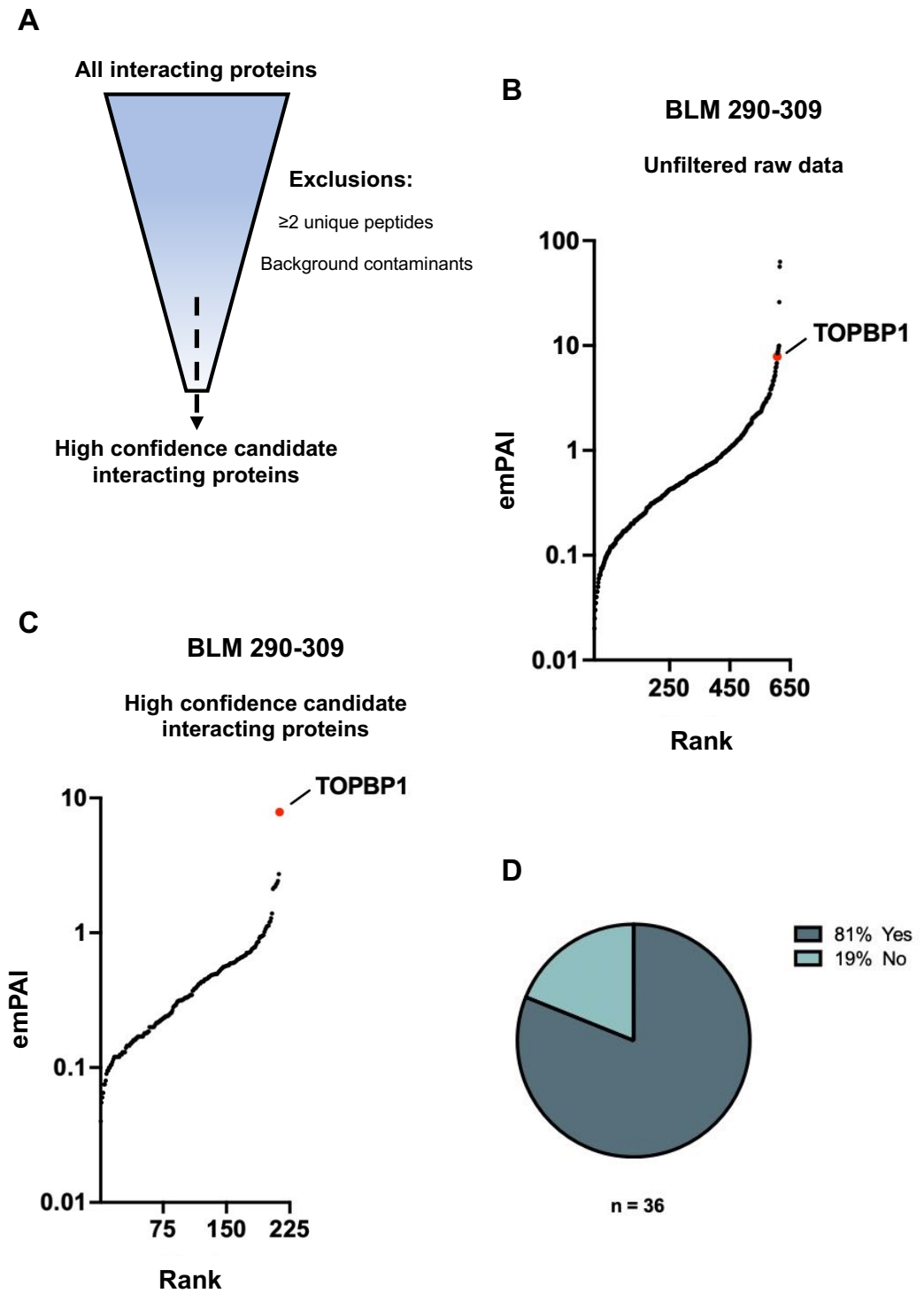


Fig. 3.5. Identification of high-confidence candidate interacting proteins.
A) Interacting proteins in LC-MS/MS samples were filtered according to the following criteria: (1) ≥2 unique peptides present and (2) not considered to be part of the background proteome. B-C) Plot of LC-MS/MS hits from BLM 290-309 peptide pulldowns, ranked by emPAI score (number of unique peptide sequences adjusted for protein size), demonstrating the impact of excluding the background proteome. emPAI score shown is the average of two independent repeats. D) Of the 36 SLiMs in the pilot screen, 81% were found to have high-confidence candidate interacting proteins, 19% did not.

Out of the 36 putative SLiMs, 81% (n = 29) had high-confidence candidate interacting proteins and 19% (n = 7) did not (**Fig. 3.5D**). The 7 putative SLiMs without high-confidence candidate interacting proteins can be found in Table 3.2. This corresponded with the silver stain gels (**Fig. 8.2.1**), demonstrating that this was not due to an error in mass spectrometry sample preparation.

Table 3.2 SLiMs without high-confidence candidate interacting proteins

Peptide	Residues
BLM	1392-1417
BRCA2	3340-3360
CtIP	651-676
RAD17	632-651
RMI1	308-329
RNF168	272-284
SHLD2	333-345

For each peptide pulldown with high-confidence candidate interacting proteins, with the exception of the RECQL4 peptide pulldowns, the interacting proteins were ranked according to their emPAI value. The emPAI value is the number of unique peptide sequences adjusted for protein size (Ishihama et al., 2005). The iBAQ value is the sum of the peak intensity values for each peptide of a given protein adjusted for protein size (Schwanhausser et al., 2011). Over the course of this DPhil, our collaborators changed their method of analysing raw mass spectrometry data and iBAQ values were given for the RECQL4 peptide pulldowns.

To aid the identification of functional protein complexes present in amongst the high-confidence candidate interacting proteins, the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database was used (Szklarczyk et al., 2021). This database is freely available at https://string-db.org. The interacting proteins for each pulldown were supplied as the input and a protein-protein interaction network was generated. The confidence score threshold was set to "low", so that protein-protein interactions identified in other organisms

were included in the output. The network subtype was changed from full (functional and physical protein associations) to physical interaction subnetworks. Lastly, the active interaction source was restricted from text-mining (co-occurrence of gene names in abstracts), neighbourhood, gene fusion, co-occurrence, co-expression, experiments and databases to experiments and databases exclusively. Full STRING networks for peptide pulldowns with high-confidence candidate interacting proteins can be found in **Fig. 8.4.1**, STRING networks for the key candidate interacting proteins can be found in this section. The thickness of the line between the proteins indicates the strength of confidence in the interaction.

55% (n = 16) of peptide pulldowns with high-confidence candidate interacting proteins had DDR proteins in the top ten candidate interacting proteins (**Table 3.3**).

Table 3.3. SLiMs with DDR proteins in the top ten high-confidence candidate interacting proteins

Peptide	Residues
53BP1	301-313
AUNIP	340-357
BARD1	140-160
BLM	137-163
BLM	290-309
BLM	552-568
BRCA2	1100-1132
ERCC6L2	796-812
MDC1	184-204
POLQ	1317-1334
RECQL4	68-87
RECQL4	101-124
RMI1	240-260
RNF168	558-571
TOPBP1	486-503
TOPBP1	774-798

Many SLiM based candidate interactions identified here were between DDR proteins and key DDR signalling conduits or “hub” proteins, the replication protein A (RPA)

complex, the MRN complex and PCNA. It was observed that certain DDR proteins, especially those that bind to DSB ends (MRE11) were part of the background proteome. If these proteins were in the top ten of high-confidence interacting proteins and other members of the complex (if applicable) were present, these proteins were considered to be high-confidence candidate interacting proteins.

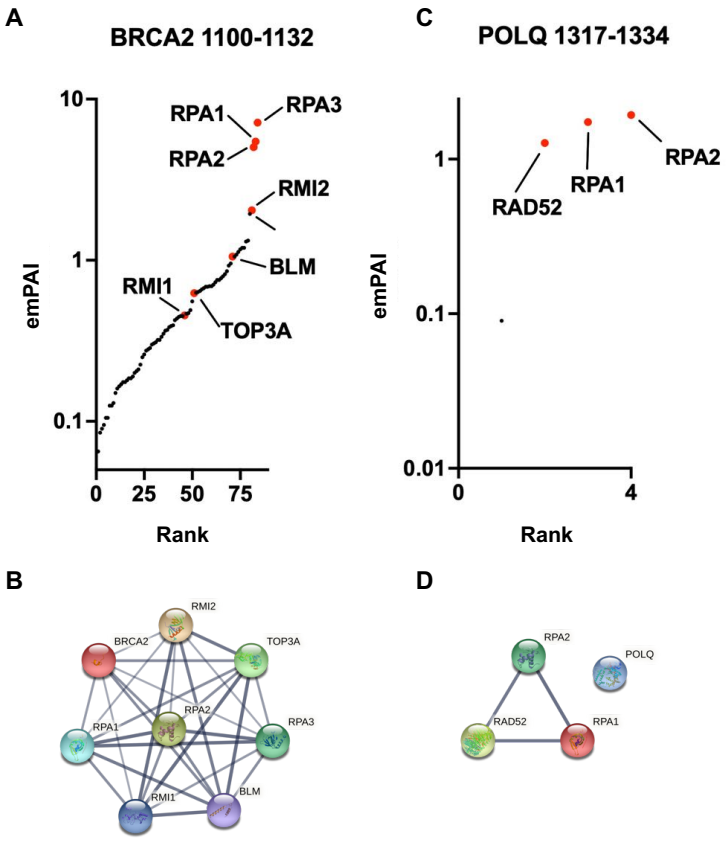


Fig. 3.6. Identification of candidate SLiM-based interactions between RPA, BRCA2 and POLQ
A) Plot of LC-MS/MS hits from peptide pulldowns of the phosphopeptide corresponding to BRCA2 1100-1132, ranked by emPAI score. The RPA complex and BTR complex subunits were high-confidence candidate interacting proteins. RPA and BTR subunits are highlighted in red. emPAI score shown is the average of two independent repeats. B) STRING network of interactions between the high-confidence candidate interacting proteins of the phosphopeptide corresponding to BRCA2 1100-1132 (Szklarczyk et al., 2021). C) Plot of LC-MS/MS hits from peptide pulldowns of the phosphopeptide corresponding to POLQ 1317-1334, ranked by emPAI score. The RPA complex subunits were high-confidence candidate interacting proteins. RPA and RAD51 are highlighted in red. emPAI score shown is the average of two independent repeats. D) STRING network of established interactions between the high-confidence candidate interacting proteins of peptides corresponding to POLQ 1317-1334 (Szklarczyk et al., 2021).

Peptides corresponding to BRCA2 1100-1132 (**Fig. 3.6A,B**), POLQ 1317-1334 (**Fig. 3.6C,D**), BLM 137-163 (**Fig. 3.22D**), BLM 552-568 (**Fig. 3.22E**) and RMI1 240-260 (**Fig. 3.22F**) have RPA components as the highest candidate interacting proteins. RMI2 was also one of the top ten high-confidence candidate interacting proteins of BRCA2 1100-1132 (**Fig. 3.6A,B**). The other BTR complex members, BLM, TOP3A and RMI1 were high-confidence candidate interacting proteins of this peptide.

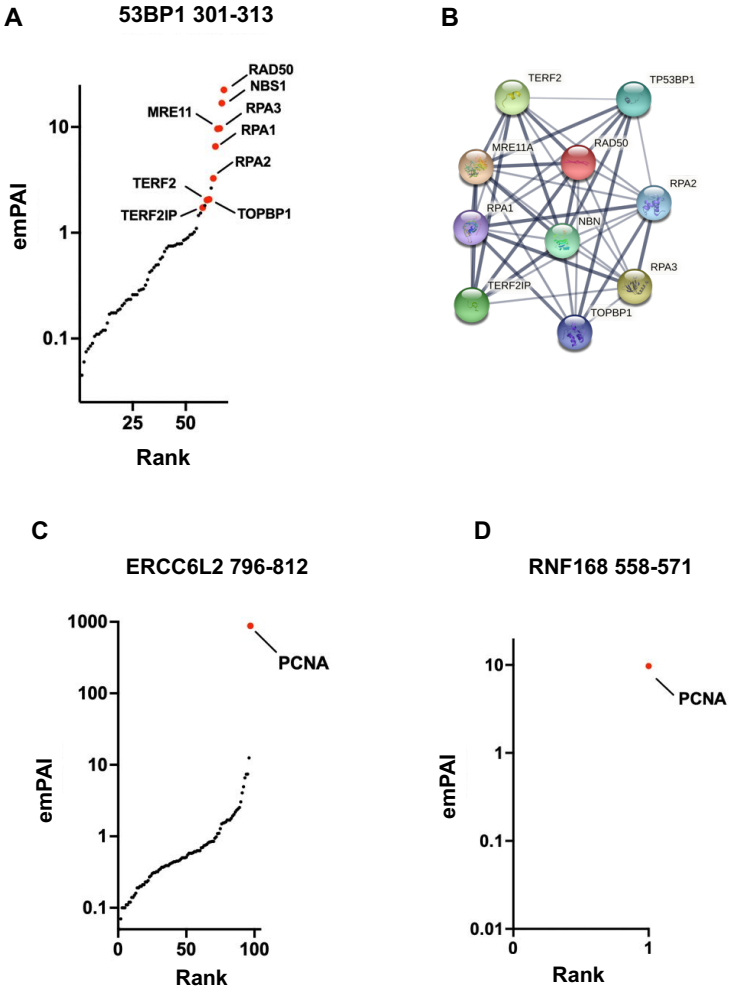


Fig. 3.7. Identification of candidate SLiM-based interactions between MRN and 53BP1, as well as PCNA, ERCC6L2 and RNF168.
A) Plot of LC-MS/MS hits from peptide pulldowns of the phosphopeptide corresponding to 53BP1 301-313, ranked by emPAI score. MRN, RPA and Shelterin complex components (highlighted in red) were high-confidence candidate interacting proteins. emPAI score shown is the average of two independent repeats. B) STRING network of interactions between the high-confidence candidate interacting proteins of the phosphopeptide corresponding to 53BP1 301-313 (Szklarczyk et al., 2021). C-D) Plots of LC-MS/MS hits from peptide pulldowns of peptides corresponding to ERCC6L2 796-812 (C) and RNF168 558-571 (D), ranked by emPAI score. PCNA (highlighted in red) was the highest candidate interacting protein for both peptide pulldowns. emPAI score shown is the average of two independent repeats.

A peptide corresponding to 53BP1 301-313 has MRN components as the highest ranked candidate interacting proteins (**Fig 3.7A,B**). The RPA complex members are also amongst the highest candidate interacting proteins, as is TOPBP1 and the Shelterin components TERF2 and RAP1 (TERF2IP). The interaction between POLQ 1317-1334 and RPA, as well as 53BP1 301-313 and MRN, is dependent on phosphorylation. The corresponding native peptides did not have any high-confidence candidate interacting proteins. Lastly, peptides corresponding to ERCC6L2 796-812 (**Fig 3.7C**) and RNF168 558-571 (**Fig 3.7D**) had PCNA as the highest candidate interacting protein.

STRING analysis revealed that peptides corresponding to BARD1 140-160 had a network of ribosomal biogenesis factors among the top ten candidate interacting proteins (**Fig. 3.8A-D**). For paired native and phosphorylated peptides, assessing fold enrichment of each protein (in the native and/or the phosphorylated peptide) can aid the identification of the biologically relevant interacting protein. This analysis demonstrated that SLX4 and its binding partner, SLX4IP, were enriched in the BARD1 140-160 peptide pulldown when the peptide was phosphorylated (**Fig 3.8E**). Out of the other SMX tri-nuclease complex components, EME1, XPF and TERF2 were also enriched in the phosphopeptide pulldown. It is important to note that this type of analysis must not be done without analysing the proteins present in native and phosphopeptide pulldowns individually. This is because by this analysis, proteins unique to the native or phosphopeptide pulldown will be missed. For example, MUS81, a SMX tri-nuclease complex component and BLM are present in the phosphopeptide pulldown (**Fig 3.8C**) but not in the native peptide pulldown, so they are not hits in **Fig 3.8E**. Of note, BLM has been shown to interact with SLX4IP (Panier et al., 2019).

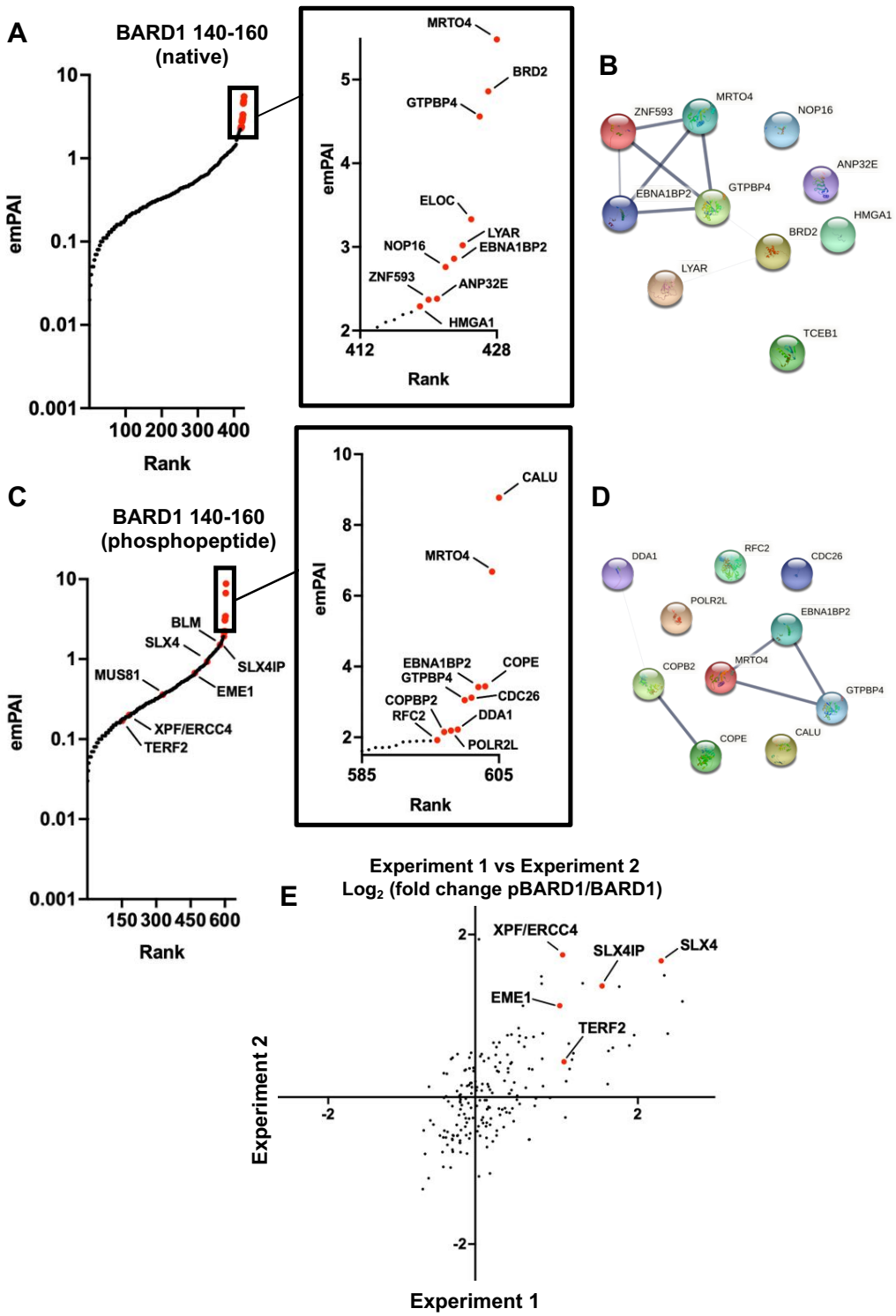


Fig. 3.8. Identification of candidate SLiM-based interactions between BARD1 and the SMX tri-nuclease complex/SLX4IP.

A) Plot of LC-MS/MS hits from peptide pulldowns of the native peptide corresponding to BARD1 140-160, ranked by emPAI score. The top ten ranked high-confidence candidate interacting proteins are highlighted in red in the zoomed in panel. B) STRING network of interactions between the top ten high-confidence candidate interacting proteins (Szklarczyk et al., 2021). C) Plot of LC-MS/MS hits from peptide pulldowns of the phosphopeptide corresponding to BARD1 140-160, ranked by emPAI score. The top ten ranked high-confidence candidate interacting proteins are highlighted in red in the zoomed in panel. Also highlighted in red are members SMX tri-nuclease complex, SLX4IP and BLM. D) STRING network of interactions between the top ten high-confidence candidate interacting proteins (Szklarczyk et al., 2021). E) SLX4 and its binding partner, SLX4IP, were enriched in a BARD1 140-160 peptide pulldown when the peptide was phosphorylated. SLX4, SLX4IP and other members of the SMX complex are highlighted in red. emPAI scores shown are the average of two independent repeats.

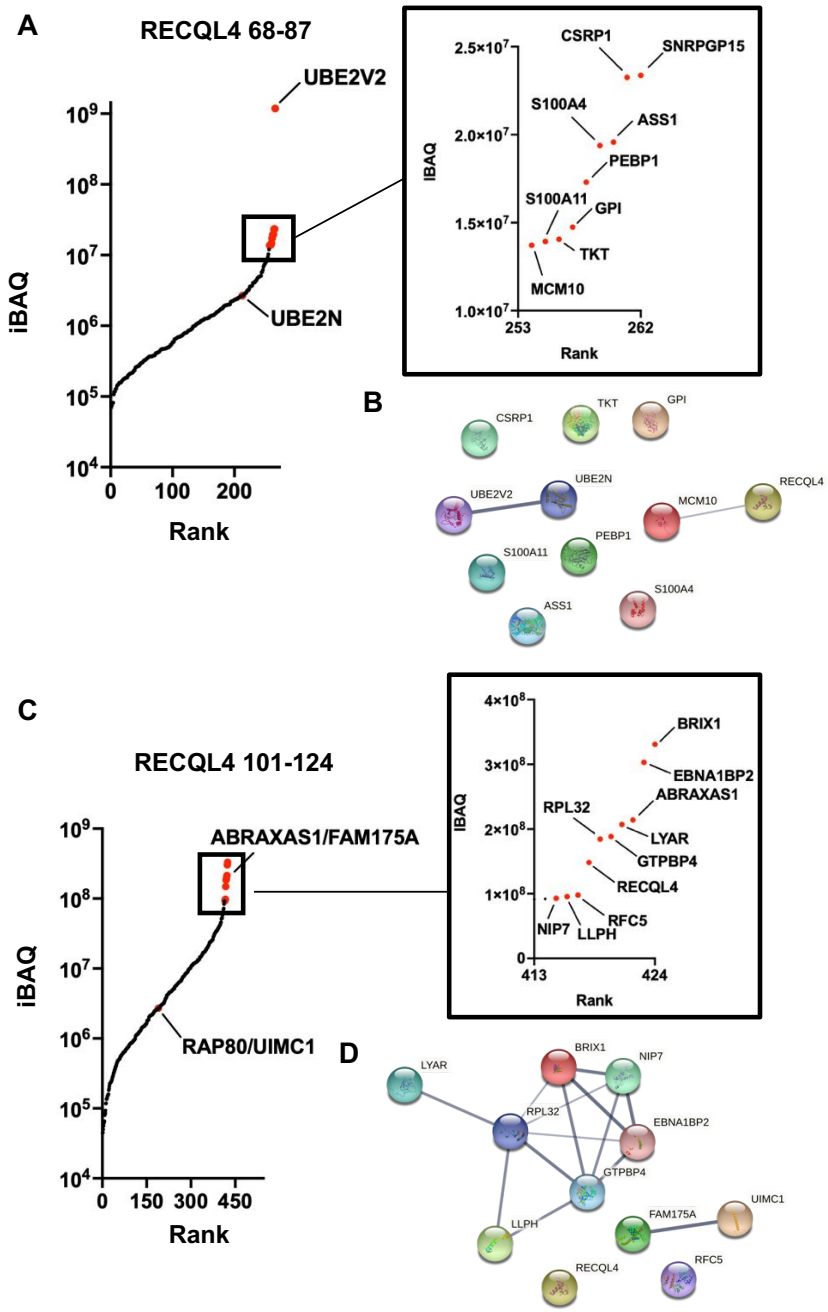


Fig. 3.9. Identification of candidate interactions of putative SLiMs from RECQL4.
A) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to RECQL4 68-87, ranked by iBAQ score. The E2 UBE2V2 (highlighted in red) was the highest candidate interacting protein. Also highlighted in red is UBE2N. The next nine ranked high-confidence candidate interacting proteins are highlighted in red in the zoomed in panel. B) STRING network of interactions between the top ten high-confidence candidate interacting proteins of the peptide corresponding to RECQL4 68-87 (Szklarczyk et al., 2021). C) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to RECQL4 101-124, ranked by iBAQ score. ABRAXAS1 (highlighted in red) was one of the highest candidate interacting protein. Also highlighted in red is RAP80. The top ten high-confidence candidate interacting proteins are highlighted in red in the zoomed in panel. D) STRING network of interactions between the top ten high-confidence candidate interacting proteins of the peptide corresponding to RECQL4 101-124 (Szklarczyk et al., 2021).

The E2 UBE2V2 was the highest candidate interacting protein of the peptide corresponding to RECQL4 68-87 (**Fig 3.9A, B**). UBE2V2 forms a heterodimer with UBE2N and catalyzes the synthesis of K63-linked polyubiquitin chains (Eddins et al., 2006; Stewart et al., 2016). UBE2N was also one of the high-confidence candidate interacting proteins. In addition, MCM10 was another of the top ten high-confidence candidate interacting proteins. This is consistent with a previous report, which has mapped the interaction between RECQL4 and MCM10 to this region in RECQL4 (Kliszczak et al., 2015). ABRAXAS/FAM175A was one of the highest candidate interacting proteins for a peptide corresponding to RECQL4 101-124 (**Fig 3.9C-D**). BRCA1, an interacting partner of ABRAXAS was not one of the high-confidence candidate interacting proteins, but RAP80/UIMC1, another interacting partner of ABRAXAS, was (Wang et al., 2007). RECQL4 was also one of the top ten high-confidence candidate interacting proteins, so this SLiM could promote RECQL4 oligomerisation.

TOPBP1 was highest candidate interacting protein for a peptide corresponding to MDC1 184-204, when phosphorylated on S196 (**Fig 3.16B**). CIP2A was highest candidate interacting protein for a peptide corresponding to TOPBP1 774-798, when phosphorylated on T779 (**Fig 3.17B**). Lastly, CtIP was one of the top ten candidate interactors for a peptide corresponding to AUNIP 340-357, when phosphorylated on S347 (**Fig 3.20C**). These interactions are validated in the following sections.

The remaining 45% (n = 13) of peptide pulldowns with high-confidence candidate interacting proteins did not contain DDR proteins in the top ten candidate interacting proteins. These were for peptides corresponding to BRCA1 455-467 (**Fig. 3.10A,B**), BRCA1 501-527 (**Fig. 3.10C,D**), BRCA2 2401-2414 (**Fig. 3.11A,B**), FANCM 1126-1143 (**Fig. 3.12A,B,C**), POLQ 1434-1454 (**Fig. 3.13A,B**), POLQ 1477-1497 (**Fig. 3.13C,D**), POLQ 1631-1652 (**Fig 3.13E,F**), RECQL4 171-197 (**Fig. 3.14A,B**), SHLD2 265-283 (**Fig. 3.14C,D**), SLX4 603-628

(**Fig. 3.15A,B**), TOPBP1 486-503 (**Fig. 3.15C,D**), XRCC4 255-278 (**Fig. 3.15E,F**), XRCC4 318-334 (**Fig. 3.15G,H**).

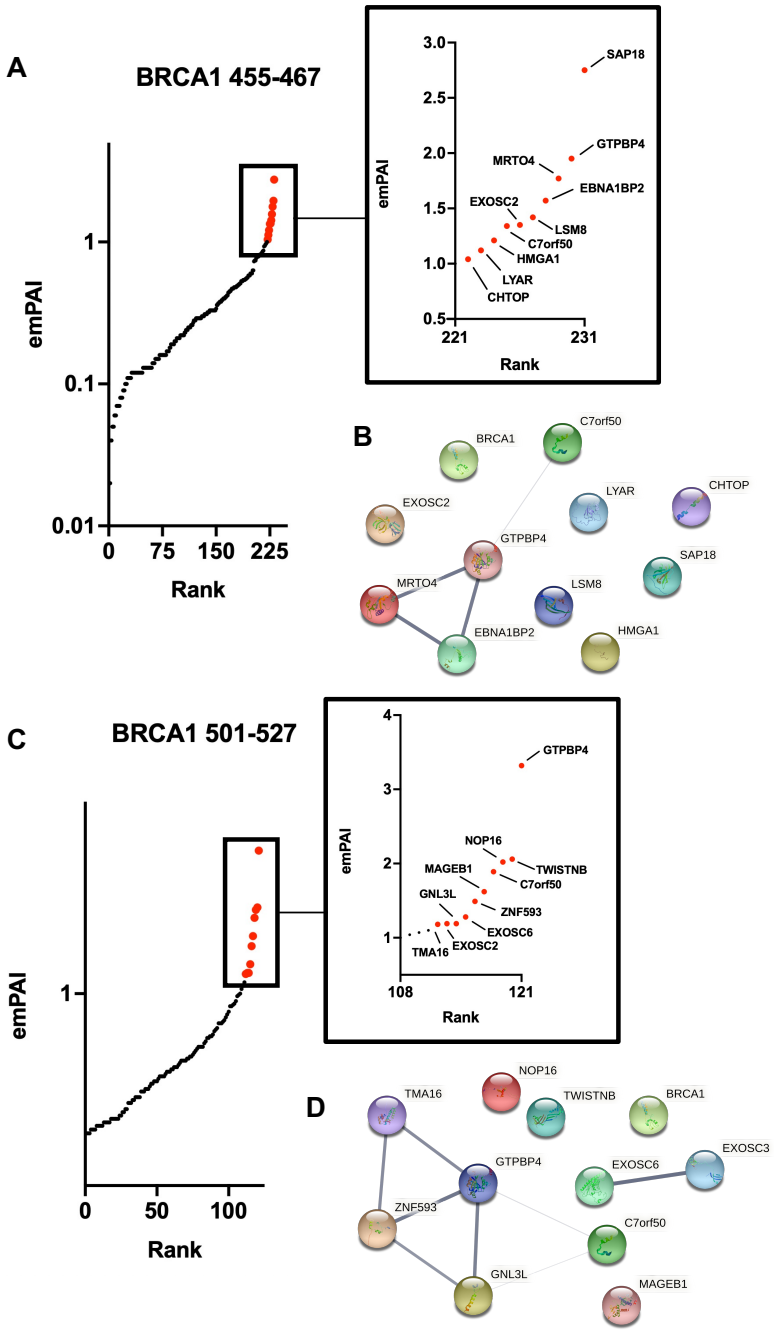


Fig. 3.10. Identification of candidate interactions of putative SLiMs from BRCA1.
A) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to BRCA1 455-467, ranked by emPAI score. The top ten high-confidence candidate interacting proteins are highlighted in red in the zoomed in panel. B) STRING network of interactions between the top ten high-confidence candidate interacting proteins of the peptide corresponding to BRCA1 455-467 (Szklarczyk et al., 2021). C) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to BRCA1 501-527, ranked by emPAI score. The top ten high-confidence candidate interacting proteins are highlighted in red in the zoomed in panel. D) STRING network of interactions between the top ten high-confidence candidate interacting proteins of the peptide corresponding to BRCA1 501-527 (Szklarczyk et al., 2021). emPAI scores shown are the average of two independent repeats.

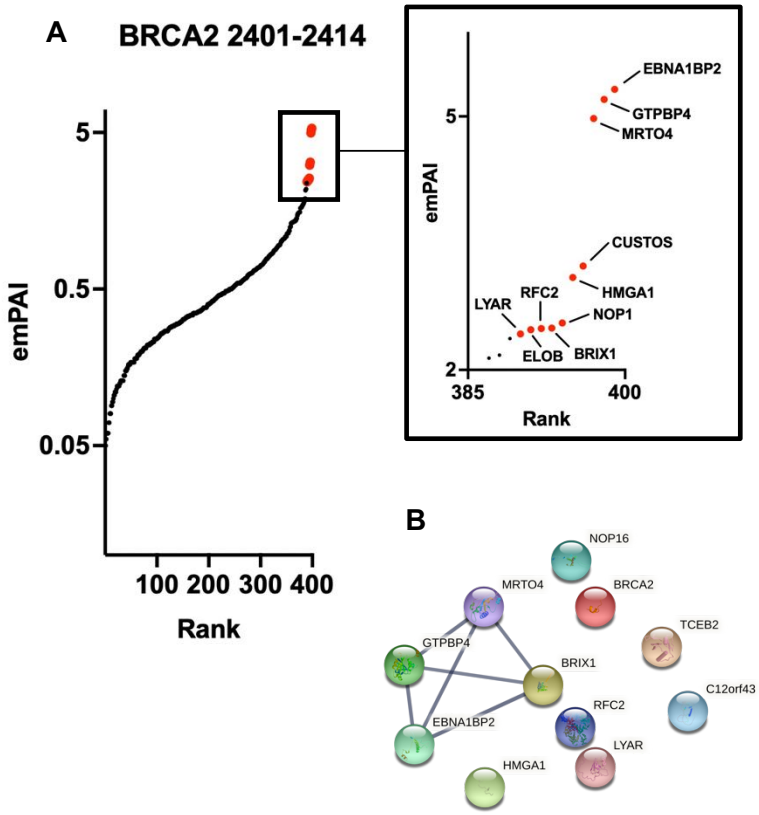


Fig. 3.11. Identification of candidate interactions of a putative SLiM from BRCA2
A) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to BRCA2 2401-2414, ranked by emPAI score. The top ten high-confidence candidate interacting proteins are highlighted in red in the zoomed in panel. emPAI scores shown are the average of two independent repeats. B) STRING network of interactions between the top ten high-confidence candidate interacting proteins of the peptide corresponding to BRCA2 2401-2414 (Szklarczyk et al., 2021).

STRING analysis highlighted that some of these peptide pulldowns contained protein complexes. Peptide pulldowns corresponding to BRCA1 455-467 (**Fig. 3.10A,B**), BRCA1 501-527 (**Fig. 3.10C,D**), BRCA2 2401-2414 (**Fig. 3.11A,B**) and XRCC4 255-278 (**Fig. 3.15E,F**) contained networks of ribosomal biogenesis factors. The peptide pulldown corresponding to BRCA1 501-527 (**Fig. 3.10C,D**) also contained members of the RNA exosome complex. Peptide pulldowns corresponding to FANCM 1126-1145 (**Fig. 3.12A,B,C**) contained proteins that promote the formation of clatherin-coated vesicles, the peptide pulldown corresponding to POLQ 1631-1652 (**Fig 3.13E,F**) contained pre-mRNA processing factors and the peptide pulldown corresponding to SLX4 603-628 (**Fig. 3.15A,B**) contained

complex members of the mitochondrial inner membrane complex. It is possible that findings represent biologically relevant SLiM-based interactions, but it would be difficult to hypothesise how they may regulate the DDR.

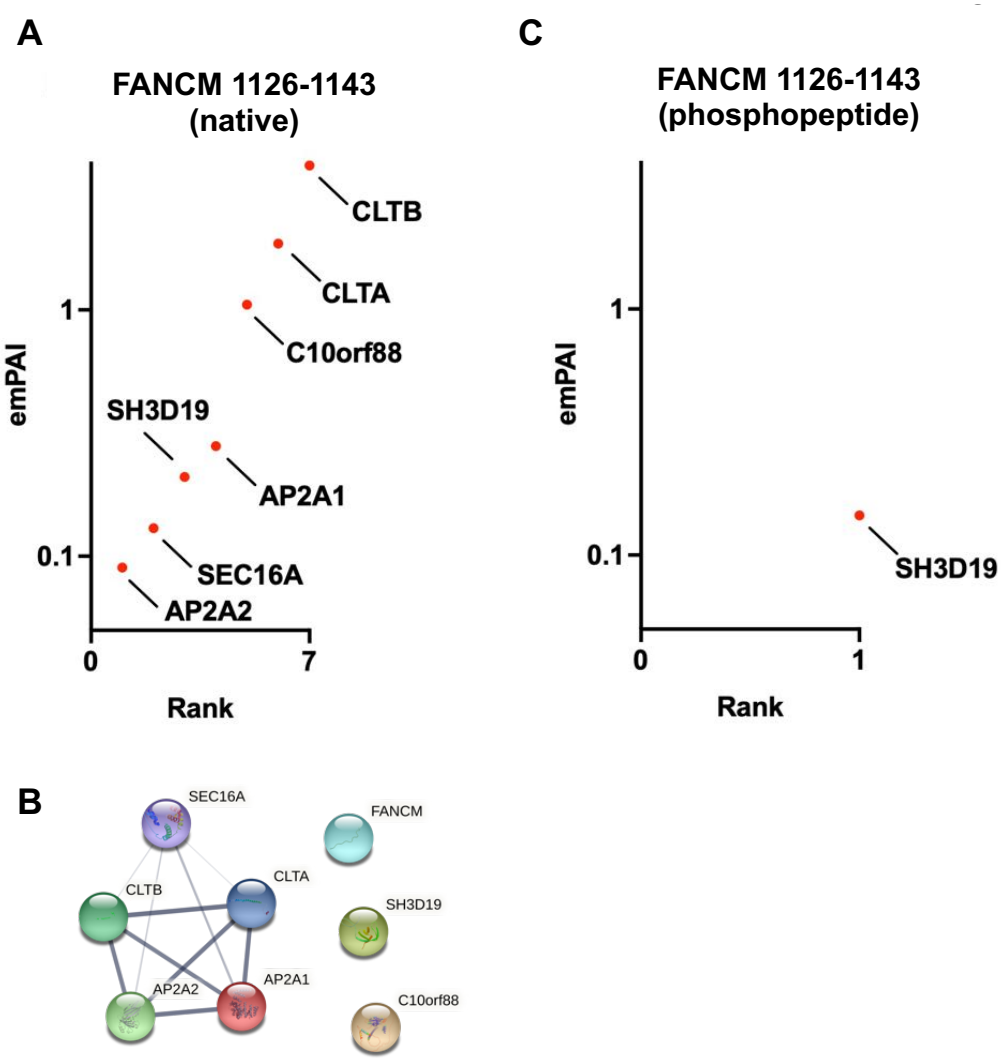


Fig. 3.12. Identification of candidate interactions of a putative SLiM from FANCM
A) Plot of LC-MS/MS hits from peptide pulldowns of the native peptide corresponding to FANCM 1126-1143, ranked by emPAI score. The top ten high-confidence candidate interacting proteins are highlighted in red. B) STRING network of interactions between the top ten high-confidence candidate interacting proteins of the native peptide corresponding to FANCM 1126-1143 (Szklarczyk et al., 2021).C) Plot of the LC-MS/MS hit from peptide pulldowns of the phosphopeptide corresponding to FANCM 1126-1143. SH3D19 is highlighted in red. emPAI scores shown are the average of two independent repeats.

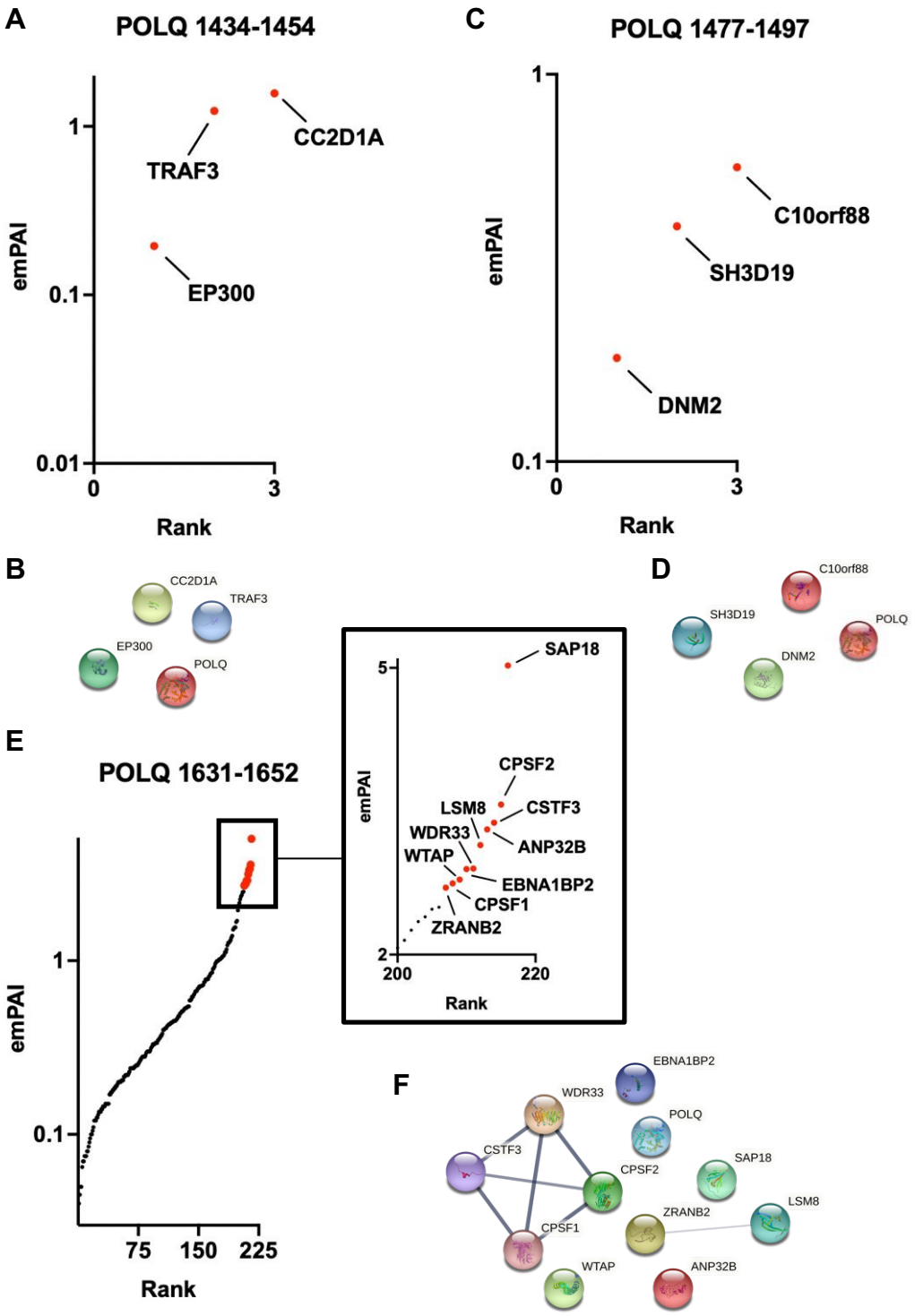


Fig. 3.13. Identification of candidate interactions of putative SLiMs from POLQ
A) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to POLQ 1434-1454, ranked by emPAI score. The high-confidence candidate interacting proteins are highlighted in red.B) STRING network of interactions between high-confidence candidate interacting proteins of the peptide corresponding to POLQ 1434-1454 (Szklarczyk et al., 2021). C) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to POLQ 1477-1497, ranked by emPAI score. The high-confidence candidate interacting proteins are highlighted in red. D) STRING network of interactions between high-confidence candidate interacting proteins of the peptide corresponding to POLQ 1477-1497 (Szklarczyk et al., 2021). E) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to POLQ 1631-1652, ranked by emPAI score. The top ten high-confidence candidate interacting proteins are highlighted in red in the zoomed in panel. F) STRING network of interactions between the top ten high-confidence candidate interacting proteins of the peptide corresponding to POLQ 1477-1497 (Szklarczyk et al., 2021). emPAI scores shown are the average of two independent repeats.

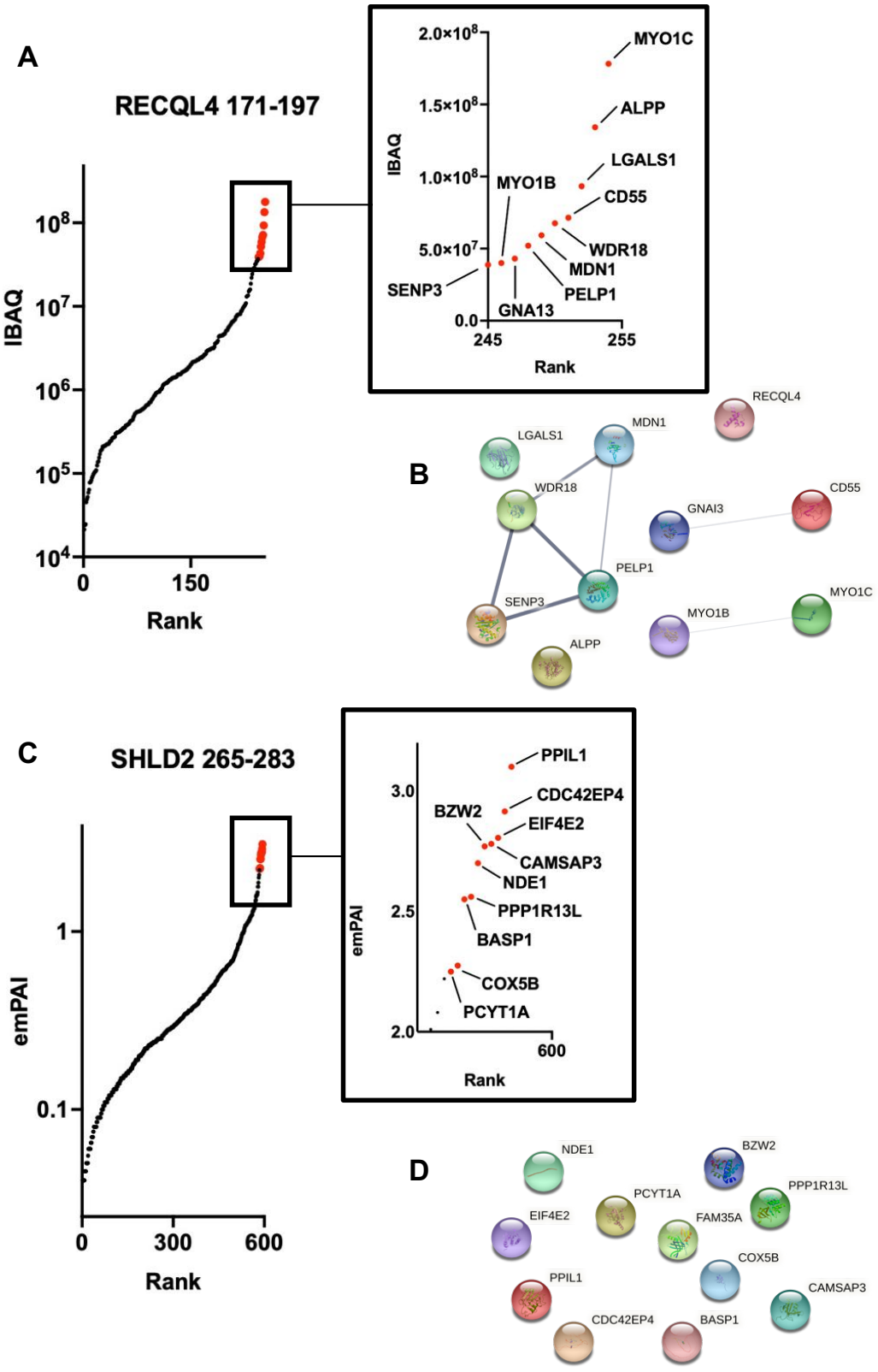


Fig. 3.14. Identification of candidate interactions of putative SLiMs from RECQL4 and SHLD2
A) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to RECQL4 171-197, ranked by iBAQ score. The top ten high-confidence candidate interacting proteins are highlighted in red in the zoomed in panel. B) STRING network of interactions between high-confidence candidate interacting proteins of the peptide corresponding to RECQL4 171-197 (Szklarczyk et al., 2021). C) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to SHLD2 265-285, ranked by emPAI score. The top ten high-confidence candidate interacting proteins are highlighted in red in the zoomed in panel. emPAI scores shown are the average of two independent repeats. D) STRING network of interactions between the top high-confidence candidate interacting proteins of the peptide corresponding to SHLD2 265-285 (Szklarczyk et al., 2021).

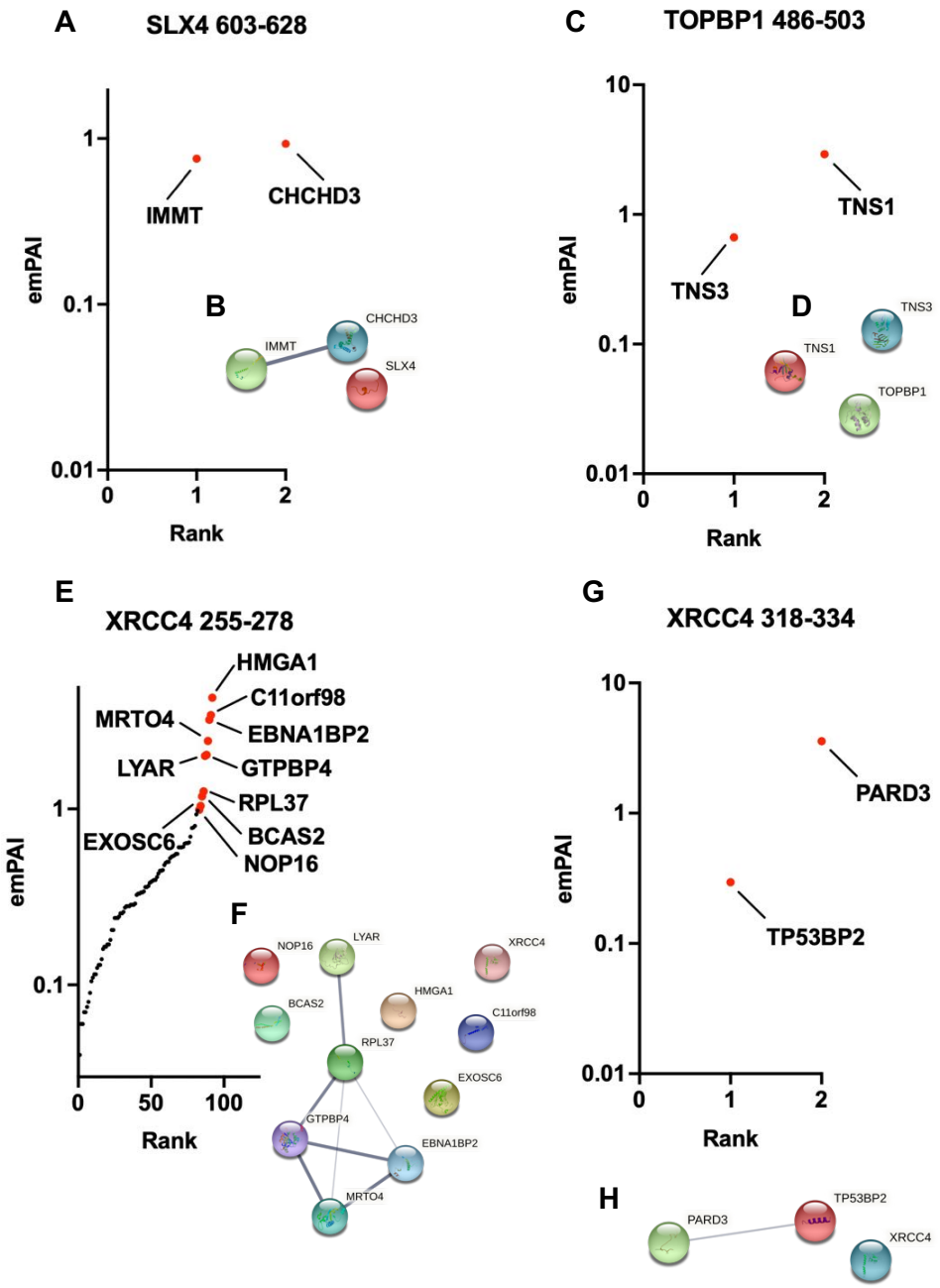


Fig. 3.15. Identification of candidate interactions of putative SLiMs from SLX4, TOPBP1 and XRCC4.

A) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to SLX4 603-628, ranked by emPAI score. The high-confidence candidate interacting proteins are highlighted in red. B) STRING network of interactions between high-confidence candidate interacting proteins of the peptide corresponding to SLX4 603-628 (Szklarczyk et al., 2021). C) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to TOPBP1 486-503, ranked by emPAI score. The high-confidence candidate interacting proteins are highlighted in red. D) STRING network of interactions between high-confidence candidate interacting proteins of the peptide corresponding to TOPBP1 486-503 (Szklarczyk et al., 2021). E) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to XRCC4 255-278, ranked by emPAI score. The top ten high-confidence candidate interacting proteins are highlighted in red. F) STRING network of interactions between the top ten high-confidence candidate interacting proteins of the peptide corresponding to XRCC4 255-278 (Szklarczyk et al., 2021).G) Plot of LC-MS/MS hits from peptide pulldowns of the peptide corresponding to XRCC4 318-334, ranked by emPAI score. The high-confidence candidate interacting proteins are highlighted in red. H) STRING network of interactions between the high-confidence candidate interacting proteins of the peptide corresponding to XRCC4 318-334 (Szklarczyk et al., 2021). emPAI scores shown are the average of two independent repeats.

The peptide corresponding to BRCA1 455-467 (**Fig. 3.10A**) also interacted with an uncharacterised proteins, C7orf50. However, C7orf50 was also enriched in the BRCA1 501-527 peptide pulldown (**Fig. 3.10C**). It is possible this could represent a biologically relevant SLiM-based interaction with an uncharacterised proteins, but it is more likely that this interactions are indirect and co-purifying with other proteins.

3.3.4 Validation of Candidate SLiM-Based Interactions

In the previous section, high-confidence candidate interacting proteins were identified for 81% of the SLiMs, but this does not prove these interacting proteins are biologically relevant. To suggest this, the SLiM-based interaction must be validated in the context of the full-length protein and further phenotypic experiments must be conducted to determine the function of the interaction. There was a need to prioritise which of the SLiM-based interactions would be validated and investigated further. There are several indicators that could suggest a specific protein was a strong candidate for being the biologically relevant interacting protein: firstly, with the background proteome eliminated, biologically relevant proteins are likely to be the most abundant; second, if the protein has a reported role in the DDR; thirdly, if multiple members of the complex were present; and fourthly, if the localisation of the protein was nuclear rather than cytoplasmic. However, it important to note that there will be exceptions to the above. For example, biologically relevant SLiM-based interactions may exist between nuclear and cytoplasmic proteins, in mitosis after the nuclear envelope has broken down.

Due to time constraints, not of all the candidate interactions described in section 3.3.3 could be validated but will form future projects for members of Blackford lab. However, a subset of these interactions were validated and are presented in the following sections: an N-terminal SLiM in MDC1 that interacts with TOPBP1 to promote chromosomal stability in mitosis (section 3.3.5), a SLiM in TOPBP1 that promotes the interaction between TOPBP1 and

the C-terminus of CIP2A (section 3.3.6), two C-terminal SLiMs in the C-terminus of AUNIP that interact with CtIP (section 3.3.7), and lastly three SLiMs in the BLM helicase complex that interact with RPA (section 3.3.8).

In summary, the bioinformatics and proteomics pipeline has identified 16 novel candidate interactions based on SLiMs and DDR. The fact that SLiMs synthesised as peptides of 25-30 amino acids interact with DNA repair factors when incubated in HeLa nuclear extract suggests that the pipeline is successful.

3.3.5 MDC1 Interacts with TOPBP1 to Promote Chromosomal Stability during Mitosis

MDC1 is a scaffold protein recruited to DSB sites, forming a platform for recruitment of DDR factors to mediate DNA repair (Stucki et al., 2005). MDC1 contains a N-terminal FHA domain, phosphorylated on T4 in an ATM-dependent manner,which promotes MDC1 dimerization and retention on damaged chromatin (Jungmichel et al., 2012; Liu et al., 2012). At its C-terminus, MDC1 contains a BRCT domain that interacts with γH2AX (Stucki et al., 2005). MDC1 contains conserved SDTD repeats constitutively phosphorylated by CK2 and are recognised by the FHA-BRCT region of the MRE11-RAD50-NBS1 (MRN) complex member, NBS1 (Chapman and Jackson, 2008; Melander et al., 2008; Spycher et al., 2008). This promotes MDC1-MRN retention on damaged chromatin. MDC1 also contains a cluster of TQXF motifs that are phosphorylated by ATM. These SLiMs are recognised by the FHA domain of the E3 ubiquitin ligase RNF8 (Huen et al., 2007; Kolas et al., 2007; Mailand et al., 2007). Recent work from the Jackson lab has demonstrated that the PST-repeats of MDC1 can promote chromatin association of MDC1 independent of H2AX by interacting directly with nucleosomes (Salguero et al., 2019).

Aiming to identify unknown protein-protein interactors of MDC1, amino acid sequence alignments of MDC1 from representative vertebrate orthologues were generated (section 8.1.9). Using the prediction criteria described previously, a potential novel SLiM was identified in MDC1. This SLiM is centred around S196 and is highly conserved in MDC1 orthologues (**Fig. 3.16A**). Furthermore, S196 has been shown to be phosphorylated in human cells in high-throughput mass spectrometry experiments (Yi *et al,* 2014). To identify whether this SLiM is an interaction site for an unidentified DDR factor, biotinylated peptides were synthesised with and without a phospho-serine at S196. These peptides were used as bait in HeLa nuclear extract. When the high-confidence candidate interacting proteins enriched in the phospho-MDC1 peptide pulldown were examined, TOPBP1 was identified as the highest confidence candidate interacting protein (**Fig. 3.16B**). TOPBP1 is a versatile scaffold protein functioning in various cellular pathways to promote genome stability, including DSB repair and ATR activation (Wardlaw et al., 2014).

Given the role of TOPBP1 in the DDR, this interaction was pursued further. Western blotting confirmed TOPBP1 was specifically pulled down by the phospho-MDC1 peptide (**Fig. 3.16C**). NBS1, which interacts with the SDTD repeats downstream of S196 in the full-length protein (Chapman and Jackson, 2008; Melander et al., 2008; Spycher et al., 2008), did not interact with either peptide. Next, it was investigated whether S196 is required for MDC1-TOPBP1 interaction in the context of full-length MDC1. S196 was mutated to alanine (S196A) in constructs expressing HA-tagged wild-type (WT) MDC1 and these constructs were transfected into 293FT cells. Immunoprecipitation (IP) with HA-affinity beads showed that the MDC1-S196A mutation abrogated TOPBP1 binding (**Fig. 3.16D**). Conversely, the interaction of MDC1 with γH2AX, known to bind to distinct regions of MDC1 (the C-terminal BRCT domains) was not affected (Stucki et al., 2005). This demonstrated that

mutating S196 to alanine in the context of full-length MDC1 does not cause general misfolding of the protein.

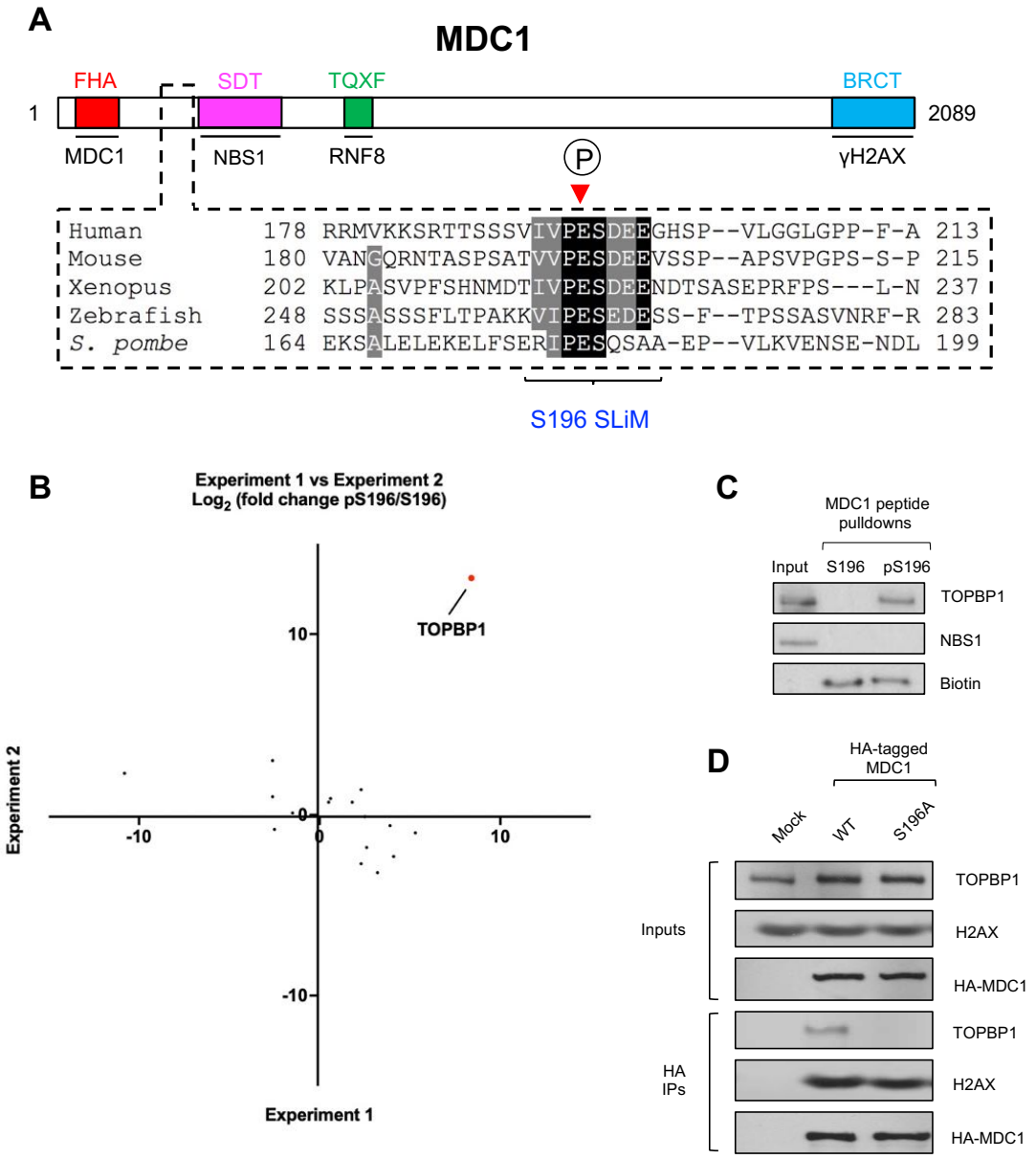


Fig 3.16. MDC1 interacts with TOPBP1 to promote chromosomal stability during mitosis.
A) Schematic showing the layout of conserved domains and SLiMs in MDC1. Names of the known MDC1 binding partners NBS1, RNF8 and H2AX are shown below the regions of MDC1 they interact with. The FHA domain promotes MDC1 dimerization, hence its binding partner is MDC1. A region of the sequence alignment for MDC1, centered around the S196 SLiM, is shown. P, phosphorylation. B) Identification of TOPBP1 as a specific interactor for the MDC1-S196 phosphopeptide by LC–MS/MS. TOPBP1 is highlighted in red. C) Peptide pulldowns from HeLa nuclear extracts using peptides corresponding to residues surrounding MDC1-S196, verifying that TOPBP1 is a specific interactor for the MDC1-S196 phosphopeptide. D) HA-IPs from 293FT cells transfected with the indicated HA-tagged MDC1 variants, showing how S196A abrogates TOPBP1 binding, whilst NBS1 and H2AX binding remains intact.

Further work sought to understand the biological significance of the MDC1-TOPBP1 interaction. Characterisation of this interaction revealed that MDC1 interacts with TOPBP1 to promote chromosomal stability during mitosis (Leimbacher et al., 2019). TOPBP1 was found to form filamentous structures that are capable of bridging MDC1 foci in mitotic cells, suggesting that MDC1-TOPBP1 complexes tether DSBs until repair is reactivated in the following G1 phase.

3.3.6 TOPBP1 Interacts with the C-terminus of CIP2A

TOPBP1 is a versatile scaffold protein functioning in various cellular pathways to promote genome stability (Wardlaw et al., 2014). Human TOPBP1 contains nine BRCT domains and an ATR activation domain (AAD; **Fig 3.17A**). Four BRCT domains contain the lysine residues required for phosphopeptide binding (BRCT 1, 2, 5 and 7). BRCT0-2 interact with phosphorylated RAD9, Treslin or MDC1 (demonstrated in the previous section; (Delacroix et al., 2007; Kumagai et al., 2010; Lees-Miller et al., 1990; Leimbacher et al., 2019), BRCT4 and 5 interact with phosphorylated 53BP1 or BLM (Blackford et al., 2015; Cescutti et al., 2010) and BRCT7 and 8 interact with phosphorylated FANCJ (Gong et al., 2010). Aiming to identify unknown protein-protein interactors of TOPBP1, amino acid sequence alignments of TOPBP1 from representative vertebrate orthologues were generated (section 8.1.17) Using the prediction criteria described previously, a potential novel SLiM was identified in TOPBP1. Centred around T779, the SLiM is highly conserved in TOPBP1 orthologues (**Fig. 3.17A**) and T779 has been shown to be phosphorylated in 13 studies recorded on PhosphoSitePlus (Dephoure et al., 2008).

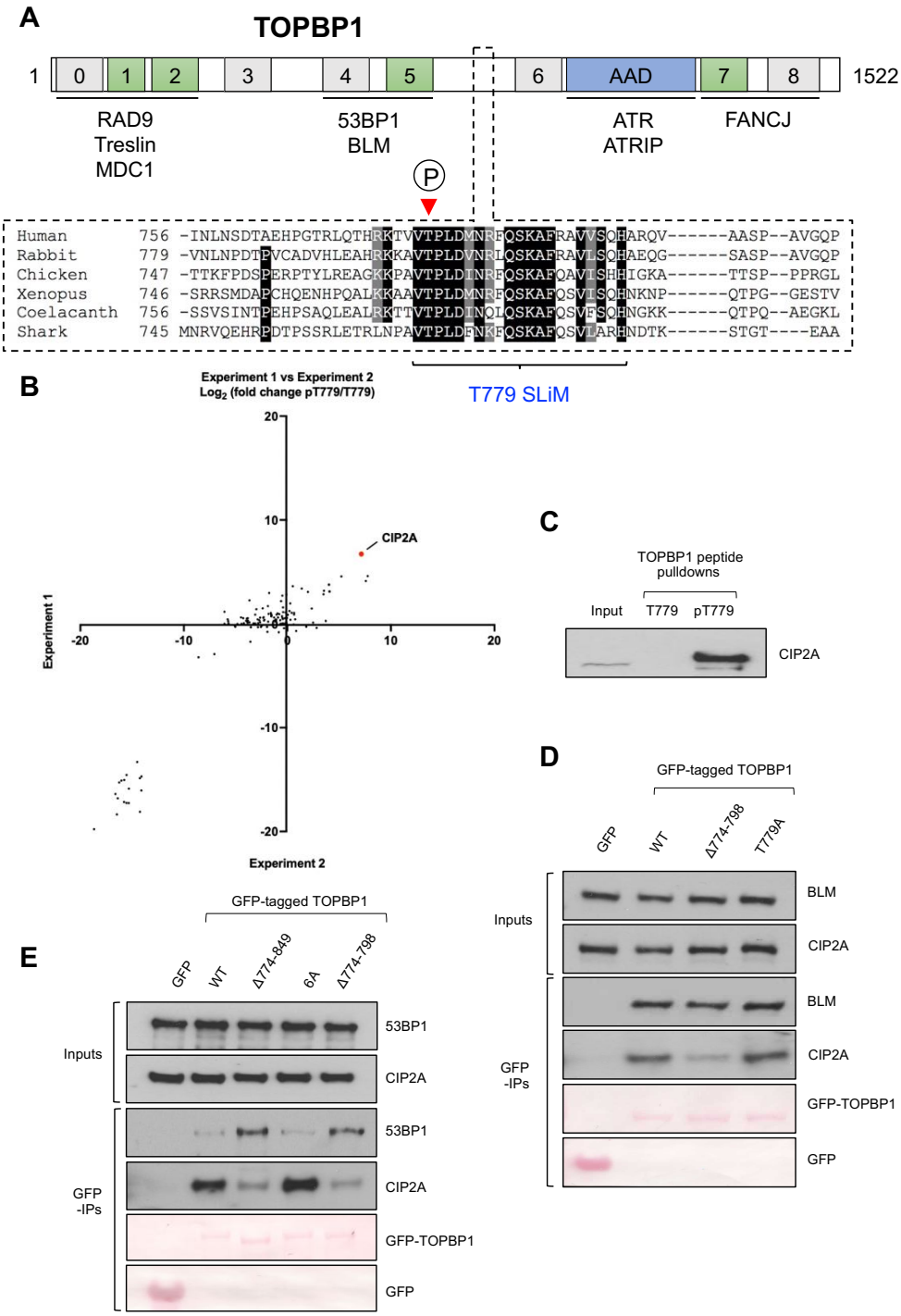


Fig. 3.17. TOPBP1 interacts with CIP2A.
A) Schematic showing the layout of conserved domains and SLiMs in TOPBP1. Numbered boxes represent BRCT domains, with phosphopeptide-binding domains in green and domains lacking phosphopeptide-binding activity in grey. Names of known TOPBP1 binding partners are shown below the domains they interact with. AAD, ATR-activation domain. A region of the sequence alignment for TOPBP1, centered around the T779 SLiM, is shown. P, phosphorylation. B) Identification of CIP2A as a specific interactor for the TOPBP1-T779 phosphopeptide by LC–MS/MS. CIP2A is highlighted in red. C) Peptide pulldowns from HeLa nuclear extracts using peptides corresponding to residues surrounding TOPBP1-T779, verifying that CIP2A is a specific interactor for the TOPBP1-T779 phosphopeptide. D) GFP-IPs from 293FT cells transfected with the indicated GFP-tagged TOPBP1 variants, showing how T779A does not abrogate CIP2A binding, whilst deletion of the entire SLiM (Δ774-798) reduced but did not abrogate CIP2A-binding. E) GFP-IPs from 293FT cells transfected with the indicated GFP-tagged TOPBP1 variants, showing how 6A does not abrogate CIP2A binding, whilst Δ774-849 reduced but did not abrogate CIP2A binding.

To identify interacting proteins for this SLiM, biotinylated peptides were synthesised with and without a phospho-threonine at the T779 position. These peptides were used as bait in HeLa nuclear extract, and the proteins enriched in the phospho-TOPBP1 peptide pulldown were examined. The protein CIP2A was the highest confidence candidate interacting protein in two independent repeats (**Fig 3.17B**).

CIP2A was originally described as an inhibitor of protein phosphatase 2A (PP2A; (Junttila et al., 2007)). CIP2A was shown to physically interact with the oncogenic transcription factor c-Myc, inhibiting PP2A activity toward c-Myc Ser62 and preventing c-Myc proteolytic degradation. CIP2A is overexpressed in various cancers, including head and neck squamous cell carcinoma (HNSCC), colon cancer and basal-like breast cancer (Junttila et al., 2007). Of note, recent genotoxin sensitivity profiling by CRISPR-Cas9 drop-out screens demonstrated that CIP2A deficiency renders cells highly sensitivity to ATR inhibitors or drugs that induce DSBs in actively replicating cells, such as topoisomerase I or II inhibitors (Hustedt et al., 2019; Olivieri et al., 2019). Lastly, two recent reports also identified the interaction between TOPBP1 and CIP2A in HEK293 cells, verifying these findings (Adam et al., 2021b; Laine et al., 2021b).

Given several studies suggest that CIP2A may have roles in the DDR, the interaction between CIP2A and TOPBP1 was pursued further. Western blotting confirmed that CIP2A was specifically pulled down by the phospho-TOPBP1 peptide (**Fig 3.17C**). Next, it was investigated whether T779 was required for the TOPBP1-CIP2A interaction in the context of full-length TOPBP1. To do this, T779 was mutated to alanine (T779A) or the entire SLiM was deleted (Δ774-798) in constructs expressing GFP-tagged WT TOPBP1 and then transfected into 293FT cells. GFP IPs from lysates harvested from these cells demonstrated that loss of the SLiM reduced but did not abrogate CIP2A-binding (**Fig. 3.17D**). Conversely, the interaction of TOPBP1 with BLM, which binds to a distinct region of TOPBP1 (BRCT5) was not affected (Blackford et al., 2015). This demonstrated that deletion of the SLiM does not cause general

misfolding of TOPBP1. Surprisingly, in contrast to the mass spectrometry data, GFP IPs showed that the TOPBP1-T779A mutation in the context of the full-length protein does not affect the TOPBP1-CIP2A interaction (**Fig. 3.17D**). This demonstrated that the peptide binding data does fully recapitulate the TOPBP1-CIP2A interaction in the context of full-length TOPBP1. One potential explanation for this is that CIP2A could interact with another region of TOPBP1, rendering the phosphorylation site redundant in the context of the full-length protein.

Further analysis of the TOPBP1 sequence alignment revealed between the T779 SLiM and BRCT6, there are 5 further S/TP sites, 4 of which are highly conserved across vertebrate evolution (section 8.1.17). Therefore, the possibility that the CIP2A interaction region spans all the way from the SLiM to just upstream of BRCT6 was considered. Accordingly, these 6 S/T residues were mutated to alanine (6A), or the entire region was deleted (Δ774-849) in constructs expressing GFP-tagged WT TOPBP1. The TOPBP1 constructs were then transfected into 293FT cells and GFP IPs from lysates harvested from these cells demonstrated that Δ774-849 reduced CIP2A-binding to the same extent as Δ774-798 (**Fig. 3.17E**). Again, surprisingly, GFP IPs showed that the 6A mutation in the context of the full-length protein did not affect the TOPBP1-CIP2A interaction. This suggested that these phosphorylated residues are not required for the TOPBP1-CIP2A interaction.

Interestingly, the interaction of TOPBP1 with 53BP1, which binds to another region of TOPBP1 (BRCT5) was enhanced with the mutants with a reduced interaction with TOPBP1 (Δ774-798 and Δ774-849), when compared to WT (Cescutti et al., 2010). This suggests that Δ774-849 and 6A do not cause general misfolding of TOPBP1, but also implies that CIP2A and 53BP1 may compete for binding to TOPBP1. The same pattern was not seen with the TOPBP1-BLM interaction (**Fig. 3.17D**), which also interacts with BRCT5, suggesting that this phenomenon is specific to 53BP1 (Blackford et al., 2015).

CIP2A is a 905 amino acid protein, roughly composed of two structurally distinct regions: an N-terminal armadillo repeat core and a C-terminal predicted coiled-coil region ((Wang et al., 2017); **Fig. 3.18A**). The N-terminal armadillo repeat core mediates homodimerization of CIP2A and PP2A binding. The crystal structure of the N-terminus (1-560) reveals a highly conserved region between residues 159 and 245 forming a positively charged surface (Wang et al., 2017). N230 is a highly conserved amino acid in the centre of the positively charged region and is critical for the interaction with PP2A (B56α), as N230E significantly inhibited the CIP2A interaction with PP2A (B56α). To map the TOPBP1 binding site in CIP2A, a series of Spot-mRuby3 CIP2A constructs were designed and synthesised **(Fig. 3.18A)**: WT, N230E, 1-571 (Armadillo repeat core) and 561-905 (coiled-coil domain). The CIP2A constructs were then transfected into 293FT cells, and the harvested lysates were incubated with biotinylated peptides derived from the T779 SLiM (with and without a phospho-threonine at the T779) for peptide pulldowns. Western blotting demonstrated the WT, N230E and 561-904 CIP2A constructs were pulled down by the phospho-TOPBP1 peptide, whereas the 1-571 CIP2A construct lacking the C-terminal coiled-coil domain was not (**Fig 3.18B**). In addition, the CIP2A constructs were transfected into 293FT cells and Spot-IPs from lysates harvested from these cells demonstrated that the WT, N230E and 561-904 CIP2A constructs could interact with TOPBP1 (**Fig 3.18C**). Again, the 1-571 CIP2A construct lacking the C-terminal coiled-coil domain could not. Together, these findings suggest that the C-terminal coiled-coil domain of CIP2A interacts with TOPBP1.

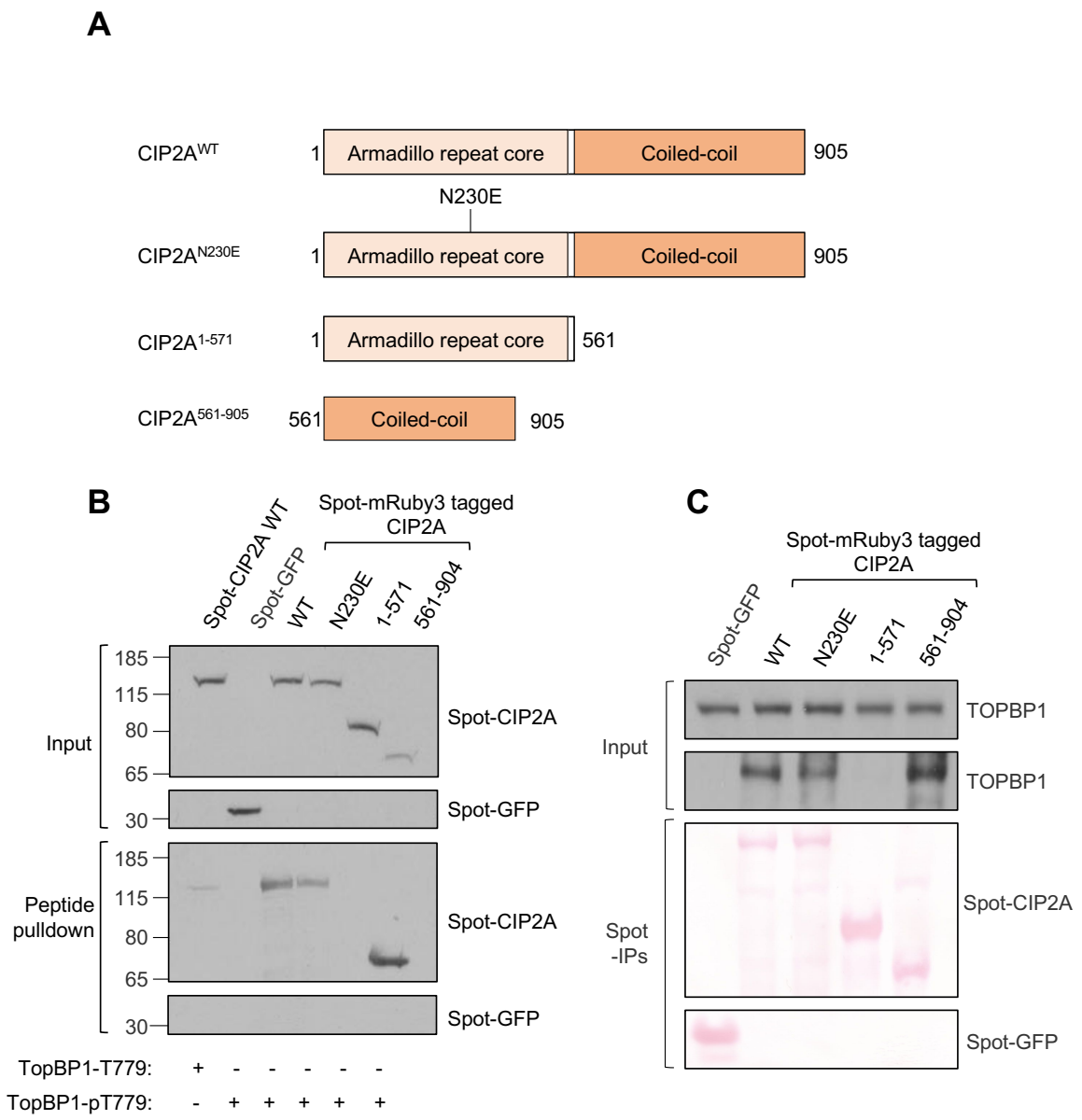


Fig. 3.18. The C-terminus of CIP2A mediates the interaction between CIP2A and TOPBP1
A) Schematic of Spot-mRuby tagged CIP2A constructs used in this study: WT, N230E, 1-571 (Armadillo repeat core) and 561-905 (coiled-coil domain). N230 is a highly conserved amino acid in the center of the positively charged regions and is critical for the interaction with PP2A. B) Peptide pulldowns from lysates of 239FT cells transfected with the indicated CIP2A constructs, using peptides derived from the T779 SLiM, revealed the WT, N230E and 561-904 CIP2A constructs interact phospho-TOPBP1 peptide, whereas the construct lacking the C-terminal coiled-coil domain does not. C) Spot-IPs from 293FT cells transfected with the indicated Spot-tagged CIP2A variants verified that the construct lacking the C-terminal coiled-coil domain cannot interact with TOPBP1.

Lastly, to provide further evidence that these proteins interact *in vivo,* U2OS cells stably expressing GFP-TOPBP1 were transfected with various Spot-mRuby3-CIP2A variants (WT, 1-571 and 561-904). Given that CIP2A is a cytoplasmic protein and TOPBP1 is a nuclear protein, this suggested that the interaction between CIP2A and TOPBP1 is most likely to occur

in mitosis. Accordingly, the cells were synchronised in prometaphase using nocodazole and irradiated. Subsequent high-resolution Airyscan imaging of mRuby3-CIP2A and GFP-TOPBP1 foci was performed in mitotic U2OS cells (**Fig 3.19**). Consistent with the mapping data, CIP2A^{WT} and CIP2A$^{561-904}$ co-localised with TOPBP1, whilst CIP2A$^{1-571}$ did not.

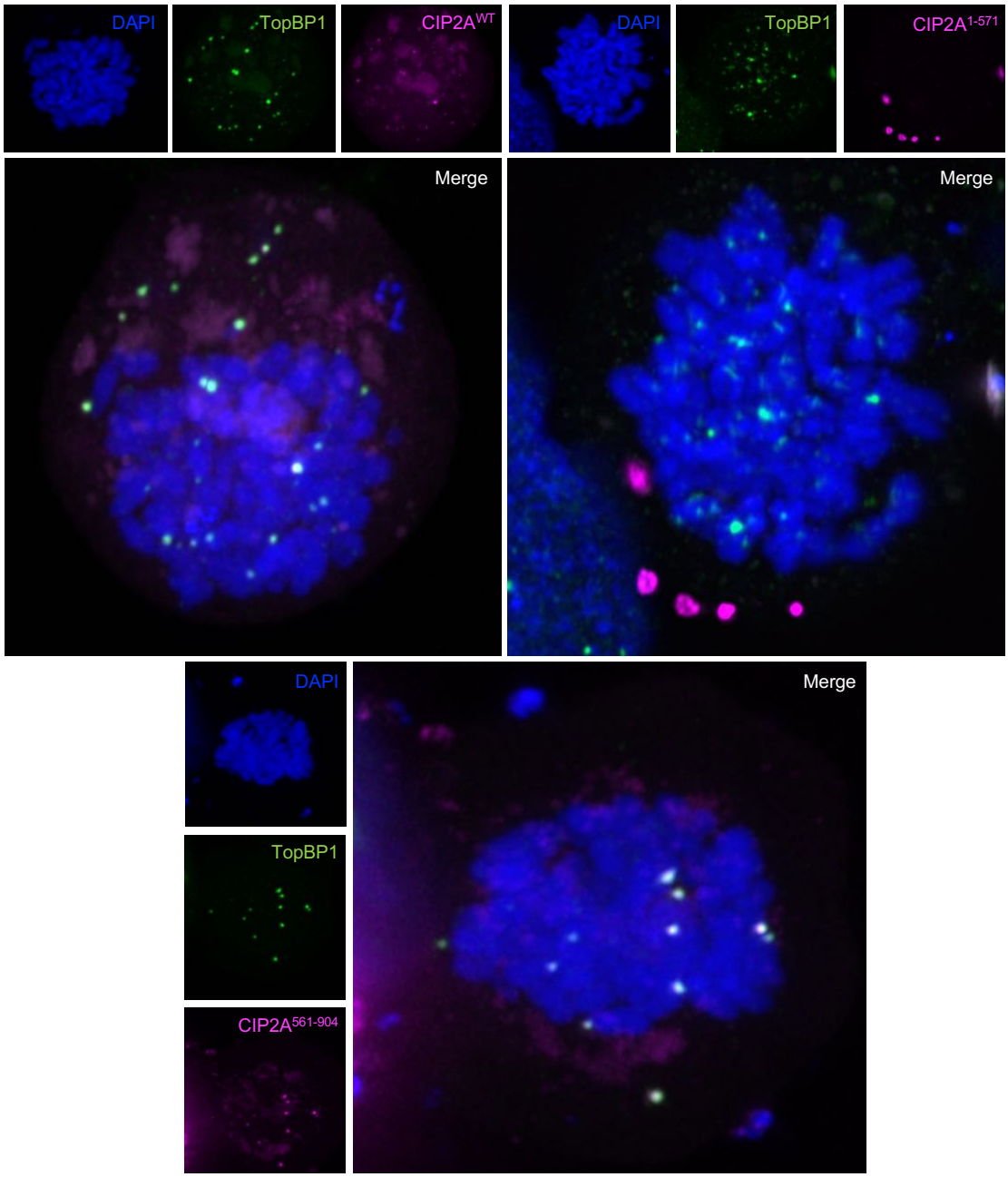


Fig. 3.19. High-resolution airyscan imaging of mRuby3-CIP2A and GFP-TOPBP1 foci in mitosis. Samples fixed and stained 1 h after 1 Gy IR. CIP2A^{WT} and CIP2A$^{561-904}$ co-localized with TOPBP1, whilst CIP2A$^{1-571}$ did not (n=2).

Together this data suggests that the conserved SLiM in TOPBP1 is not necessary for the interaction between TOPBP1 and the C-terminal coiled coil domain of CIP2A in the context of the full-length protein. This suggests that other regions of TOPBP1 may contribute to the TOPBP1-CIP2A interaction in the context of the full-length protein. Further work conducted by collaborators has sought to identify the functional significance of this interaction and the manuscript is currently under review. They found that CIP2A functions downstream of MDC1 to mediate TOPBP1 recruitment to sites of chromosome breaks in mitosis.

3.3.7 Two SLiMs in the C-terminus of AUNIP interact with CtIP

AUNIP promotes CtIP-dependent DNA end resection at DNA replication forks (Lou et al., 2017). It contains an N-terminal DNA binding domain, as well a C-terminal CtIP-interacting domain (**Fig 3.20A**). Other than this, little is known about this protein and so to gain insight into the mechanisms by which AUNIP promotes DNA end resection, amino acid sequence alignments of AUNIP from representative vertebrate orthologues were generated (section 8.1.2). Using the prediction criteria described previously, two SLiMs highly conserved across vertebrate evolution were identified at the C-terminus of the protein. Both SLiMs contained a highly conserved serine and threonine residue. The SLiMs are homologous and so three synthetic biotinylated peptides corresponding to only the extreme C-terminal SLiM were synthesised: one native, one incorporating a phospho-threonine at the T344 and one incorporating a phospho-serine at the S347. The biotinylated peptides were conjugated to streptavidin-coupled beads and incubated in HeLa nuclear extract. Potential interacting partners were purified using a non-denaturing wash and then identified by mass spectrometry.

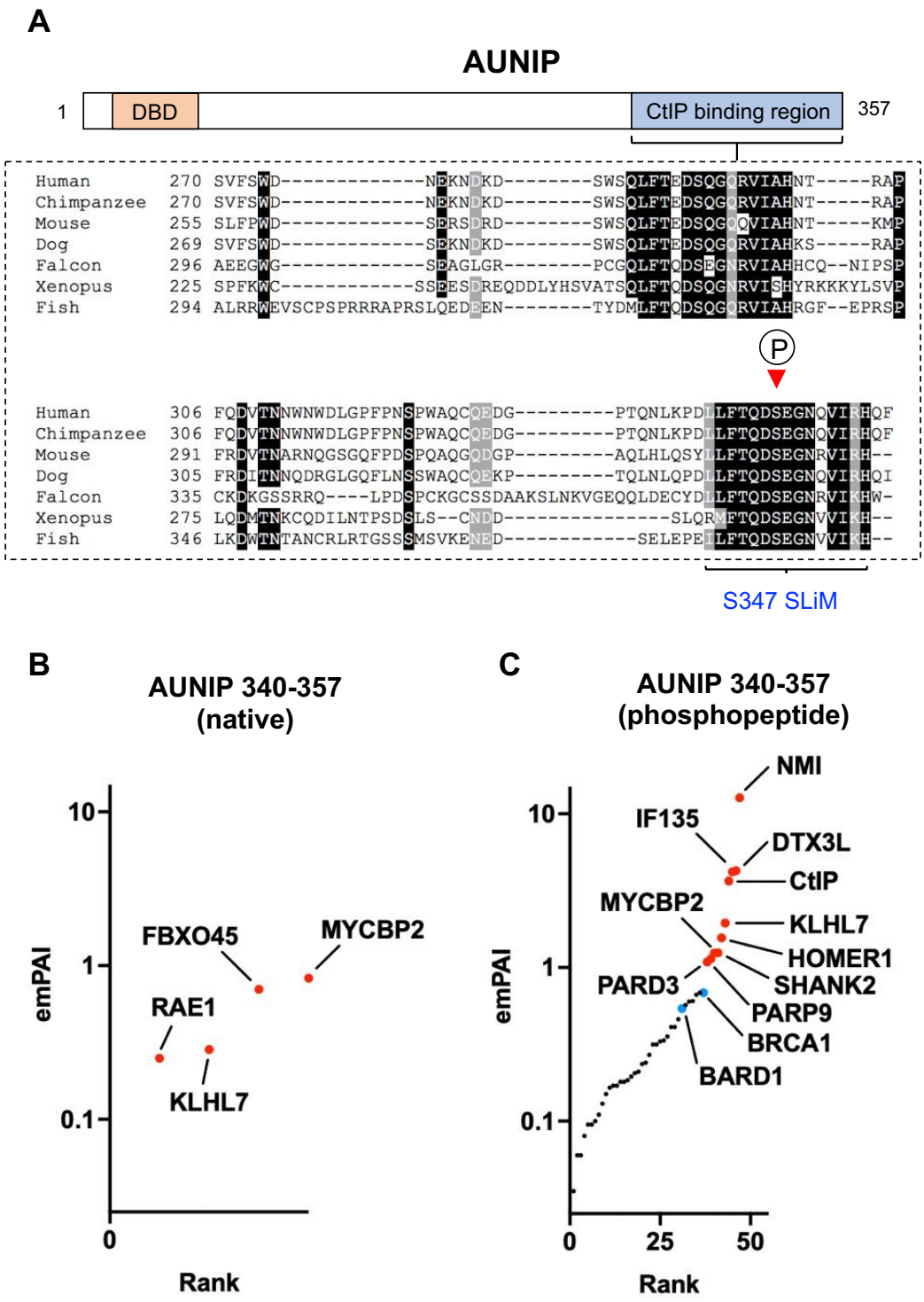


Fig. 3.20. A SLiM in the C-terminus of AUNIP interacts with CtIP.
A) Schematic showing the layout of AUNIP. The DNA binding domain (DBD) and the CtIP interacting domain is shown. A region of the sequence alignment for AUNIP, centred around the two C-terminal SLiMs, is shown. P, phosphorylation. B) Plot of LC-MS/MS hits from native peptide pulldown of the peptide corresponding to AUNIP 340-357, ranked by emPAI. The high-confidence candidate interacting proteins are highlighted in red. C) Plot of LC-MS/MS hits from native peptide pulldown of the phosphopeptide corresponding to AUNIP 340-357, ranked by emPAI. The top ten high-confidence candidate interacting proteins are highlighted in red. This identified CtIP as a specific interactor for the AUNIP-S347 phosphopeptide. BRCA1, known to interact with CtIP and BARD1, which forms a heterodimer with BRCA1 are highlighted in blue.

When the proteins unique to the phospho-S347 AUNIP peptide pulldown were examined, the protein CtIP was identified as one of the highest confidence candidate interacting proteins in two independent repeats (**Fig, 3.20B,C**). BRCA1 (known to interact with CtIP) and BARD1 (known to interact with BRCA1) were also found amongst the top hits. Being consistent with literature and seeking to understand the regulation of the interaction between CtIP and AUNIP, this interaction was validated (Lou et al., 2017). Other high-confidence candidate interaction proteins were identified, N-myc interactor (NMI) and the E3 ubiquitin-protein ligase, DTX3L. These interactions will be validated in further studies. Western blotting confirmed that CtIP was specifically pulled down by the phospho-S347 AUNIP peptide pulldown (**Fig. 3.21A**). Next, it was investigated whether pS347 was required for the AUNIP-CtIP interaction in the context of full-length AUNIP.

Given the similarity to the SLiM N-terminal of the S347 binding SLiM, the possibility that these two SLiMs may act redundantly was considered. To test this, a series of GFP-tagged AUNIP constructs were designed and synthesised: WT, a construct with both SLiMs deleted (Δ1/2) and a construct with the four conserved serines or threonines mutated to alanine, as well as the less well conserved serine at the beginning of the first SLiM (5A). These constructs were transfected into 293FT cells and GFP IPs from lysates harvested from these cells demonstrated that Δ1/2 and the phospho-dead mutant (5A) abrogated the CtIP interaction (**Fig. 3.21B**). To test the requirements for each SLiM individually, SLiM 1 (Δ285-300) or SLiM 2 (Δ340-357) were deleted in the context of full length GFP-AUNIP. Furthermore, S292 or S347 was mutated to alanine in the context of full length GFP-AUNIP. These constructs were transfected into 293FT cells and GFP IPs from lysates harvested from these cells demonstrated that deletion of SLiM 1 (Δ285-300) or SLiM 2 (Δ340-357) caused a partial defect in the AUNIP-CtIP interaction when compared to deletion of both SLiMs (**Fig. 3.21C**). This suggests that both SLiMs contribute to CtIP interaction. Mutation of S292 or S347 to alanine caused a mild defect

in the AUNIP-CtIP interaction, suggesting that in isolation these individual phosphorylation sites are not necessary for the interaction between CtIP and AUNIP.

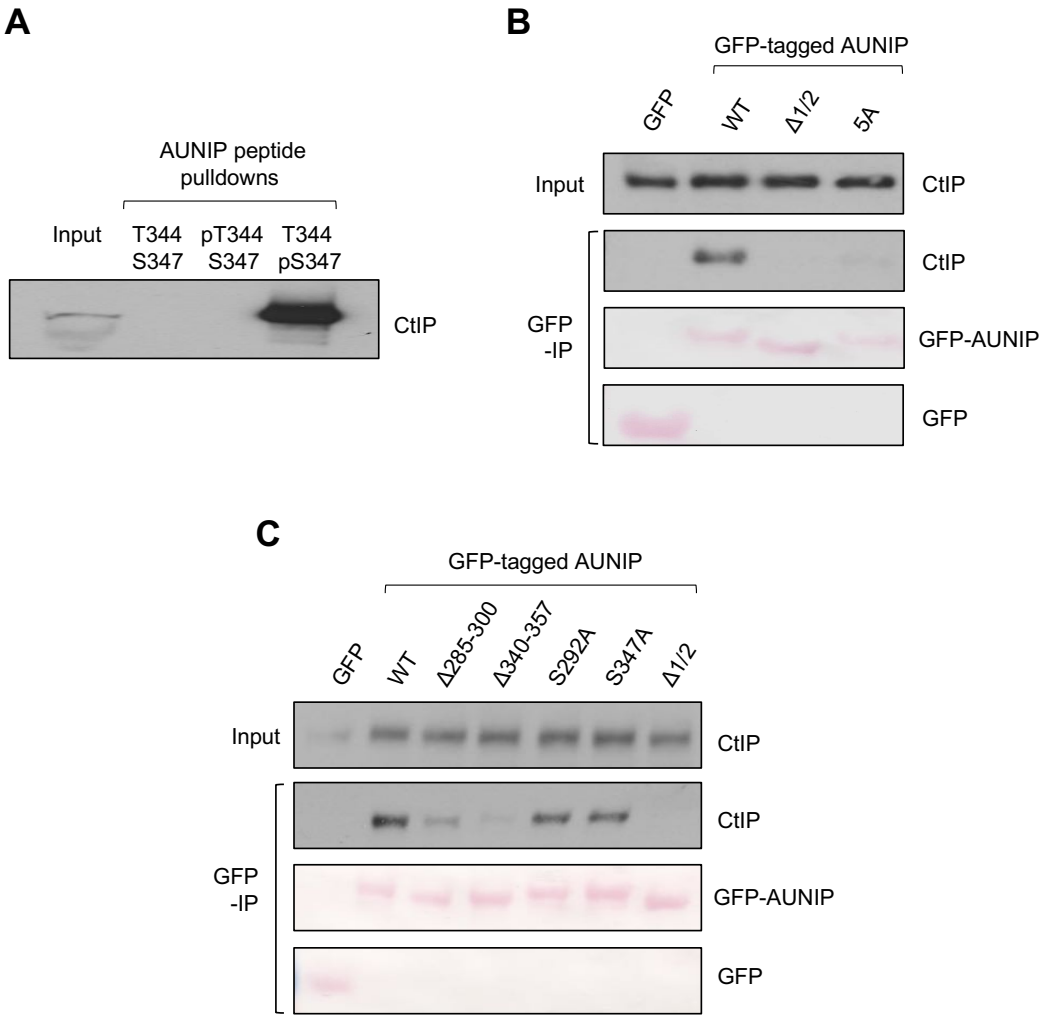


Fig. 3.21. Validation of the CtIP-binding SLiM in the context of full-length AUNIP
A) Peptide pulldowns from HeLa nuclear extracts using peptides corresponding to residues surrounding AUNIP-S347, verifying that CtIP is a specific interactor for the AUNIP-S347 phosphopeptide. B) GFP-IPs from 293FT cells transfected with the indicated GFP-tagged AUNIP variants, demonstrated that Δ1/2 and the phospho-dead mutant (5A) abrogated the CtIP interaction. C) GFP-IPs from 293FT cells transfected with the indicated GFP-tagged AUNIP variants, demonstrated that deletion of SLiM 1 (Δ285-300) or SLiM 2 (Δ340-357) caused a partial defect in the AUNIP-CtIP interaction when compared to deletion of both SLiMs. Mutation of S292 or S347 to alanine caused a mild defect in the AUNIP-CtIP interaction.

In summary, this data suggests that two SLiMs in the C-terminus of AUNIP are required for the interaction between AUNIP and CtIP. Published literature has suggested that

the function of this interaction is to promote CtIP-dependent DNA end resection at DNA replication forks (Lou et al., 2017). As the phospho-dead mutant (5A) abrogated the CtIP interaction, this has demonstrated that the interaction is phosphorylation-dependent, and furthers studies are required to identify the kinase responsible for the phosphorylation. This could be done by transfecting 293FT cells with GFP-AUNIP WT and prior to harvesting the lysates treating the cells with various different kinase inhibitors, such as ATRi, to determine if inhibition of any specific kinase abrogates the interaction. This will facilitate insight into the regulation of this interaction, as well as the regulation of CtIP-dependent DNA end resection at DNA replication forks.

3.3.8 Identification of Three Conserved RPA-binding SLiMs in the BTR Complex

The RecQ helicase BLM functions in cells with topoisomerase III alpha (TOP3A) and RecQ-mediated genome instability proteins 1 and 2 (RMI1 and RMI2, respectively). Together, these proteins form the BTR "dissolvasome" complex (Bizard and Hickson, 2014). Despite a comprehensive investigation of the structured regions of the BTR complex, the function of intrinsically disordered regions within the BTR complex is still poorly understood. Therefore, regions of intrinsic structural disorder within the BTR complex were analysed to identify SLiMs.

To identify SLiMs in the BTR complex, amino acid sequence alignments of all four BTR subunits (BLM, TOP3A, RMI1 and RMI2) were generated. Using the prediction criteria previously described, six SLiMs were identified within the complex, all highly conserved across evolution and in unstructured regions of the proteins; four of which were found in BLM (section 8.1.4**, Fig. 3.22A**) and two in RMI1 (section 8.1.13**, Fig. 3.22A**). No putative peptide SLiMs were identified in RMI2 or TOP3A. In accordance with this, TOP3A consists of a topoisomerase domain and a series of structured zinc-finger domains.

RMI2 consists of a single oligonucleotide-binding (OB)-fold domain.

The possibility that these SLiMs in BLM and RMI1 could constitute previously uncharacterised interaction surfaces for other proteins was considered. In support of this hypothesis, BLM 290-309 (herein referred to as SLiM 2) has been shown to bind to TOPBP1 when phosphorylated on S304 (**Fig. 3.22A**; Blackford et al., 2015; Sun et al., 2017). Therefore, synthetic biotinylated peptides corresponding to all six SLiMs were designed for use as baits in pulldown experiments. Native and phospho-S304 versions of SLiM 2 were synthesised for use in subsequent experiments. Biotinylated peptides were conjugated to streptavidin-coupled beads and incubated in HeLa nuclear extract. Potential interacting partners were purified using a non-denaturing wash and then identified by LC-MS/MS. TOPBP1 was identified as the highest confidence candidate interacting protein of SLiM 2, but only when the peptide corresponding to this SLiM was phosphorylated on S304 (**Fig 3.22B**) and this was confirmed by western blotting (**Fig 3.22C**). In contrast, peptides corresponding to BLM 1392-1417 (SLiM 4) and RMI1 308-329 (SLiM 6) did not interact with any proteins other than those designated to be common contaminants. Conversely, when peptides corresponding to BLM 137-165 (SLiM 1), BLM 552-568 (SLiM 3) and RMI1 240-260 (SLiM 5) were used as baits, all three subunits of replication protein A (RPA) were the highest confidence candidate interacting proteins in the pulldown (**Fig. 3.22D,E,F**). Of note, in two out of the three peptide pulldowns, the RPA binding protein RAD52, was also one of the highest confidence candidate interacting proteins.

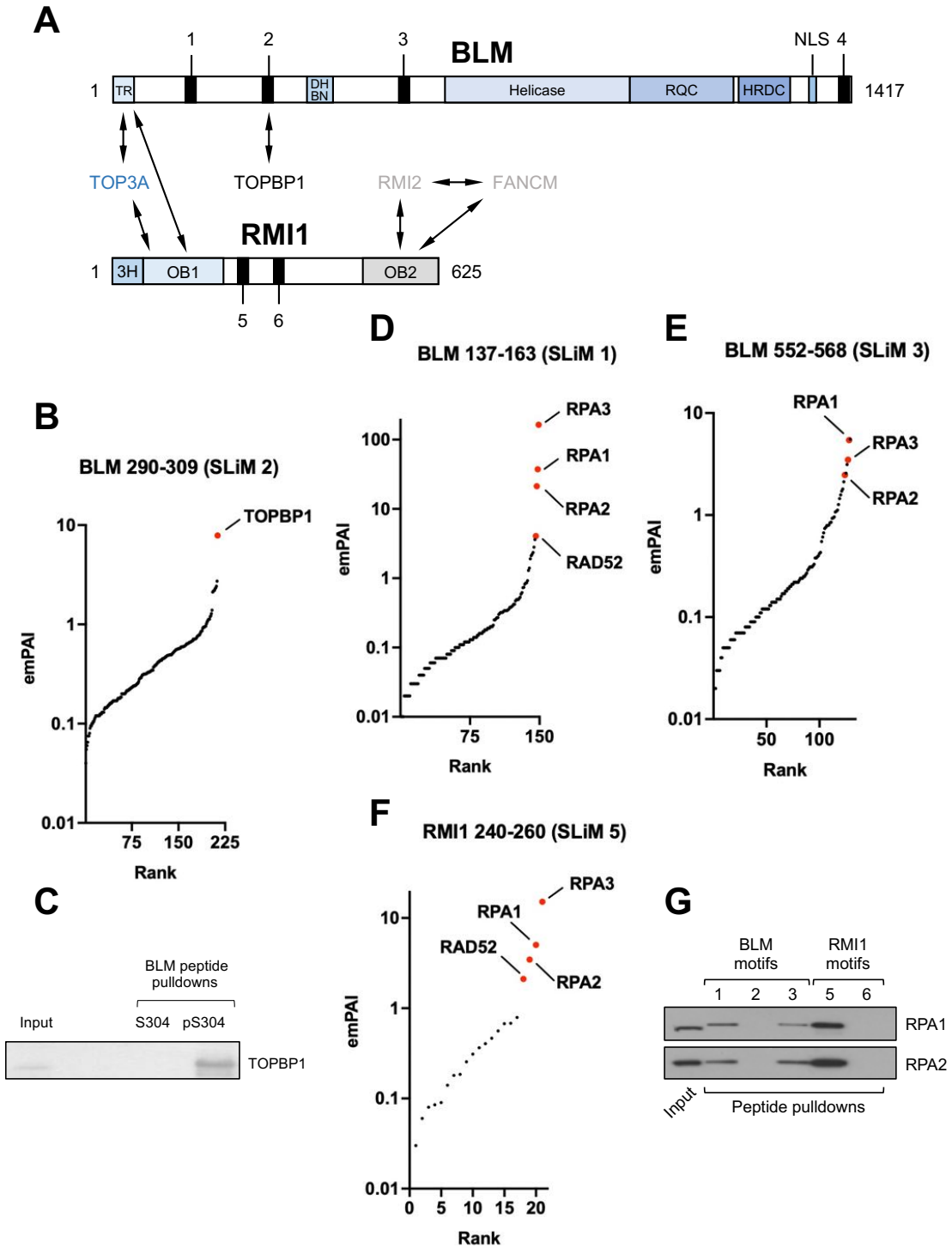


Fig. 3.22. Identification of conserved RPA-binding SLiMs in the BTR complex.
A) Schematic of proteins, domains and SLiMs in the BTR complex. TR, TOP3A-RMI1 binding domain; DHBN, dimerization helical bundle in N-terminal; RQC, RecQ C-terminal; HRDC, helicase and RNaseD C-terminal; 3H, 3-helix bundle; OB, oligonucleotide-binding. Numbered SLiMs in BLM and RMI1 correspond to sequences synthesized as biotinylated peptides. B) Identification of TOPBP1 as a specific interactor for the BLM-S304 phosphopeptide by LC–MS/MS. TOPBP1 is highlighted in red. C) Validation of MS results by western blotting, showing that TOPBP1 is a specific interactor for the BLM-S304 phosphopeptide. D-F) Identification of RPA as the highest confidence candidate interactor for SLiMs 1, 3 and 5 in the BTR complex. RPA subunits are highlighted in red. Experiments in D & E were conducted by Dr. Andrew Blackford. G) Validation of MS results by western blotting, showing that RPA from HeLa nuclear extracts is specifically pulled down by biotinylated peptides based on SLiMs 1, 3, and 5 but not SLiMs 2 and 6.

Western blotting confirmed that RPA specifically interacted with biotinylated peptides based on SLiMs 1, 3 and 5 but not SLiMs 2 and 6 (**Fig. 3.22G**). Therefore, it could be speculated that the BTR complex contains at least three separate interaction sites for RPA; two in BLM, one in RMI1.

In summary, this data suggests that three SLiMs in the BTR complex interact with RPA. The validation of these interactions in the context of full-length BLM/RMI1 and their functional significance will be explored in the next chapter.

3.4 Discussion

Protein-protein interactions play essential roles in many biological processes, including the DDR (Ding and Kihara, 2019). Identification of protein-protein interactions is useful for defining multimeric protein complexes and biological pathways, predicting protein function as well as highlighting candidates for the development of new therapies. In this chapter, a bioinformatics and proteomics pipeline was developed to identify and subsequently characterize SLiMs in DSB repair proteins in a pilot screen. To identify SLiMs, multiple sequence alignments had to be generated. Manual sequences alignments are labour intensive but guarantee the inclusion of sequences spanning an appropriate evolutionary distance, as well as the ability to exclude partial sequences. Being labour intensive, if they are being generated on a larger scale, there is a limit in the number of vertebrate sequences that can be acquired and in the total number of alignments that can be feasibly generated. Therefore, the higher quality of these alignments has to be balanced with the speed of their generation, with the quantity of alignments feasibly generated being lower and with relatively few vertebrate orthologues included. Various open access bioinformatic tools exist, such as the ConSurf web server (http://consurf.tau.ac.il), which generate sequence alignments automatically. This method is far less labour intensive and allows for the generation of sequence alignments

spanning a large number of orthologues. However, it tends to include incomplete protein sequences and may not include species spanning a suitable evolutionary distance. This can occlude the identification of SLiMs in that protein of interest. Therefore, with automated sequence alignments, the higher quantity of the alignments that can be generated is compromised by the quality of alignments being lower. One way of overcoming the limitations of both methods would be to develop our own tool that generates sequence alignments based on the criteria used to generate manual sequence alignments. The aim of this web-based tool would be to facilitate the generation of high quality but also a high quantity of sequence alignments for the identification of SLiMs. A web-based tool has been developed for this purpose, and its generation is detailed in Chapter 5. The SLiMCon tool predicts SLiMs present within the sequence alignments, reducing the likelihood of personal bias when identifying SLiMs from the manual sequence alignments.

Using manually generated sequence alignments, 36 potential SLiMs were identified and synthesised as biotinylated peptides and used as baits to identify physiological binding partners from HeLa nuclear extract by mass spectrometry. Some peptide pulldowns were found to have over 1500 interacting proteins by mass spectrometry, highlighting the need for a method to distinguish between high-confidence candidate interacting proteins and non-specific interacting proteins. For a protein to be classified as a high-confidence candidate interacting protein, it could not be part of the background proteome. The designation of a protein to be a part of the background proteome and its subsequent elimination from the dataset is contentious to some, as one cannot rule out the possibility that in that particular incidence the interaction is biologically relevant. However, it was clear that in many peptide pulldowns, the most abundant proteins in the samples were common contaminants such as keratins and ribonucleic proteins, which are unlikely to represent biologically relevant interactions, given the nature of this study seeking to identify protein-protein interactions mediated by DDR factors. The most

challenging aspect of defining which proteins were considered common contaminants was selecting a threshold of peptide pulldowns a protein had to be present in to be considered part of the background proteome. To ensure that this threshold was suitable, proteins within the background proteome were excluded from hits in peptide pulldowns with validated interactors, such as the SLiMs in BLM demonstrated to interact with TOPBP1 and RPA. Once this had been done, the most abundant protein in the sample was the expected validated interactor. This gave some assurance that the threshold set was not too lenient or too stringent.

Of the 36 SLiMs, 81% (n = 29) had high-confidence candidate interacting proteins and 19% (n = 7) did not (**Fig 3.5D**). This could suggest that the 7 identified putative SLiMs without high-confidence candidate interacting proteins are not SLiMs. However, there are several other reasons why these SLiMs do not have high-confidence interacting proteins by this method. A limitation of this technique is that it relies on the assumption that peptides corresponding to each SLiM can form the same protein-protein interactions as *in vivo*. However, this may not be the case for various reasons. Firstly, within the context of the full-length protein the SLiM may co-operate with other regions of the protein to promote the protein-protein interaction in peptide pulldowns in HeLa nuclear extract. Another limitation of this technique is that it cannot probe how the sequences surrounding the SLiM affect their interactions, outside of the immediate surrounding sequence. Early crystal structures of SLiM-binding domains bound to SLiMs typically demonstrated that three to six residues of the SLiM interacted with the SLiM-binding domain (Lim et al., 1994; Macias et al., 1996). These findings lead to the dogma that the three to six core positions confer much of the interaction specificity. This led to the question as to how high-fidelity interactions are maintained between low complexity SLiMs and their recognition domains but there is increasing evidence that the surrounding sequence critically influences SLiM interaction affinity and specificity (Hwang et al., 2022; 2021). To take this into account, the immediate surrounding sequence of the SLiM was incorporated into the

peptide. However, in the context of the full-length protein, distal regions of the protein may be in immediate proximity to the SLiM and may contribute to the interaction. The peptide-pulldown method cannot take this into account and is therefore a limitation of the method.

In some cases, the synthetic peptide corresponding to the SLiM was sufficient to promote the interaction, even if other regions of the protein promote the protein-protein interaction. This was the case with the interaction between RPA and the three RPA-binding SLiMs in BTR complex (**Fig. 3.22**). In Chapter 4, it is demonstrated that all three SLiMs contribute to the interaction between the BTR complex and RPA in the context of the full-length protein. However, each SLiM was sufficient to promote the interaction with SLiM in peptide pulldown experiments (**Fig. 3.22D,E,F**). In other cases, the synthesised peptide corresponding to the SLiM may not be sufficient to promote the interaction in peptide pulldowns in HeLa nuclear extract as other regions of the protein may be required to facilitate the protein-protein interaction. One example of this is peptide corresponding to SLX4 603-628. When it was used as a bait in peptide pulldowns it was found to interact with members of the mitochondrial inner membrane complex (**Fig. 3.15A,B**). In the context of full-length SLX4, the SLiM is 80 amino acids upstream of the oligomerization domain (the BTB domain), which promotes SLX4 dimerization. In a recently published report, this region (the SLiM and the BTB domain) was shown to be required for the interaction between SLX4 and the helicase RTEL1 (Takedachi et al., 2020). Point mutations of key residues within SLX4 603-628 abrogated the interaction between RTEL1 and SLX4. However, by our method, an interaction between to SLX4 603-628 and RTEL1 was not identified. This is an example of a SLiM whose interacting partner cannot be identified by our method, as in this case, the SLiM may not have high enough affinity for the interacting partner in isolation and requires other regions of the protein to promote the interaction.

Various examples in this chapter confirm the hypothesis that within the context of the full-length protein the SLiM may co-operate with other regions of the protein. When peptides corresponding to the MDC1 pS196 SLiM were used as bait in peptide pulldowns, TOPBP1 was enriched in the pulldown (**Fig 3.16B**). This interaction was subsequently validated in the context of the full-length protein (**Fig 3.16C**). However, further mapping experiments that aimed to determine the region of TOPBP1 that bound to the pS196 SLiM of MDC1 revealed that BRCT domains 1 and 2 of TOPBP1 interact with MDC1 (Leimbacher et al., 2019). The explanation for this finding was that there is an additional TOPBP1 binding site in MDC1. Further analysis of the multiple sequence alignment for MDC1 revealed the existence of another SLiM located just upstream of S196, consisting of a serine residue (S168 in human MDC1) embedded within an acidic sequence environment. Fluorescence polarisation experiments suggested that TOPBP1 BRCT1 interacts with phosphorylated S168 of MDC1 and TOPBP1 BRCT2 interacts with phosphorylated S196. This demonstrated that in the context of the full-length protein, both SLiMs participate in the interaction between MDC1 and TOPBP1. The interaction between TOPBP1 and CIP2A is another example of a more complex interaction (**Fig. 3.17**). When peptides corresponding to the TOPBP1 pT779 SLiM were used as baits in peptide pulldowns, CIP2A was enriched in the pulldown. However, when this SLiM was deleted in the context of full-length TOPBP1, the CIP2A interaction was reduced but not abrogated. This suggests that other regions of TOPBP1 (not identified here) also contribute towards the interaction between CIP2A and TOPBP1. Therefore, whilst many SLiMs can function in isolation, the data presented here suggests that SLiM-mediated interactions may act cooperatively, with multiple SLiMs/domains contributing to a given binding event. Cooperative binding can increase the specificity of the interaction, permitting high fidelity recognition of biologically relevant binding partners despite the large number of sequences present within the proteome that fit the specificity requirements of a single SLiM (Van Roey

and Davey, 2015). The same mechanisms allow the intrinsically weak affinities of a single SLiM to be increased by binding multivalently.

Another reason why a SLiM may not be found to have high-confidence candidate interacting proteins by this method is because the interaction could be dependent on a PTM other than phosphorylation, such as methylation, ubiquitylation or SUMOylation. The pipeline described here was limited to phosphorylation, for ease of peptide synthesis. If a SLiM contains a PTM consensus sequence other than phosphorylation, one could attempt to synthesise peptides with other modifications. However, it is important to note that it may be impossible to synthesise peptides incorporating certain modifications and that is a limitation of this method. For example, a ubiquitylated peptide incorporating the 9 kDa ubiquitin protein would be difficult to synthesise.

In addition, the peptide pulldown buffer may not reconstitute the intracellular conditions required for the protein-protein interaction and result in no identifiable interacting partners by mass spectrometry. Compartmentalisation within the cell has been recognised for many years as a key organisational principle of life (Aguzzi and Altmeyer, 2016). However, it has only been in recent years that subcellular partitioning has started to be understood at the molecular level, especially when it occurs independently of membranes. Recent work has provided evidence that membraneless organelles assemble by intracellular phase separation and behave like liquid droplets inside the cell. The propensity to phase separate seems to be particularly high for intrinsically disordered proteins. Once phase separation has occurred, proteins within the liquid environment are likely to find themselves in a different solvent environment compared to their surroundings. Accordingly, although the salt concentration within the peptide pulldown buffer (150 mM NaCl) was chosen to mimic physiological salt, it is possible that intracellularly (due to phase separation) the interaction forms in the presence of a different salt concentration. To maximise the chances of identifying the physiological

binding partner of a SLiM, peptide pulldowns in the presence of a range of salt concentrations could be trialled, such as 150 mM, 300 mM and 450 mM salt.

HeLa nuclear extract is used in this study to recapitulate the nucleus *in vitro*. Consistent with their biological role, the vast majority of DDR proteins are nuclear, justifying the use of a nuclear extract, as opposed to whole cell extract (Pearl et al., 2015). However, a peptide corresponding to a SLiM may have no identifiable interacting partners by mass spectrometry because the SLiM interacts with a cytoplasmic protein *in vivo*. The SLiM based interaction could be specific to mitosis, an effective method of regulating protein-protein interactions, as it would restrict the protein-protein interactions to after the nuclear membrane has broken down. Use of a nuclear extract could preclude identification of the biological binding partner. To overcome this, peptide pulldowns could instead be carried in lysates prepared from whole cell extract. However, there is some suggestion the nuclear extract used in this study contains some contamination from cytoplasmic proteins. CIP2A is a cytoplasmic protein yet when peptides corresponding to the TOPBP1 pT779 SLiM are used as baits in peptide pulldowns, CIP2A is enriched in the pulldown.

A strength of this method is that protein-protein interactions that are dependent on modifications that only occur in specific cell cycle stages (e.g., CDK-dependent phosphorylation events) or in response to stimuli such as DNA damage (e.g., PIKK-dependent phosphorylation events), which may normally be present in very low amounts or only transiently in cells, using a peptide containing such a modification within a PTM consensus sequence. However, the interacting partner could be post-translationally modified in response to DDR stimuli or the cell cycle rather than the SLiM. To overcome this, peptide pulldowns could be carried out in lysates prepared from whole cells treated with specific types of DNA damage (such as irradiation) or synchronised in a specific cell cycle stage. However, it is important to note that HeLa nuclear extract originates from a cervical cancer cell line (HeLas)

and cancer cells have a higher level of intrinsic DNA damage when compared to untransformed cells.

Alternative methods of identifying SLiM based interactions exist, including proteomic peptide-phage display (Teyra et al., 2020). In this method, a phage-encoded peptide library is designed to display disordered regions of a target proteome. These peptide libraries are incubated with pre-defined, immobilised SLiM-binding domains and SLiM-based interactions are identified by sequencing the DNA of the captured phages. These methods form a powerful pipeline for proteome-wide discovery of SLiM-based interactions but can only identify interactions for pre-defined SLiM-binding domains unlike this method.

In conclusion, a combined bioinformatics and proteomics approach was developed and optimised to identify SLiMs in DSB repair proteins, as well as their interacting partner(s). One interesting finding was the identification of three SLiMs in the BLM helicase complex that interact with RPA (**Fig 3.22**). In Chapter 4, these interactions were characterized further, and the biological function was investigated.

Chapter 4: The Bloom Syndrome Complex Senses RPA Coated Single-Stranded DNA to Restart Stalled Replication Forks

4.1 Authorship Statement

The GFP immunoprecipitations in **Fig. 4.4C** and **Fig 4.5D** were conducted by Dr. Andrew Blackford, as was the GFP-BLM recruitment time course (**Fig 4.8A**). Validation of *BLM*$^{-/-}$ RPE-1 cells by western blot (**Fig 4.6B**) was conducted by Dr. Kaima Tsukada. All SCE assays in this chapter (**Fig 4.6D, 4.8B**) were conducted by Dr. Samuel Jones & Dr. Kaima Tsukada. Lentiviruses were produced by Dr. Samuel Jones, as was the generation of all complemented *BLM*$^{-/-}$ RPE-1 cell lines in section 4.3.7. The survival assay assessing HU sensitivity (**Fig. 4.8G**) was conducted by Zoulikha Belblidia.

4.2 Introduction

4.2.1 Bloom Syndrome is Characterised by Developmental Defects and Cancer Predisposition

Bloom syndrome is a rare hereditary chromosomal instability disorder, first described in 1954 by the dermatologist David Bloom in New York City (Bloom, 1954). It is an autosomal recessive disease, predominately caused by loss-of-function mutations in the *BLM* gene located on chromosome 15q26.1 (Ellis et al., 1995). Clinically, Bloom syndrome is characterised by growth retardation, hypersensitivity to sunlight, immunodeficiency, early-onset diabetes, infections of the middle ear and respiratory tract, infertility and a greatly increased cancer

predisposition (German, 1993). Strikingly, Bloom syndrome patients are predisposed to the full spectrum of malignancies found in the general population, at roughly the same relative frequencies but developed at a much younger age. In line with this, leukemias tend to predominate in childhood, with lymphomas appearing in both children and adults. Carcinomas begin to appear during adolescence and predominate thereafter. In contrast, other inherited cancer-predisposing disorders tend to be associated with malignancies of specific tissue types (Ciccia and Elledge, 2010). For example, recessive mutations in the *FANC* family of genes predispose patients to AML, as well as head and neck carcinomas, whereas mutations in mismatch repair genes predispose patients to gastrointestinal tumours.

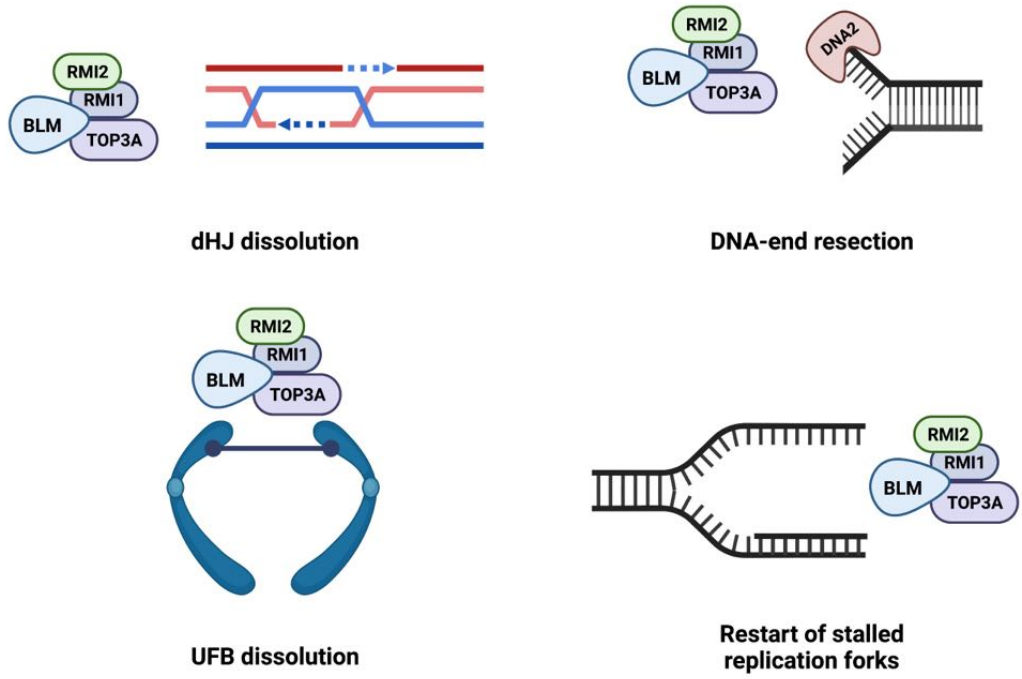


Fig. 4.1. The cellular functions of the BTR dissolvasome complex.
The BTR complex has been implicated in dHJ dissolution, stimulating DNA-end resection, the resolution of ultrafine DNA bridges (UFBs) between sister chromatids in anaphase and promoting the restart of stalled replication forks. Created with BioRender.com.

All known cellular functions of the BTR complex require the helicase domain of BLM, but BTR complex members can interact with other proteins via other domains, which may contribute to the roles of the complex in specific pathways. At the onset of this project, little was known about the relative contributions of BLM in dHJ dissolution, DNA-end resection, UFB resolution and stalled fork restart in promoting tumour suppression (**Fig. 4.1**). Furthermore, little was understood as to how the BTR complex is regulated in response to genotoxic stress, and how BLM can differentiate between unperturbed DNA replication and DNA damage or stalled replication forks.

4.2.2 Aims and Objectives

In the previous chapter, three SLiMs from the BTR complex were shown to interact with RPA (**Fig. 3.22**). RPA is an abundant, high-affinity single-stranded DNA (ssDNA)-binding complex that recognizes ssDNA rapidly in cells, whether generated during DNA replication or as a result of DNA damage. It consists of three subunits, RPA1, RPA2 and RPA3 (also known as RPA70, RPA32 and RPA14 respectively). All three subunits contain oligonucleotide (OB)-binding folds that either bind to ssDNA or mediate protein-protein interactions, with the N-terminal OB-fold of RPA1 and the C-terminal winged-helix domain of RPA2 being the major interaction sites recognized by proteins associated with the DDR or DNA replication stress response (Bhat and Cortez, 2018). BLM complexes immunoprecipitated from cellular extracts contain RPA, and BLM and RMI1 bind to RPA *in vitro* (Meetei et al., 2003; Brosh et al., 2000, Xue et al., 2013). However, the RPA binding sites within the BTR complex have not yet been precisely mapped and thus the physiological relevance of its association with RPA has not been determined. With RPA being highly enriched in the peptide pulldowns, these three potential RPA-binding SLiMs were investigated further.

The first aim was to validate these interactions in the context of full-length BLM/RMI1 proteins. As little is known of the mechanisms by which the BTR complex is regulated *in vivo*, the second aim was to identify the functional significance of the interaction between the BTR complex and RPA. This involved the use of a combination of CRISPR-Cas9, RNA interference and ectopic expression of recombinant proteins delivered by lentivirus, to generate cell lines expressing a BTR complex that cannot bind RPA. Subsequently, functional assays were used to identify which cellular function(s) the interaction between the BTR complex, and RPA were required for.

4.3 Results

4.3.1 Identification of Three Conserved RPA-binding SLiMs in the BTR Complex

In the previous chapter, three SLiMs from the BTR complex were shown to interact with RPA (**Fig. 3.22**). To test whether these SLiMs are necessary to mediate the interaction between BLM/RMI1 and RPA, each SLiM was mutated in the context of full-length BLM and RMI1. Firstly, SLiMs 5 or 6 were deleted individually in constructs expressing GFP-tagged wild-type RMI1 using site-directed mutagenesis (**Fig. 4.2A**) and then transfected into 293FT cells. The interaction between RPA and SLiM 6 was assessed because a recent study suggested this region of RMI1 contains the RPA binding site (Xue et al., 2013). Verifying peptide pulldown experiments (**Fig. 3.22G**), GFP IPs from lysates harvested from these cells demonstrated that deletion of SLiM 5 but not SLiM 6 abrogated RPA binding (**Fig. 4.2B**). Conversely, the interaction of RMI1 with BLM or RMI2, which bind to distinct regions of RMI1 (the OB1 and OB2-fold domains respectively) was not affected (Raynard et al., 2008; Xu et al., 2008a). This demonstrated that deletion of SLiM 5 or 6 does not cause general misfolding of RMI1.

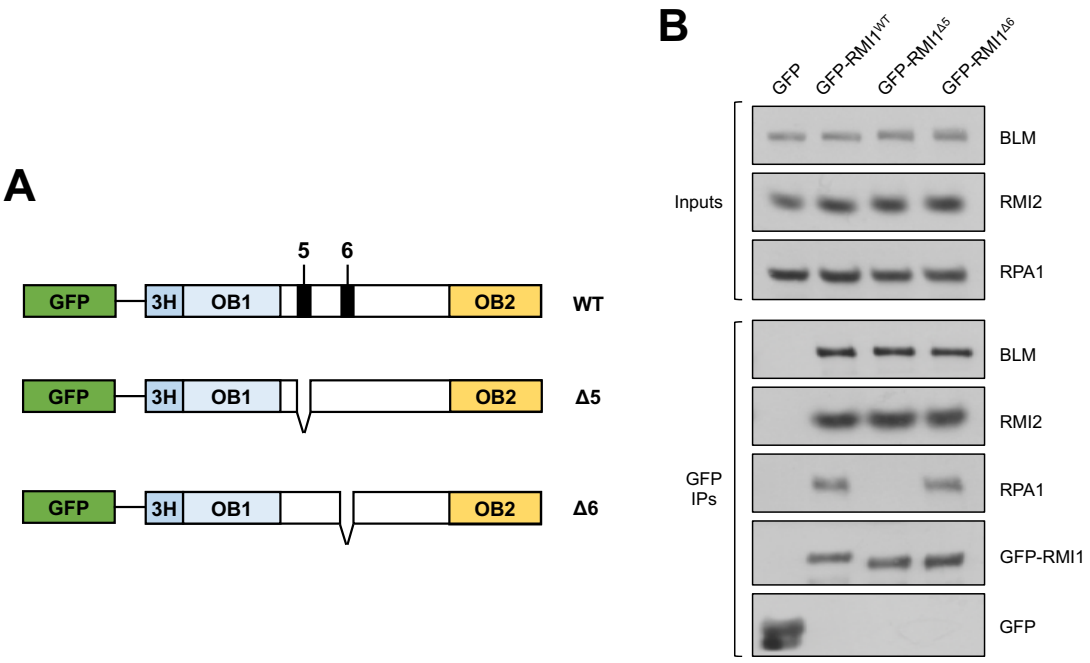


Fig. 4.2. Validation of RPA-binding SLiMs in the context of full-length RMI1.
A) Schematic of GFP-tagged RMI1 constructs used in this study. 3H, 3-helix bundle; OB, oligonucleotide-binding. B) GFP-pulldowns from 293FT cells transfected with constructs expressing GFP or the indicated GFP-tagged RMI1 variants, showing that loss of SLiM 5 disrupts RPA binding, whereas loss of SLiM 6 does not.

Next, SLiMs 1, 2 and 3 were deleted individually and in combination in constructs expressing FLAG-tagged BLM via site-directed mutagenesis (**Fig. 4.3A**) and these constructs were co-transfected into 293FT with GFP or GFP-RPA1. GFP IPs from lysates from these cells revealed that deletion of SLiM 1 or SLiM 3 mildly reduced the interaction between SFB-BLM and GFP-RPA1, whereas loss of SLiM 2 did not. SLiM 2 was assessed because a recent study demonstrated that peptides from BLM (similar to SLiMs 1 and 2) interact with a N-terminal fragment of RPA1 (Kang et al., 2018). Deletion of SLiM 1 and 3 together resulted in a more pronounced defect in RPA binding, which was not increased by also deleting SLiM 2 (**Fig. 4.3B**).

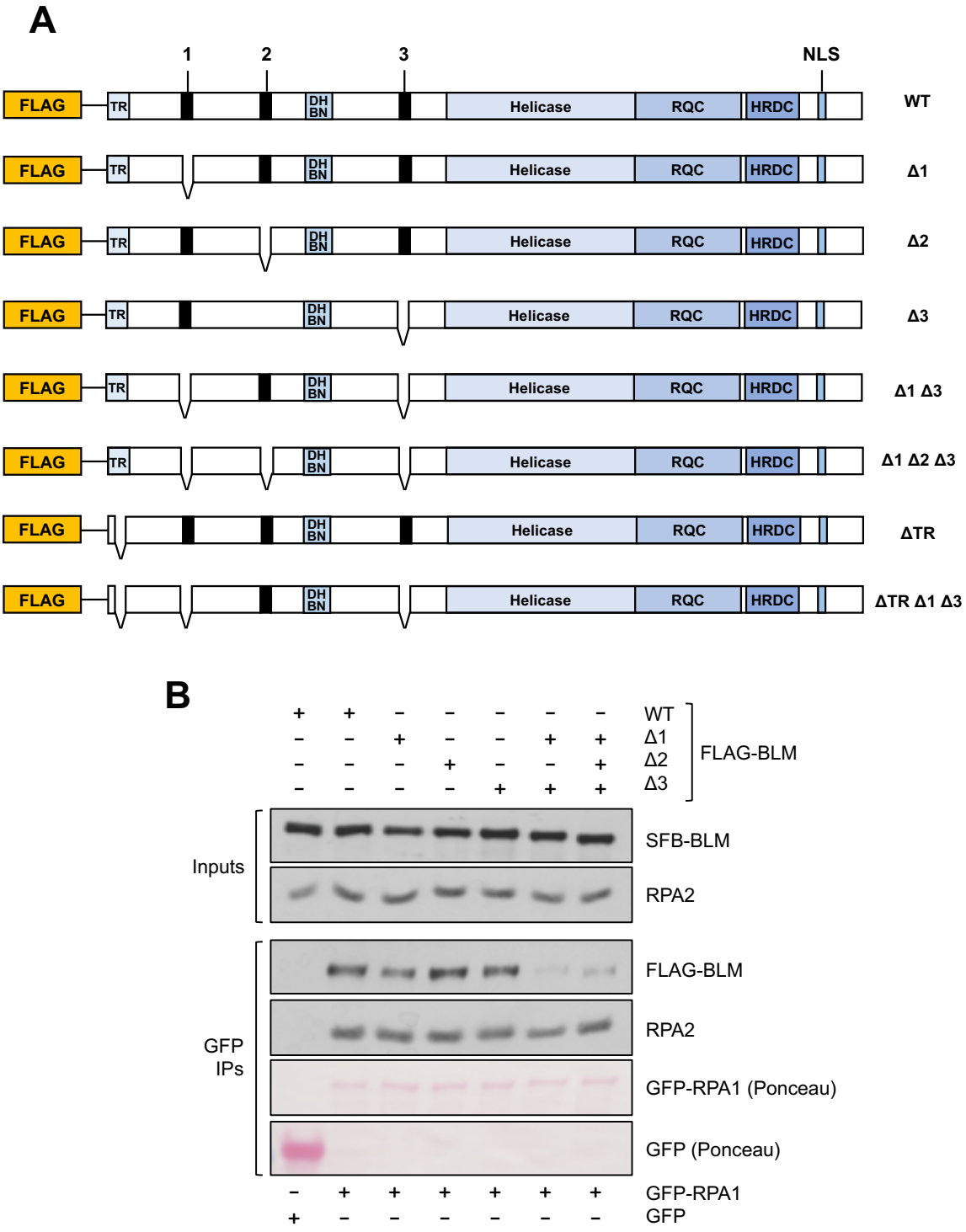


Fig. 4.3. Validation of RPA-binding SLiMs in the context of full-length BLM.
A) Schematic of SFB-tagged BLM constructs used in this study. TR, TOP3A-RMI1 binding domain; DHBN, dimerization helical bundle in N-terminal; RQC, RecQ C-terminal; HRDC, helicase and RNaseD C-terminal. B) GFP-pulldowns from 293FT cells transfected with constructs expressing either GFP or GFP-RPA1, and the indicated FLAG-tagged BLM variants, showing how loss of SLiMs 1 and 3 but not SLiM 2 of BLM impacts on RPA binding.

RMI1 contains an RPA binding SLiM, so the presence of endogenous RMI1 in the BTR complex could be responsible for the residual BLM-RPA binding observed in the absence of both SLiMs 1 and 3. To test this, a BLM construct defective for the RMI1 interaction was required. The first 133 residues of BLM (the TR domain; **Fig. 3.22A**) have been shown to promote the interaction between RMI1 and the rest of the BTR complex (Blackford et al., 2015). However, SLiM 1 is close to the TR domain and so deletion of the whole TR domain (residues 1-133) could interfere with BLM-RPA binding by affecting SLiM 1. Two regions (residues 9-17 and 25-38) within the N-terminus of Sgs1, the *S. pombe* orthologue of BLM, are required for the interaction between Top3 and Rmi1 with Sgs1. Point mutations in conserved residues in these regions of Sgs1 abrogate the interaction with Top3/Rmi1 (Kennedy et al., 2013; **Fig. 4.4A**). Therefore, to narrow down the TOP3A/RMI1 binding site in BLM, the human BLM TR domain was examined to identify analogous residues within the human protein. Accordingly, a region in the TR domain was deleted, spanning residues 4-37 of BLM, to determine if this disrupted association of BLM with TOP3A, RMI1 and RMI2. This was confirmed by GFP IP (**Fig. 4.4B**). Conversely, the interaction of BLM with TOPBP1, which binds to a distinct region of BLM (SLiM 2; Blackford et al., 2015) was not affected, demonstrating deletion of these residues does not cause general misfolding of BLM. The effect of this mutation (ΔTR) on interaction between BLM and RPA was examined alone and in combination with deletions of SLiMs 1 and 3 (**Fig. 4.4C**). Alone, this mutation resulted in a small decrease in RPA binding, suggesting that the interaction between BLM and RMI1 depends to some extent on the presence of the RPA-binding SLiM in RMI1 (**Fig. 4.4C**). Furthermore, deletion of the TR domain together with deletion of SLiMs 1 and 3 abrogated almost all of the binding of RPA to BLM. Together, these results demonstrate all three RPA-binding SLiMs in the BTR complex contribute to maximise the BLM-RPA interaction.

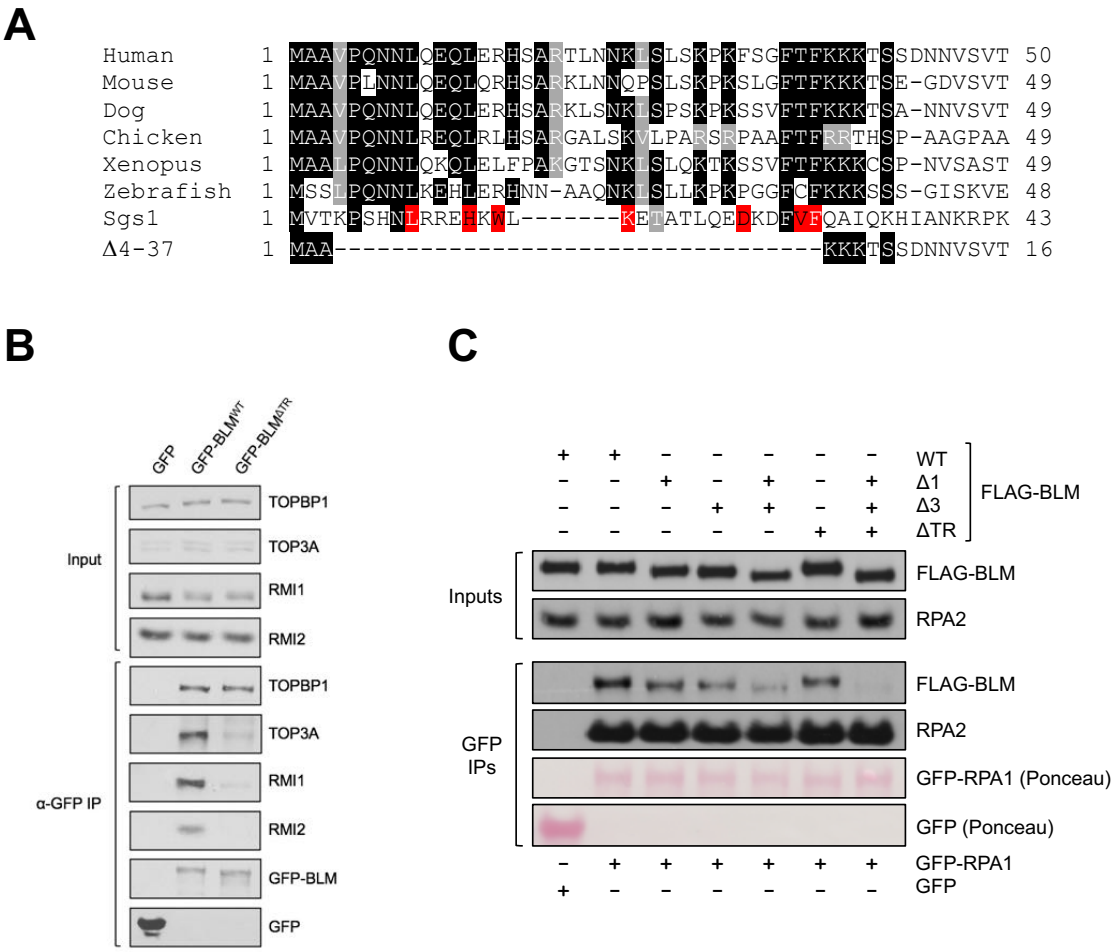


Fig. 4.4. The RPA-binding SLiMs interact cooperatively with RPA.
A) Sequence alignment of the N-terminal TR region of vertebrate BLM orthologs and Sgs1. Residues in red indicate those required for Sgs1 to promote resistance to HU (Kennedy et al., 2013) B) 293FT cells were transfected with the indicated constructs and GFP-pulldowns were carried out from cell lysates, followed by western blotting for the indicated proteins demonstrated that deletion of residues 4-37 in the TR domain of BLM abrogates interaction with the BTR complex. C) GFP-pulldowns from 293FT cells transfected with constructs expressing either GFP or GFP-RPA1, and the indicated FLAG-tagged BLM variants, showing how loss of SLiMs 1 and/or 3, and the N-terminal TOP3A-RMI1 (TR) binding domain of BLM impacts on RPA binding. Experiment conducted by Dr. Andrew Blackford.

4.3.2 The RPA-binding SLiMs in BLM and RMI1 interact with the RPA1 N-terminal OB-fold domain

Next, it was investigated whether the three RPA-binding SLiMs interact with distinct regions of the RPA heterotrimer, or the same region. To answer this, pulldown experiments with peptides derived from SLiMs 1, 3 and 5, in which an aromatic residue was substituted with the synthetic photoreactive amino acid p-benzoyl-L-phenylalanine (BPA),

were carried out. Substitution of tryptophan or phenylalanine with BPA can facilitate site-directed photo-crosslinking between the peptide and the direct binding partner. Subsequent denaturing washes (1% SDS in TBS-T) should disrupt all but the direct binding partner, covalently linked to the peptide (Arnoult et al., 2017). UV crosslinking and subsequent denaturing washes revealed the RPA1 subunit was retained by these peptides (**Fig. 4.5A**), but not RPA2 or RPA3. This suggested all three SLiMs directly interaction with RPA1, but not RPA2 or RPA3. RPA1 contains an N-terminal OB-fold domain with a basic binding cleft known to mediate interactions with multiple DNA damage checkpoint and repair factors including p53, MRE11, ATRIP, RAD9, DNA2, ETAA1 and PRIMPOL (Bass et al., 2016; Bochkareva et al., 2005; Feng et al., 2016; Guilliam et al., 2017; Haahr et al., 2016; Olson et al., 2007; Xu et al., 2008b; Zhou et al., 2015). Comparing the sequences of the RPA-binding SLiMs in BLM/RMI1 with the RPA-binding SLiMs from these proteins demonstrated significant sequence similarity (**Fig. 4.5B**).

To investigate whether BLM and RMI1 might interact with the N-terminal OB fold domain of RPA1, the RPA1-RPA2-RPA3 heterotrimer was synthesised with either a WT or a mutant version of RPA1 in which key residues in the N-terminal OB-fold of RPA1 were mutated (R41E/Y42F; Haring et al., 2007) using *in vitro* translation. These translated protein products were incubated with biotinylated peptides derived from the three SLiMs for peptide pulldowns. Although all three peptides could interact with WT RPA, they could not interact with R41E/Y42F RPA1 (**Fig. 4.5C**).

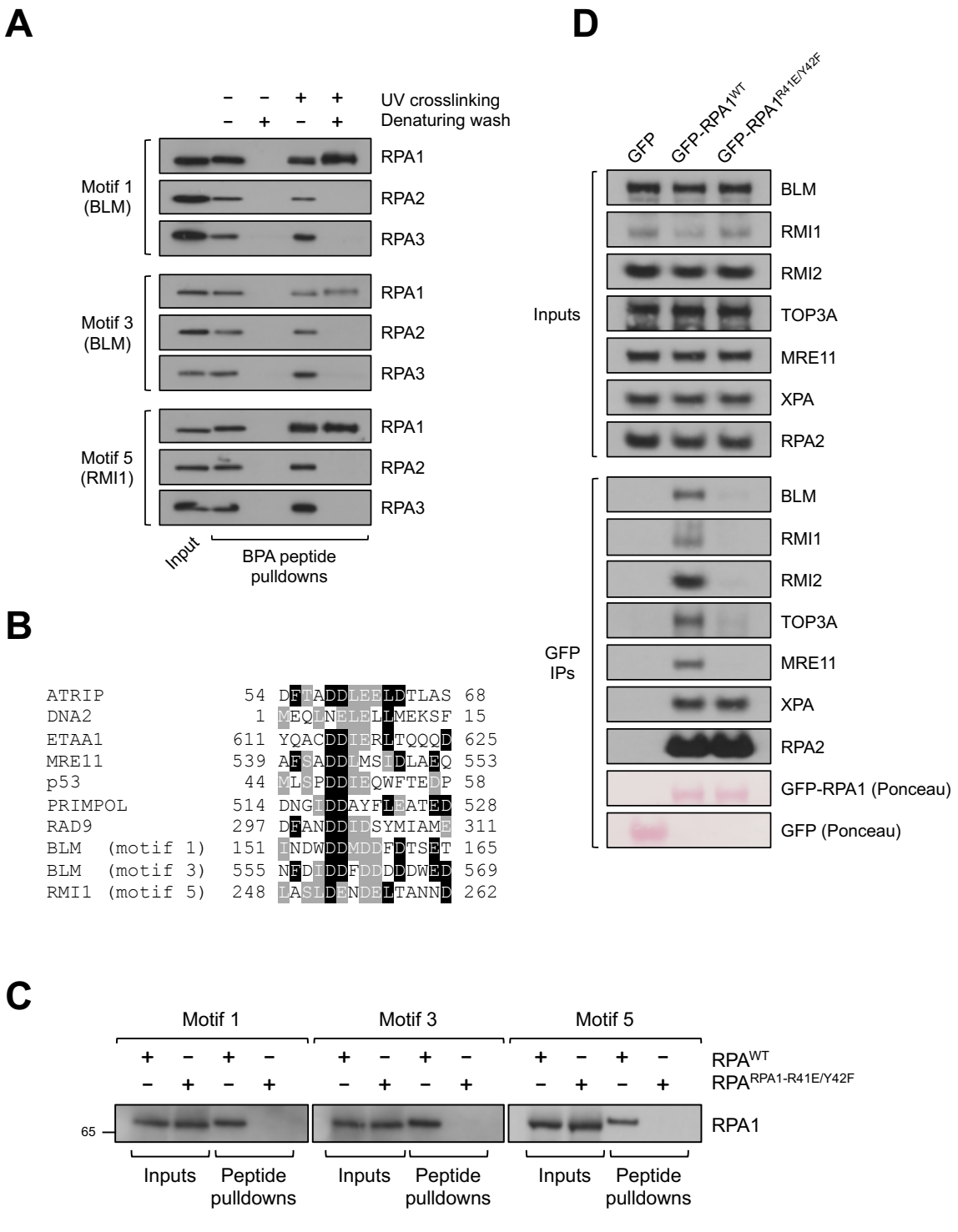


Fig. 4.5. The RPA-binding SLiMs in BLM and RMI1 interact with the RPA1 N-terminal OB-fold. A) BPA-modified peptide pulldowns from HeLa nuclear extracts with UV-crosslinking and denaturing washes where indicated, demonstrating that SLiMs 1, 3 and 5 interact directly with the RPA1 subunit of RPA. B) Sequence alignments showing the similarity of RPA1-binding SLiMs in DNA damage response proteins. C) Peptide pulldowns from rabbit reticulocyte lysates containing in-vitro translated heterotrimeric RPA complexes incorporating either WT or R41E/Y42F RPA1, showing that all three BLM/RMI1 SLiMs cannot interact with the mutant RPA complex. D) GFP-pulldowns from 293FT cells transfected with constructs expressing GFP or the indicated GFP-tagged RPA1 variants, showing that mutation of the N-terminal OB-fold of RPA1 (R41E/Y42F) disrupts binding to the BTR complex and MRE11. XPA and RPA2 serve as negative controls because they bind to RPA via different domains. Experiment conducted by Dr. Andrew Blackford.

To verify the importance of the N-terminal OB fold in the context of full-length RPA1 for the interaction with BLM/RMI1, 293FT cells were transfected with constructs expressing GFP-tagged WT RPA1 or R41E/Y42F RPA1. GFP IPs from lysates from these cells revealed mutating the N-terminal OB-fold of RPA1 inhibited the interaction with MRE11 in accordance with previous work (Olson et al., 2007). The mutation also completely abrogated the interaction between RPA1 and the BTR complex (**Fig. 4.5D**). Importantly, the RPA1 R41E/Y42F mutant was still able to bind to XPA, which associates with RPA via the central region of RPA1 and the C-terminus of RPA2 (Stigger et al., 1998; Walther et al., 1999). This demonstrated that the RPA1-R41E/Y42F mutation does not cause general misfolding of RPA1.

In summary, this data demonstrates that the BTR complex contains three SLiMs (two in BLM and one in RMI1) that interact with the same basic binding cleft in the N-terminal OB-fold of RPA1. Accordingly, it is possible at least three separate RPA complexes associate simultaneously with the BTR complex.

4.3.3 Generation of a Cell-Based System to Test the Physiological Relevance of BTR-RPA binding

To investigate whether RPA binding contributes to any of the functions of the BTR complex, cells expressing mutants of BLM and RMI1 in which all three RPA-binding SLiMs are deleted were required. This required the generation of cells deficient in BLM and RMI1 that could be complemented with RPA-binding mutants. Human retinal pigment epithelial (RPE-1) cells are immortalised by expression of the human telomerase catalytic component, hTERT, as opposed to transformation. Furthermore, they have a normal karyotype and express WT p53 (Bodnar et al., 1998; Jiang et al., 1999). For these reasons this cell line was used as it may best mimic normal cellular physiology, and with the aim of generating a *BLM/RMI1* knockout line, CRISPR-Cas9 mediated genome editing was attempted.

CRISPR-Cas9 systems function in bacteria and archaea to provide adaptive immunity against foreign nucleic acids, such as viruses and plasmid DNA (Barrangou et al., 2007; Wiedenheft et al., 2012). In recent years, these systems have been harnessed to mediate genome engineering within mammalian cells via the induction of sequence-specific DSBs (Cong et al., 2013; Mali et al., 2013a). In the simplest and most widely used CRISPR-Cas9 system, based upon that of the *Streptococcus pyogenes* type II system, a nuclease (Cas9) can be directed to any genomic locus that is followed by a 5' NGG proto-spacer adjacent SLiM (PAM) by a chimeric single-guide RNA (sgRNA). The sgRNA contains a 20-nucleotide sequence, complementary to the targeted genomic site via canonical Watson-Crick base pairing, which directs the Cas9 nuclease to generate a blunt DSB at the desired genomic locus (**Fig. 4.6A**). DSB repair mechanisms such as NHEJ or HDR induce error-prone or defined alterations at these DSB sites (Mali et al., 2013b).

While the sgRNA should be designed to have high specificity for the target site, similar sequences often exist within large genomes, such as those of mammalian cells. Aiming to improve the specificity of Cas9-mediated genome editing, Ran *et al.,* developed a modified strategy combining a mutated form of Cas9, Cas9-D10A, with a pair of offset sgRNAs complementary to opposite strands of the target site. When directed to the target DNA sequence, with one active nuclease domain deactivated, Cas9-D10A cuts only one strand of DNA. As single-stranded breaks are repaired with high fidelity, simultaneous nicking via appropriately offset sgRNAs is required for DSB formation, effectively doubling the number of specifically recognised bases that are required for target cleavage. Accordingly, this system reduces off-target activity by 50- to 1,500-fold in cell lines without sacrificing on-target cleavage efficiency (Ran et al., 2013; **Fig. 4.6A**).

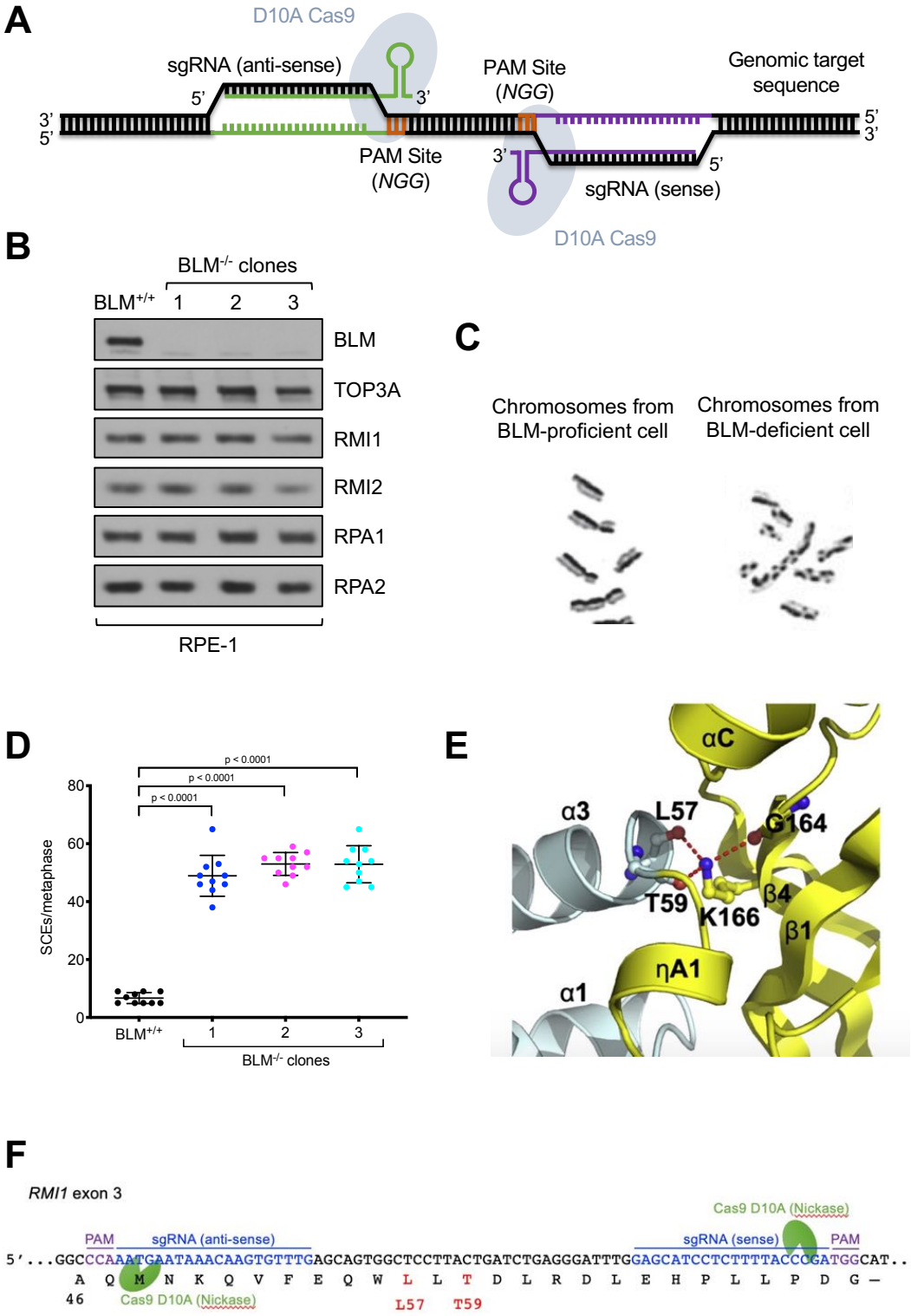


Fig. 4.6. Generation of BLM/RMI1-deficient RPE-1 cells

A) Schematic of CRISPR-Cas9 mediated genome engineering (Cong et al., 2013; Mali et al., 2013). B) Western blots showing that BLM expression is undetectable in three separate BLM knockout RPE-1 clones. Experiment conducted by Dr. Kaima Tsukada. C) Comparison of typical chromosomes and chromosomes containing SCEs. D) SCE assay showing that all three BLM-/- clones have very high levels of SCEs (n=3, one representative experiment shown). Significance was determined using the Mann-Whitney U test. Experiment conducted by Dr. Samuel Jones & Dr. Kaima Tsukada. E) RMI1N crystal structure demonstrating that K166 forms hydrogen bonds with three other amino acids within the N-terminus of RMI1: L57, T59 and G164. These interactions, amongst others, are proposed to stabilize the structural integrity of the OB-fold. Reprinted from Structure, 18, Wang, F., Yang, Y., Singh, T.R., Busygina, V., Guo, R., Wan, K., Wang, W., Sung, P., Meetei, A.R., and Lei, M. (2010), Crystal structures of RMI1 and RMI2, two OB-fold regulatory subunits of the BLM complex, 1159–1170, Copyright (2010), with permission from Elsevier. F) sgRNAs were designed to disrupt the OB-fold of RMI1 by targeting Cas9 to the vicinity of amino acids L57 and T59 (highlighted).

Three independently derived BLM-deficient RPE-1 clones, with no detectable BLM protein expression were further characterized (**Fig. 4.6B**). To identify the mutations present, genomic DNA (gDNA) from these three *BLM*$^{-/-}$ RPE-1 clones was amplified using primers flanking the predicted Cas9 cleavage site in *BLM* exon 4. Purified PCR products were TOPO-cloned, and their sequences were analysed by Sanger sequencing. Both alleles in all three clones had frameshift mutations, generating a premature stop codon (**Table 4.1**).

Table 4.1. BLM allele sequencing

Clone	Allele	Mutation	Protein
1	1	NM_000057.4:c.946_947insTCTGCTTCTTCTTCCTCTTTTATTTCTGCTTCTTCTTCCTCTTCAAAATGCATTTCTGCTTCTTCTTCCTCTTTCTGCTTCTTCTTCCTCTTTTATTTCTGCTTCTTCTTCCTCTT	p.(Ser316PhefsTer46)
1	2	NM_000057.4:c.883_931delinsTATGATACGGATTTTGTTCCACCTTCTCCAGAAGAAATTATTTCTTTCTGCTTTCTGCTTCTTCTTCCTCTTATTTCTTTCTGCTTTCT	p.(Asp295TyrfsTer42)
2	1	NM_000057.4:c.825_951delinsATTTCTGCTTCTTCTTCCACGTT	p.(Glu276PhefsTer10)
2	2	NM_000057.4:c.954_955insTGCTTCTTCTTCCTCTTCAAAATGC	p.(Leu319CysfsTer10)
3	1	NM_000057.4:c.947_948insTGCTTCTTCTTCCTCTTCAAATTTCTGCTTCTTCTTCCTCTTCTGCTTCTTCTTCCTCTTC	p.(Lys317AlafsTer24)
3	2	NM_000057.4:c.948_949insGAGAAATTATTTCTGCTTCTTCTTCCTCTTCACAGAAGAAATTATTTCTGCTTCTTCTTCCTCTTCA	p.(Lys317GlufsTer26)

To verify whether they were functional knockouts, the number of SCEs was measured in the three *BLM*$^{-/-}$ RPE-1 clones (**Fig. 4.6C,D**). The SCE assay differentially stains sister chromatids, allowing for microscopic detection of the physical exchange of DNA that occurs with crossover events between sister chromatids. Of note, the standard protocol employed here detects spontaneous crossover events rather than those induced by DNA damage (Stults et al., 2020). Each clone had a >10-fold increase in SCEs (**Fig. 4.6D**), similar to that of Bloom syndrome patient cells, *Blm*$^{-/-}$ mouse embryonic fibroblasts and BLM knockout chicken DT40 cells (Chester et al., 1998; Wang et al., 2000). Together, this provided genetic and functional confirmation that these clones are true *BLM* knockout cells.

To investigate the physiological relevance of RPA binding to the BTR complex, cells deficient in both BLM and RMI1 were required. To generate double *BLM/RMI1* double knockout RPE-1 cells, two sgRNAs designed to target offset sequences within exon 3 of *RMI1* were cloned into an All-in-One plasmid vector encoding the Cas9-D10A nickase and a GFP marker (Chiang, 2016). This approach circumvented compromised transfection efficiency and thus efficiency of mutagenesis by the co-transfection of multiple plasmids. This plasmid was transfected into *BLM*$^{-/-}$ RPE-1 cells, then 72 hours later GFP-positive cells were single cell sorted into 96-well plates and allowed to grow into colonies. Fifty clones were analysed for RMI1 expression via western blot, but all clones demonstrated at least low levels of RMI1 expression, indicating this attempt at generating *BLM/RMI1* double knockout RPE-1 cells was unsuccessful despite successful editing of the *RMI1* locus. This is in line with previous studies indicating that complete loss of RMI1, or its stable binding partner TOP3A, is lethal in vertebrate cells (Chen et al., 2011; Li and Wang, 1998; Mönnich et al., 2010), unlike BLM.

Whilst no clones demonstrated a complete loss of RMI1 expression, one clone (*BLM*$^{-/-}$ *RMI1*$^{mut/mut}$) had extremely low levels of RMI1 when compared to WT (**Fig. 4.7B**). RMI1 interacts with both BLM and TOP3A via its N-terminal OB1-fold domain and this N-terminal third of RMI1 (RMI1N) is just as adept as the full-length protein at promoting dHJ dissolution and HJ unwinding *in vitro* (Raynard et al., 2008). Furthermore, a point mutation in RMI1 (K166A) that disrupts the interaction between TOP3A and RMI1 is unable to promote dHJ dissolution and HJ unwinding by BLM-TOP3A. This suggests that the functional integrity of the BTR complex is dependent on the interaction of RMI1 with TOP3A. Analysis of the RMI1N crystal structure demonstrated that K166 forms hydrogen bonds with three other amino acids within the N-terminus of RMI1: L57, T59 and G164 (**Fig. 4.6E;** Wang et al., 2010). These interactions, amongst others, were proposed to stabilise the structural integrity of the

OB-fold. Accordingly, disruption of this interaction network could result in a conformational change in the N-terminus of RMI1, abrogating the interaction between RMI1 and TOP3A and compromising the functional integrity of the BTR complex. The sgRNAs described above were designed to disrupt the OB-fold of RMI1 by targeting Cas9 to the vicinity of amino acids L57 and T59 (**Fig. 4.6F**), so the possibility was considered that the residual RMI1 expressed within this clone had been rendered functionally inactive.

To investigate the nature of the mutation(s) present within the clone, *BLM*$^{-/-}$ *RMI1*$^{mut/mut}$ gDNA was extracted and amplified using primers designed to flank the predicted Cas9 cleavage site in *RMI1* exon 3. The purified PCR product was TOPO cloned, and sequences were analysed by Sanger sequencing. Although one allele contained a point mutation in the critical residue, T59, the only other mutations were in-frame insertions in each allele. Accordingly, the impact of these mutations to the structural integrity of the OB1-fold in both alleles remained unclear.

As with BLM-deficient cells, RMI1-deficient cells display a large increase in the level of SCEs (Wu et al., 2006). However, this potential BLM/RMI1-deficient clone would be expected to display the characteristic increase in SCEs as it was derived from a *BLM*$^{-/-}$ cell line, irrespective of RMI1 function. Therefore, to assess the function of the residual RMI1 protein expression, BLM protein expression had to be restored within this clone. To do this, GFP-BLM was sub-cloned into a lentiviral vector under the control of a PGK promoter. The clone was then infected with viruses to stably express either WT GFP-BLM or GFP alone as control. GFP-positive cells were selected in G418 and then isolated using FACS. Once a pure population was obtained, the level of SCEs was assessed. While this clone complemented with GFP only displayed the expected 10-fold excess of SCEs and was comparable to the parental *BLM*$^{-/-}$ line, cells complemented with exogenous WT GFP-BLM had levels of SCEs similar to

WT RPE-1 cells (data not shown). These results suggest that the residual RMI1 protein expression is indeed functional, thus this clone could not be used for further assays.

To establish a complementation system to test the requirement for RPA association with the BTR complex, a lentiviral vector was designed expressing GFP-tagged BLM and FLAG-tagged RMI1 from the same promoter, separated by a P2A self-cleaving peptide sequence, to produce two separate polypeptides (**Fig. 4.7A**). The RPA-binding SLiMs in BLM and RMI1 were deleted from this plasmid using site-directed mutagenesis. *BLM*$^{-/-}$ cells were then infected with viruses to express either WT GFP-BLM and FLAG-RMI1, RPA-binding mutant versions of these proteins (BLM/RMI1$^{\Delta RPA}$), or GFP alone as control. By transiently depleting RMI1 using siRNA, cells that do not detectably express endogenous BLM or RMI1, and cells expressing only exogeneous WT or RPA-binding mutant versions of these proteins, were established for use in functional assays (**Fig. 4.7B**). Within the lentiviral vector expressing FLAG-tagged RMI1, the RMI1 DNA sequence was mutated to introduce silent mutations that did not affect the protein sequence but disrupted siRNA binding. Accordingly, the siRNA depleted endogenous RMI1 but not RMI1 expressed ectopically. Of note, levels of RMI1 post siRNA depletion were lower than the levels of RMI1 in *BLM*$^{-/-}$ *RMI1*$^{mut/mut}$, demonstrating the use of this method for functional assays (**Fig. 4.7B**). An alternative method that could have been used to deplete RMI1 is one based on inducible degradation of RMI1, such as using the auxin-inducible degron system (Nishimura et al., 2009)This method circumvents some of the limitations of RNAi, such as off-target effects and the fact that protein depletion is indirect, and thus depending on the stability of the target protein. Furthermore, it induces rapid depletion, making it possible to observe and immediate phenotype. However, the auxin-inducible degron system has two major drawbacks: leaky degradation and the requirement of a high dose of auxin, making it difficult to control the expression level of the protein of interest in living cells.

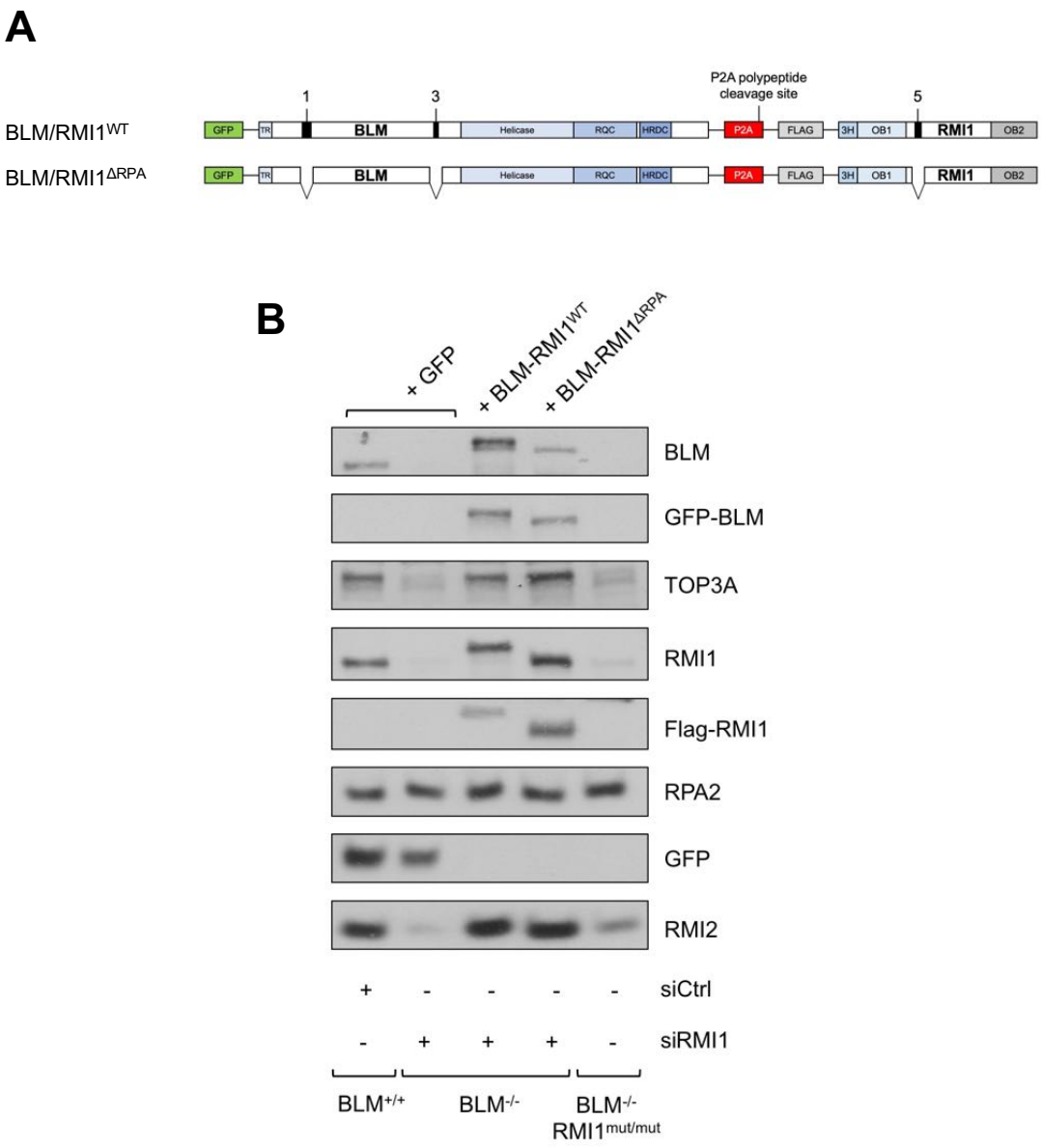


Fig. 4.7. Generation of a cell-based system to test the physiological relevance of BTR-RPA binding.
A) Schematic showing the complementation strategy for adding back WT BLM/RMI1 proteins or mutant ones lacking RPA-binding SLiMs to *BLM*$^{-/-}$ cells treated with siRMI1. B) Western blots showing the levels of exogenous GFP and BLM/RMI1 proteins delivered by lentivirus in WT or *BLM*$^{-/-}$ cells treated with siCtrl or siRMI1 as indicated. The RMI1 protein level in *BLM*$^{-/-}$ *RMI1*$^{mut/mut}$ cells was found to be lower than the RMI1 protein level in WT cells.

4.3.4 RPA Binding Stabilises BLM at ssDNA Sites but is not Essential for BLM Recruitment to DSBs

As RPA recruits multiple DNA damage response proteins in to sites of DNA damage in response to genotoxic stress, it was considered whether RPA binding promotes recruitment of

the BTR complex to DNA damage sites (Bhat and Cortez, 2018). To test this hypothesis, laser micro-irradiation was used to generate RPA-coated ssDNA. The recruitment of GFP-BLM^{WT} and GFP-BLM$^{\Delta RPA}$ was examined (**Fig. 4.8A**). Initially, GFP-BLM^{WT} and GFP-BLM$^{\Delta RPA}$ accumulated rapidly to laser lines (within 5 min), but with lower intensity and a more diffuse staining pattern compared to later timepoints. RPA recruitment was not detectable at this early time point, explaining the similarity between the recruitment patterns of GFP-BLM^{WT} and GFP-BLM$^{\Delta RPA}$. Early recruitment of BLM to DSBs is dependent on the C-terminus of BLM rather than the N-terminal region containing the RMI1 and RPA interaction sites (Karmakar et al., 2006). By 15 min, RPA microfoci appeared along the laser line. By 20 min, GFP-BLM^{WT} formed bright foci along the laser line that colocalised with RPA. In contrast, the localisation pattern for GFP-BLM$^{\Delta RPA}$ was diffuse and a lower intensity, similar to the early recruitment pattern. GFP-BLM$^{\Delta RPA}$ formed a small number of foci along the laser line at later time points, they rarely colocalised with RPA. This suggests one function of RPA binding is to stably recruit the BTR complex to sites of ssDNA damage.

4.3.5 RPA Binding is Required for the BTR Complex to Restart Stalled Replication Forks but not to Suppress SCEs

Having demonstrated stable BLM recruitment to RPA-ssDNA sites is defective in absence of the RPA-binding SLiMs in BLM/RMI1 (**Fig. 4.8A**), it was investigated whether any of the biological processes that depend on BLM were defective in cells expressing a BTR complex with abrogated RPA binding.

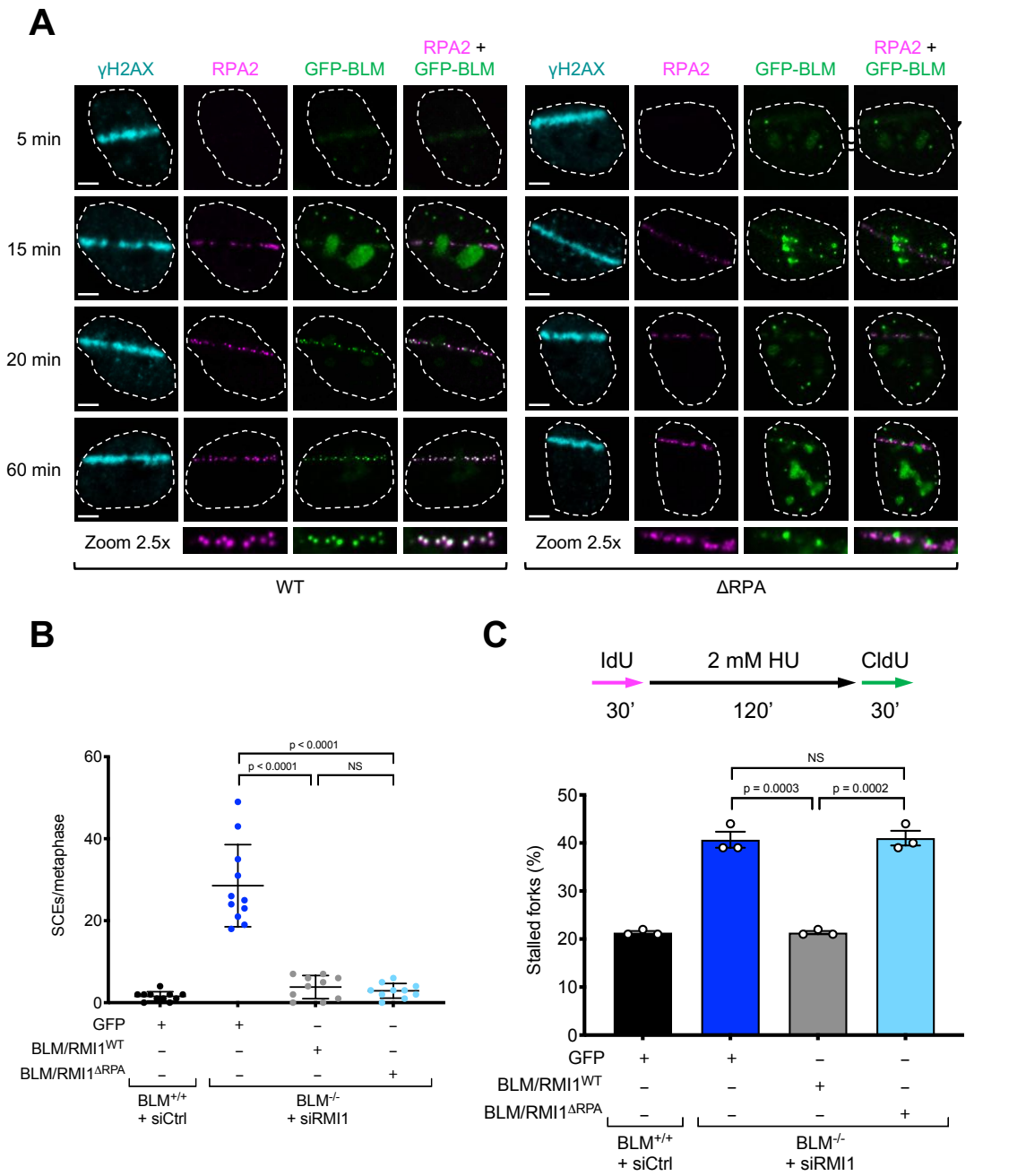


Fig. 4.8. RPA binding is not required for the BTR complex to suppress SCEs but is required for the restart of stalled replication forks.

A) Airyscan high-resolution confocal microscopy image time-course, showing the kinetics of recruitment of γH2AX, RPA2 and WT/mutant GFP-BLM proteins to laser-induced DSBs in representative cells. Zoomed-in images of laser lines are from the 60-min time-point images. Scale bars, 5 µm. Experiment conducted by Dr. Andrew Blackford. B) RPA binding by BLM and RMI1 is not required for suppression of SCEs (n=3, one representative experiment shown). Error bars denote standard deviation from the mean. Significance was determined using the Mann-Whitney *U* test. Experiment conducted by Dr. Samuel Jones & Dr. Kaima Tsukada. C) Experimental outline of the DNA fibre assay to measure replication fork restart. Quantification of the replication fork restart defect demonstrating that RPA binding by BLM and RMI1 is required to rescue the replication fork restart defect in BLM$^{-/-}$ cells. The number of IdU-positive, CldU-negative tracts was quantified as a percentage of total fibres. Significance was determined by unpaired t-test (n=3, error bars denote SEM). D) Quantification of CldU/IdU tract length ratios in DNA fibre assays showing that RPA binding by BLM and RMI1 is required for replication fork restart. Significance was determined using the Mann-Whitney U test (n=3, one representative experiment shown).

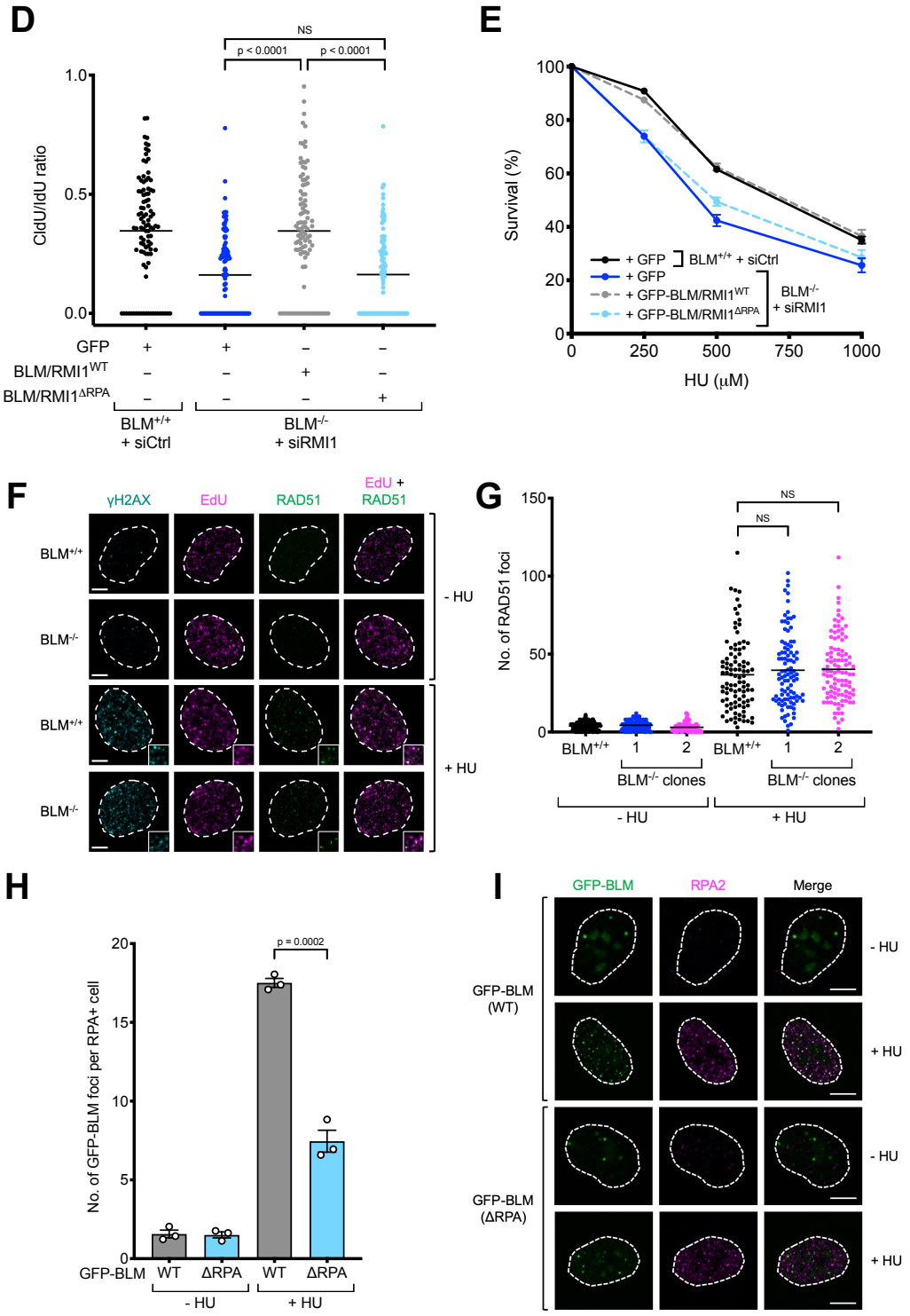


Fig. 4.8. RPA binding is not required for the BTR complex to suppress SCEs but is required for the restart of stalled replication forks.

E) Colony survival assays showing RPA binding by BLM and RMI1 is required for HU resistance (n=3, error bars denote SEM). Cells were treated with the indicated doses of HU for 24 hr. Experiment conducted by Zoulikha Belblidia. F) RAD51 foci are not affected by BLM loss in RPE-1 cells. Cells were pulsed with 10 μM EdU for 15 min to mark sites of active DNA replication and treated where indicated with 1 mM HU for 24 hr. G) Quantification of RAD51 foci in Cyclin A-positive cells, treated where indicated with 1 mM HU for 24 hr (n=2, one representative experiment shown). Significance was determined using the Mann-Whitney *U* test. H) Quantification of the number of WT and ΔRPA GFP-BLM foci in the experiment shown in c) (n=3, error bars denote SEM). I) Immunofluorescence confocal microscopy images of the indicated proteins, demonstrating that the RPA-binding BLM mutant (ΔRPA) is defective in its ability to form foci that colocalize with RPA2 in response to replication stress. Cells were treated where indicated with 2 mM HU for 24 hr.

To assess whether RPA binding to the BTR complex is required for dHJ dissolution, the number of SCEs in cells expressing RPA-binding mutant BLM/RMI1 proteins was analysed. *BLM*$^{-/-}$ cells depleted of RMI1 and complemented with GFP alone displayed the expected 10-fold increase in SCEs compared to WT, but cells complemented with either WT or the RPA-binding mutant BLM/RMI1 proteins had levels of SCEs similar to WT RPE-1 cells (**Fig. 4.8B**). Thus, the interaction between RPA and the BTR complex is not required for SCEs suppression. This can be explained by the fact that dHJs do not contain ssDNA and the earlier observation that GFP-BLM$^{\Delta RPA}$ can still localise to DSBs, although at a lower level. The requirement for RPA binding for UFB processing and for DNA-end resection was assessed and it was demonstrated that BLM-RPA binding is not required for the role of the BTR complex in promoting either cellular function (Shorrocks et al., 2021). To test the role of the BLM-RPA interaction in DNA end resection, a fluorescence-activated cell sorting (FACS)-based assay was used to measure RPA recruitment to chromatin in response to CPT treatment (Forment et al., 2012).

Some BLM-deficient cell lines are reportedly hypersensitive to DNA replication perturbing agents such as hydroxyurea (HU; Davies et al., 2004; Fukao et al., 2004; Pichierri et al., 2004), but not all studies have found this (Imamura et al., 2001). HU inhibits the ribonucleotide reductase (RNR) subunit, RRM2 (Navarra and Preziosi, 1999). RNR catalyses the conversion of rNTPs to dNTPs, which are necessary for replicative and repair DNA synthesis and when RNR is inhibited replication fork stalling occurs. The increased sensitivity of BLM-deficient cells to HU suggests BLM may promote the maintenance of genome stability by protecting cells against DNA replication stress. The DNA fibre technique allows for qualitative and quantitative *in vivo* analysis of DNA replication and involves the labelling of actively progressing replication forks via sequential incorporation of two different halogenated nucleotides, IdU (5-iodo-2′-deoxyuridine) and CldU

(5-chloro-2′-deoxyuridine), into newly replicated DNA strands and their detection via immunofluorescent staining. This technique can analyse the efficiency of replication fork restart after exposure to agents interfering with DNA replication using sequential pulsing with IdU, followed by drug treatment and then CldU labelling (**Fig. 4.8C**, Nieminuszczy et al., 2016). Using the DNA fibre assay, BLM-deficient cells are defective in replication fork restart after exposure to HU (Davies et al., 2007; Schlacher et al., 2012). Therefore, the requirement for RPA binding in BTR complex-mediated replication fork restart was investigated. Using DNA fibre analyses as described above, $BLM^{-/-}$ cells depleted of RMI1 and complemented with GFP alone had a 50% defect in fork restart (**Fig. 4.8C, D**), corroborating earlier reports. Representative images of DNA fibres can be found in section **8.5**. In contrast to WT proteins, which could complement replication fork restart in BLM/RMI1-deficient cells, complementation with the BLM/$RMI1^{\Delta RPA}$ mutant were as defective in fork restart and displayed a similar increase in stalled forks after HU as cells lacking BLM/RMI1 expression altogether (**Fig. 4.8C, D**).

A failure to restart stalled replication forks can result in HU sensitivity, implying that the gene-of-interest is important for genome stability and cellular survival in the face of replication stress. Corroborating a subset of published reports referenced above, $BLM^{-/-}$ cells depleted of RMI1 and complemented with GFP alone displayed HU hypersensitivity. As with the DNA fibre analyses, the RPA-binding BLM/RMI1 mutant did not complement the HU hypersensitivity of BLM/RMI1-deficient cells, unlike the WT proteins (**Fig. 4.8E**).

Given that BLM has been suggested to counteract RAD51 loading on ssDNA (Patel et al., 2017), it is possible that increased RAD51 accumulation at stalled forks might explain the HU sensitivity in BLM-deficient cells. However, no significant increase in the number of RAD51 foci was observed after exposure to HU in two independent *$BLM^{-/-}$* clones (**Fig. 4.8F,**

G). This suggests that HU hypersensitivity is not due to the increased RAD51 accumulation at stalled forks.

4.3.6 RPA Binding is Required for the Recruitment of BLM to Stalled Replication Forks

Nucleotide depletion by HU induces helicase-polymerase uncoupling and the accumulation of RPA-ssDNA (Byun et al., 2005b). As stable recruitment of BLM to ssDNA foci at DSB sites is promoted by RPA binding (**Fig 4.8A**), the possibility that RPA binding may support the recruitment of BLM to stalled replication forks in response to HU was considered. To test this hypothesis, the recruitment of GFP-BLM^{WT} and GFP-BLM$^{\Delta RPA}$ to stalled forks in response to HU was examined using immunofluorescence microscopy. This demonstrated that there was a significant defect in the ability of GFP-BLM$^{\Delta RPA}$ to form foci induced by HU (**Fig. 4.8H**, **I**).

In conclusion, this suggests that RPA binding is required for the promotion of replication fork restart, but not for its role in the suppression of SCEs, UFB processing or DNA-end resection. Loss of RPA binding thus demonstrates that the functions of the BTR complex in maintenance of genome stability can be separated.

4.3.7 BTR-mediated Replication Fork Restart is Dependent on BLM Helicase Activity and the interaction between BLM and TOP3A/RMI1/RMI2

In addition to RPA, other regions of BLM are known to form complexes with various proteins (**Fig. 4.9A**), including TOP3A/RMI1 via the TOP3A/RMI1-interacting (TR) domain (Blackford et al., 2015) and TOPBP1 via SLiM 2 phosphorylated on S304 (**Fig. 3.22B;** (Blackford et al., 2015)). In addition, BLM is a substrate for the small ubiquitin-like modifier (SUMO) modification, with lysines at K317, K331, K334 and K347 being the preferred sites

of modification (Eladad et al., 2005). Further work identified tandem SUMO-interacting motifs (SIMs) that are essential for SUMO binding, with the modification ΔSIM able to disrupt SUMO binding and inhibit SUMOylation of BLM (Zhu et al., 2008). Interestingly, in budding yeast, Ubc9-mediated SUMOylation of Sgs1 is required to prevent the accumulation of cruciform structures at stalled replication forks (Branzei et al., 2006). Furthermore, it is still unclear what the stoichiometry is of the various subunits within the BTR complex, but BLM for example can dimerize via its DHBN domain (Hodson et al., 2022). Accordingly, within cells the BTR complex could potentially associate with more than three RPA complexes via BLM oligomerization. To gain further insight into the wider requirements for BLM-mediated replication fork restart, other factors such as specific BLM-interacting proteins, higher complex stoichiometry and BLM helicase activity were assessed to determine if they are required for the BTR-complex to process stalled replication forks.

To assess the wider requirements for fork restart, the following regions were mutated individually in constructs expressing GFP-tagged BLM via site-directed mutagenesis to generate: ΔTR (lacking the TOP3A/RMI1 interacting region), ΔSIM, S304A (disrupting the TOPBP1-BLM interaction), ΔDHBN (BLM dimerization domain), as well as K695R (a helicase-dead mutant). K695R mutant BLM is devoid of ATPase and helicase activity *in vitro,* with the mutant showing no more than 1% of the WT level of ATP hydrolysis (Bussen et al., 2007). Lastly, data presented in section 4.3.5 demonstrated that deletion of all three RPA binding motifs caused a similar fork restart defect to BLM deficiency (**Fig. 4.8C, D**). To gain insight into the number of RPA complexes required for BTR-mediated replication fork restart, a mutant disrupting the RPA-binding SLiMs in BLM alone (Δ13) was generated.

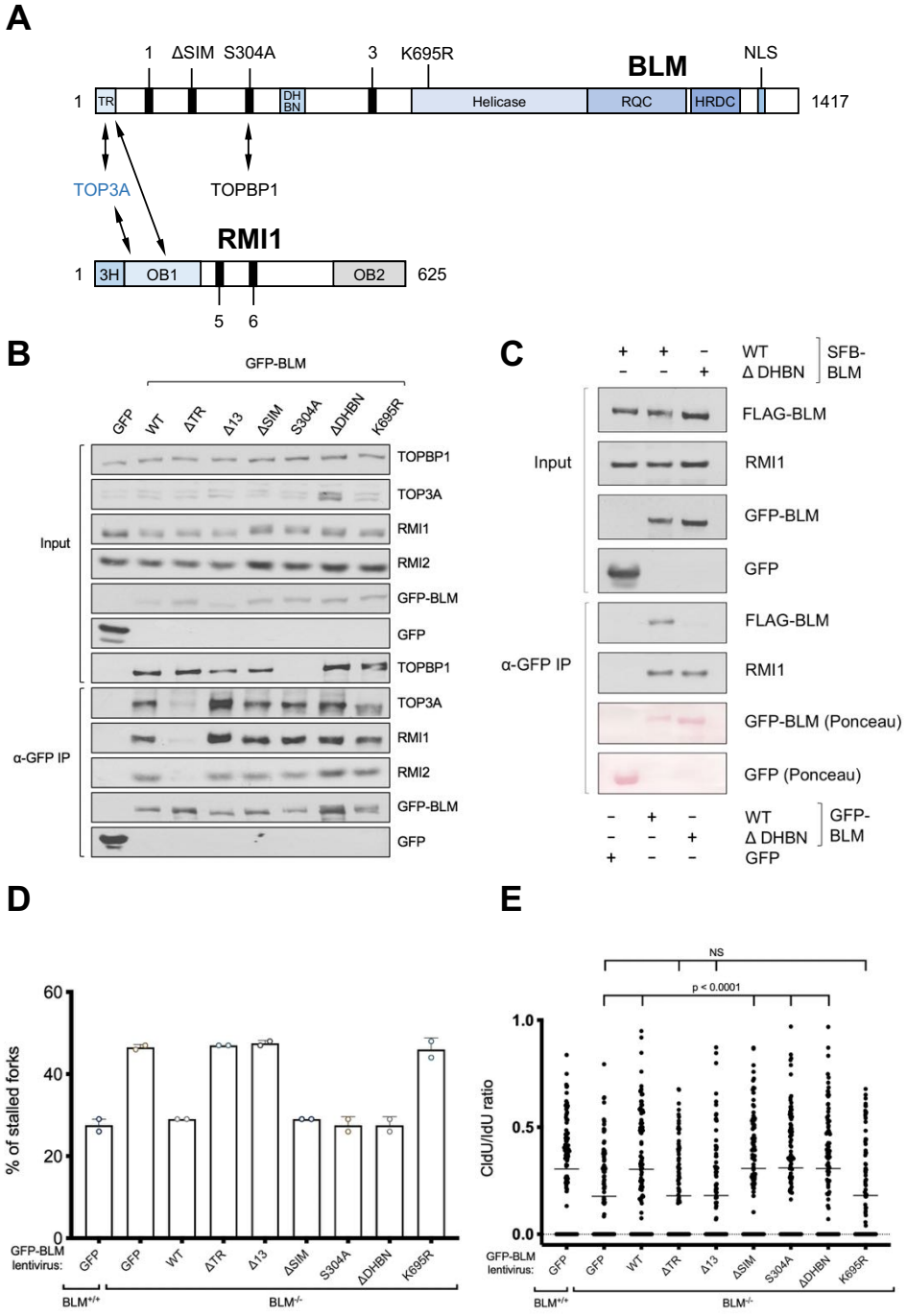


Fig. 4.9. BTR-mediated replication fork restart is dependent on a physical interaction between BLM and TOP3A/RMI1/RMI2 and its helicase activity.
A) Schematic of proteins, domains and SLiMs in the BTR complex. B) GFP-pulldowns from 293FT cells transfected with constructs expressing GFP or the indicated GFP-tagged BLM variants, showing that ΔTR disrupts RMI1, RMI2 and TOP3A binding and S304A disrupts TOPBP1 binding. All other interactions are unaffected. C) GFP-pulldowns from 293FT cells transfected with constructs expressing either GFP or the indicated GFP-BLM variants, and the indicated FLAG-tagged BLM variants, showing how loss of the DHBN domain disrupts BLM oligomerization. D) Quantification of the replication fork restart defect demonstrating that the TOP3A/RMI1/RMI2 interaction and BLM's helicase activity is required to rescue the replication fork restart defect in $BLM^{-/-}$ cells. The number of IdU-positive, CldU-negative tracts was quantified as a percentage of total fibres (n=2, error bars denote SD). E) Quantification of CldU/IdU tract length ratios in DNA fibre assays showing that the TOP3A/RMI1/RMI2 interaction and BLM's helicase activity is required for replication fork restart. Significance was determined using the Mann-Whitney *U* test (n=2, one representative experiment shown).

These mutants were transfected into 293FTs alongside GFP-BLM WT. GFP IPs from lysates harvested from these cells revealed verified that these mutations disrupted only the expected interactions (**Fig. 4.9B**). This demonstrated that the mutations did not cause general misfolding of BLM. The association between BLM and TOP3A, RMI1 and RMI2 was disrupted in 293FT cells transfected with ΔTR, whilst the interaction between BLM and TOPBP1 remained intact. Conversely, as expected, S304A could no longer interact with TOPBP1 but the interaction between BLM and TOP3A, RMI1 and RMI2 remained intact. To assess whether mutating the DHBN domain disrupted BLM dimerization, the DHBN domain was mutated in constructs expressing FLAG-tagged BLM via site-directed mutagenesis. These constructs were co-transfected into 293FT cells with GFP or GFP-BLM variants. GFP IPs from lysates from these cells revealed that loss of the DHBN domain disrupted BLM dimerization (**Fig. 4.9C)**.

To generate stable cell lines expressing these mutants, *BLM*$^{-/-}$ RPE-1 cells were complemented with each of the different mutants. To assess the requirement of each region in promoting BTR complex mediated replication fork restart, DNA replication fibre assays were used. Complementation of *BLM*$^{-/-}$ RPE-1 cells with ΔSIM, S304A or ΔDHBN rescued the replication fork restart defect in BLM-deficient cells and restored the defect in the number of stalled forks after HU to WT levels (**Fig. 4.9D,E**). This suggests that SUMO binding/SUMOylation of BLM, a physical interaction between BLM and TOPBP1 or BLM dimerization is not required to promote replication fork restart. Conversely, complementation of *BLM*$^{-/-}$ RPE-1 cells with ΔTR, Δ13 or K695R did not rescue the defect in replication fork restart in BLM-deficient cells or restore the defect in the number of stalled forks after HU to WT levels (**Fig. 4.9D,E**). This suggests that a physical interaction between BLM and TOP3A/RMI1/RMI2, as well as BLM helicase activity is required to promote replication fork restart. Interestingly, it also suggests that disrupting the RPA-binding SLiMs in BLM (Δ13)

alone is sufficient to disrupt the ability of the BTR complex to promote replication fork restart, even though the RPA binding site in RMI1 remained intact. An important implication of this is that one RPA binding site within the BTR complex is not sufficient to promote replication fork restart alone.

In conclusion, this suggests that BTR-mediated fork restart is dependent on a physical interaction with TOP3A/RMI1/RMI2, as well as the helicase activity of the complex. However, the physical interaction between BLM with SUMO, TOPBP1 and BLM dimerization is dispensable.

4.4 Discussion

To provide insight into which cellular functions contribute to phenotypes associated with Bloom syndrome and the mechanisms by which the BTR complex is regulated *in vivo*, intrinsically disordered regions within the BTR complex were analysed and investigated. In chapter 3, the BTR complex was shown to contain three RPA-binding SLiMs, two in BLM and one in RMI1 (**Fig. 3.22**). Building on these findings, the aims of the work in this chapter were twofold: firstly, to validate these interactions in the context of full-length BLM/ RMI1 and secondly, to identify the functional significance of the interaction between the BTR complex and RPA. In this chapter, all three SLiMs were shown to contribute to the RPA interaction with the BTR complex in the context of full-length BLM/RMI1. Furthermore, they all interact with same site in RPA1, suggesting the BTR complex can associate with three discrete RPA complexes. Then using a combination of CRISPR-Cas9, RNAi and ectopic expression of recombinant proteins delivered by lentivirus, cells expressing a BTR complex defective in RPA binding were generated. Subsequent characterization of these cells demonstrated RPA binding is required for stable recruitment of BLM to ssDNA, replication fork restart and to promote cellular resistance to replication stress, but not for the roles of the

BTR complex ultra-fine bridge resolution or HR. Accordingly, in this chapter the question as to how the BTR complex is recruited and activated in response to replication stress has been addressed.

BLM was shown to interact with RPA using purified recombinant proteins (Brosh et al., 2000) Subsequent studies demonstrated that RPA is present in BTR complexes isolated from cellular extracts, and that the N-terminus of BLM is required for RPA binding (Meetei et al., 2003; Doherty et al., 2005). However, the interaction has not been precisely mapped, with published data being contradictory and contentious. Acidic peptides from BLM (similar to SLiMs 1 and 2 here; **Fig. 3.22A**) were shown to interact with a recombinant N-terminal fragment of RPA1 in a recent report (Kang et al., 2018). However, in this chapter and the one previously, SLiM 2 was shown to interact with TOPBP1 but not RPA (**Fig. 3.22G**). In addition, deletion of this SLiM in the context of full-length BLM did not impact RPA binding (**Fig. 4.3A.B**). Another group could not detect any interaction between recombinant BLM and RPA, and identified an interaction site for RPA on RMI1 (Xue et al., 2013). That site corresponds to SLiM 6 in this chapter (**Fig. 3.22A**). The peptide corresponding to this SLiM was not found to interact with RPA (**Fig. 3.22F**) and deletion of this SLiM in full-length RMI1 does not abrogate on RPA binding, unlike deletion of SLiM 5, which almost abrogates the RMI1-RPA interaction (**Fig. 4.2A,B**). Of note, all cited studies were conducted using purified proteins. Whilst investigation of protein-protein interactions using this approach has much value, these assays (including our own peptide pulldown assay) do not necessary recapitulate *in vivo* mechanisms. Misleading or erroneous results can easily arise if the reaction conditions are not optimal for the interaction and furthermore *in vitro* reconstitution may lack biological relevance given the absence of the spatiotemporal regulation present *in vivo*. In contrast, IPs from cellular extracts results in the maintenance of physiological complexes. Therefore, validation of *in vitro* findings via this

method gives valuable insight into the biological relevance of protein-protein interactions identified *in vitro*.

All three RPA-binding SLiMs in BLM and RMI1 recognize the same region of RPA1, which is known to mediate interactions with multiple DDR factors (Bhat and Cortez, 2018). As all three SLiMs contribute to RPA binding in the context of full-length BLM and RMI1 proteins, this indicates that they are not functionally redundant for binding (**Fig. 4.4C**). Instead, it suggests at least three separate RPA complexes can interact simultaneously with the BTR complex.

The BTR complex has various cellular functions and in this chapter it was investigated whether the interaction between the BTR complex and RPA promotes any of these cellular functions. Some biochemical studies have suggested that RPA binding stimulates BTR complex mediated dHJ dissolution (Xue et al., 2013; Plank et al., 2006), but others have contradicted this (Cejka et al., 2010; Martin et al., 2018). Data in this chapter suggests *in vivo*, RPA binding by the BTR complex is probably not required for dHJ dissolution because cells expressing mutant BLM /RMI1 that cannot interact with RPA do not display increased SCEs (**Fig. 4.8B**). In line with this, dHJ structures do not contain ssDNA and would therefore not be expected to require RPA binding to recruit the BTR complex to dissolve them. In contrast, RPA binding was shown promote replication fork restart and cellular resistance to HU (**Fig. 4.8C-E**).

Investigation of the wider complex requirements that promote replication fork restart demonstrated that BTR-mediated fork restart is dependent on a physical interaction with TOP3A/RMI1/RMI2 (Fig. 4.9D-E). This is consistent with published data, demonstrating that BLM functions with TOP3A, RMI1 and RMI2 to resolve recombination intermediates (Bizard and Hickson, 2014). These recombination intermediates could arise during the processing of stalled replication forks, or during the repair of DSBs following replication fork

Furthermore, like BLM, RMI1 promotes replication fork recovery from aphidicolin-mediated replication stress, in a manner dependent on a physical interaction with TOP3A (Yang et al., 2012). Alternatively, the requirement for the TR domain for BTR-mediated fork restart could be the need for the third RPA binding site in the BTR-complex (SLiM 5 in RMI1). Interestingly, it was observed that disrupting the RPA-binding SLiMs in BLM alone (Δ13) is sufficient to disrupt the ability of the BTR complex to promote replication fork restart, even though the RPA binding site in RMI1 remained intact. An important implication of this is that one RPA binding site within the BTR complex is not sufficient to promote replication fork restart alone. Lastly, the helicase activity of the BTR complex was required for BTR-mediated fork restart. Again, this is in accordance with published reports. *In vitro,* BLM can promote fork regression in an ATP-dependent fashion, suggesting that BLM-mediated fork regression is dependent on its helicase activity (Ralf et al., 2006). In addition, the helicase-dead mutant of BLM, D795A, could not promote the regression of a model replication fork *in vitro* (Machwe et al., 2006). Moreover, *in vivo,* mutation of the ATP binding site (K695T) is as defective in replication fork restart as BLM-deficient cells (Davies et al., 2007). In contrast, SUMO binding/SUMOylation of BLM, a physical interaction between BLM and TOPBP1 or BLM dimerization is not required to promote replication fork restart. This suggests that even if the BTR complex associates with more than three RPA complexes via BLM oligomerization *in vivo*, this is not required for BTR-mediated replication fork restart.

The precise mechanism as to why BLM is required for fork restart is still unclear but could involve fork remodelling by promoting fork regression or unwinding of aberrant DNA structures, such as G-quadruplexes (Sun et al., 1998). In addition to the SWI/SNF helicase-like proteins, BLM can catalyze the regression of a model replication fork in vitro (Machwe et al., 2006). Alternatively, BLM helicase has been proposed to promote replication fork stabilisation and restart by restoring the regressed fork via its Holliday junction migration

activity (Petermann et al., 2010). To identify whether the RPA-BTR complex interaction is required for fork regression or restoring regressed forks, further studies *in vitro* are required to provide key mechanistic insights. Of note, recent work from the Tainer lab has demonstrated that EXO5, a structure-specific nuclease, is phosphorylated by ATR in response to replication stress and forms a complex with BLM and RPA to resect the 5' ends and promote the restart of stalled replication forks (Hambarde et al., 2021). It is tempting to speculate that the function of the BLM-RPA interaction at stalled forks is recruit EXO5, promoting the resection and restart of stalled forks. In support of this, in response to HU treatment, co-depletion of EXO5 and BLM did not enhance cell death, nor did it further reduce replication fork recovery (Hambarde et al., 2021). This suggests the EXO5 and BLM function in the same pathway in response to replication stress.

Even in BLM/RMI1-deficient cells, a proportion of stalled forks restart (**Fig 4.8C,D**). It could be that the requirement for the BTR complex in the restart of stalled forks is context specific. DNA fibre assays could be repeated in the presence of different genotoxic agents that perturb DNA replication to give insight into the nature of lesion(s) the BTR complex acts to resolve. In addition, further studies *in vitro* could provide key mechanistic insight into the nature of the stalled fork structures the BTR complex is active upon. Equally, however, these findings could reflect the functional redundancy that exists with other DNA helicases, such as WRN helicase or the SWI/SNF helicase-like proteins (Machwe et al., 2006). Interestingly, both WRN and SMARCAL1 interact with RPA and both proteins can catalyze the regression of a model replication fork *in vitro* (Ciccia et al., 2009; Yeom et al., 2019). It is therefore tempting to speculate that in that in the absence of the BTR complex, SMARCAL1 or WRN could be recruited to promote replication fork restart.

Data in this chapter gives some insight into the regulation of the BTR complex *in vivo*. It is shown that, mechanistically, a role of RPA binding is to stably recruit the BTR complex to ssDNA sites (**Fig. 4.8A**). Furthermore, the BTR complex contains three discrete SLiMs that bind cooperatively to the N-terminal OB-fold of RPA1. This could suggest a model whereby cells "measure" the length of RPA-coated ssDNA at a lesion such as a stalled or collapsed replication fork and recruit the BTR complex to promote its repair, while preventing aberrant recruitment and activation of BLM at ongoing replication forks where it is not required (**Fig. 4.10**).

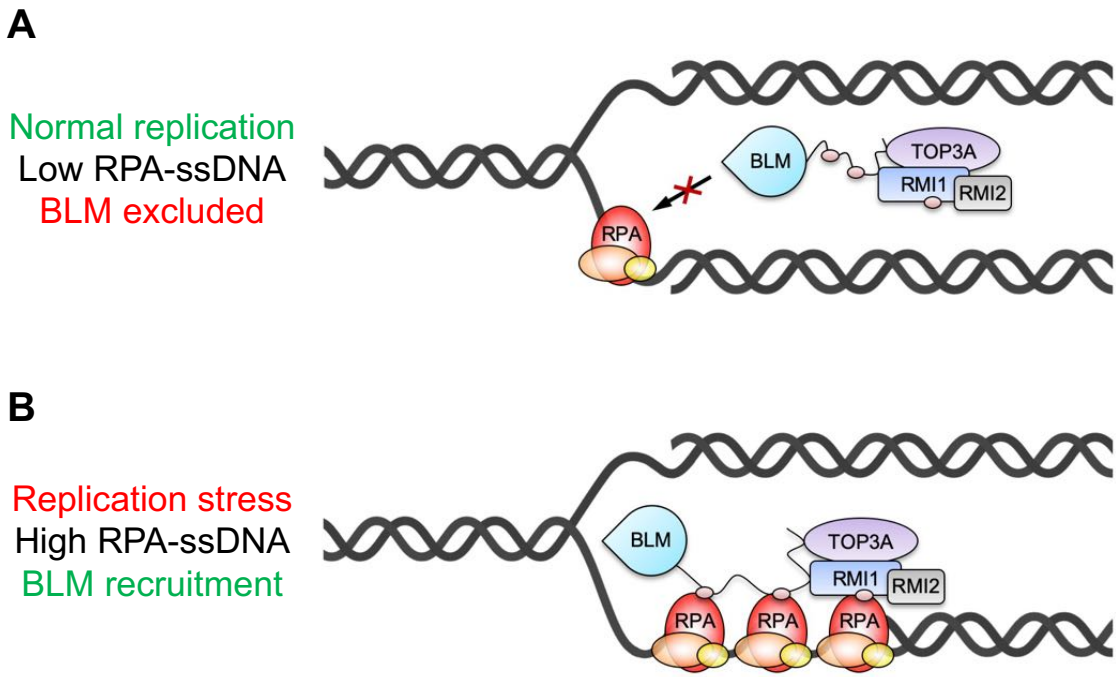


Fig. 4.10. Model for the role of RPA binding in specifically recruiting the BTR complex to stalled replication forks.
A) During normal DNA replication, the BTR complex is not recruited to ongoing replication forks because not enough RPA-coated ssDNA complexes are present. B) When a replication fork stalls, increased RPA-ssDNA leads to stable recruitment of the BTR complex and replication fork restart.

To validate this mechanism, lentiviral vectors expressing GFP-tagged BLM and FLAG-tagged RMI1 would be required, containing individual mutations in SLiM 1, 3 or 5. BLM-/- cells would need to be infected with these viruses, RMI1 depleted by siRNA and the functional assays repeated. If the above mechanism is indeed correct, one would expect complementation with BLM/RMI1 proteins with mutations in the individual RPA binding SLiMs to partially, but not completely, rescue the defect in replication fork restart and cellular resistance to HU.

Bloom syndrome is a heterogeneous condition with many associated pathologies and the BTR complex has a vast array of cellular functions. BLM mutations found in patients result in loss of BLM protein or inhibit its helicase activity, which is required for all its known cellular functions. It is thus unclear which cellular role(s) of BLM are required to prevent the various symptoms of the disease. Furthermore, many questions remain as to how the complex is regulated in vivo.

Together, the data in this chapter demonstrates loss of the interaction between RPA and the BTR complex constitutes a clear separation-of-function mutation with regard to the cellular functions of BLM, given that RPA binding is required for replication fork restart and promoting cellular resistance to HU but not for SCE suppression. A mutation of T99 in the N-terminus of human BLM has shown to cause a defect in replication fork restart but not increase levels of SCEs (Davies et al., 2007). However, T99 is not conserved beyond primates so cannot be used to investigate the role of this residue in animal models. In contrast, further studies in animal models with deletions in the RPA-binding SLiMs in BLM/RMI1 could identify which function(s) of BLM are required to prevent the various pathologies with Bloom syndrome, with possible therapeutic relevance for BLM syndrome patients. Given that BLM deficient cells complemented with a BLM mutant lacking SLiMs 1 and 3 were as defective in fork restart and displayed a similar increase in stalled forks after HU as cells lacking BLM expression altogether (**Fig. 4.9D-E**), this suggests only the RPA-binding SLiMs in BLM would

need to be disrupted to generate a fork restart-deficient mouse model. This would make studies in animal models more logistically possible as only one gene would need to be edited instead of two.

By successfully providing insight into the questions outlined above, work in this chapter and the one before together serve as a proof-of-principle that the bioinformatics and proteomics pipeline is a powerful method to identify functional protein-protein interactions mediated by SLiMs in a high-throughput fashion, as well as provide key insight into the regulation of DNA repair complexes. Given the success of the pilot screen, in Chapter 5, a web-based tool to predict SLiMs automatically was developed.

Chapter 5: Towards a DNA Repair Interactome Based on SLiMs

5.1 Authorship Statement

All coding for the development of the SLiMCon tool was performed by Dr Simon McGowan from the Computational Biology Research Group (CBRG) at the Weatherall Institute of Molecular Medicine (WIMM).

5.2 Introduction

5.2.1 Computational Approaches for Predicting SLiMs

Within members of any given family, protein domains possess well defined three-dimensional structures (Bhowmick et al., 2015). As such, domain signatures are relatively easy to detect from the primary sequence alone using information contained in databases such as Pfam (El-Gebali et al., 2018). Conversely, SLiMs are difficult to identify based on their primary sequence alone due to their low sequence complexity. Given their short size and the fact that generally only 3-4 amino acids confer key specificity and/or affinity determinants (or consensus sequence), the proteome will contain a vast number of sequences matching this consensus sequence (Gibson et al., 2015). This suggests that most of these sequences matching a given consensus sequence in the human proteome cannot be biologically relevant. When the number of false positives greatly exceeds the number of functional SLiMs, the poor signal-to-noise ratio will interfere greatly with the computational discovery of bona fide novel SLiMs. Despite this, several web-based tools are currently available for the prediction of putative SLiMs with characterised specificity determinants. Like the bioinformatics approach outlined

in chapter 3, these computational approaches use known attributes of SLiMs to aid in the prediction of functional SLiMs: the capacity to encode a functional interaction interface within a short sequence (typically up to 30 amino acids); that SLiMs involved in critical processes tend to be conserved over large evolutionary distances and thirdly, that SLiMs tend to be enriched in intrinsically disordered regions of proteins. These web-based tools can be split into methods that scan a single protein with a set of predefined functionally characterised SLiM consensuses (such as the Eukaryotic Linear Motif (ELM) resource (Kumar et al., 2021) and ScanSite (Obenauer et al., 2003); those that identify short conserved sequences in protein sequences spanning vertebrate evolution (such as the Multiple EM for Motif Elicitation (MEME) resource (Bailey et al., 2015) and SLiMPrint (Davey et al., 2012a)). These four SLiM discovery resources are further detailed below.

5.2.1.1 Eukaryotic Linear Motif (ELM)

The Eukaryotic Linear Motif (ELM) resource (http://elm.eu.org) is a manually curated set of SLiMs derived from experimental literature (Kumar et al., 2021). All SLiMs that share a biological function, binding partner or recognition feature are grouped into an "ELM motif class". Each class has a manually created SLiM pattern defining the key specificity and affinity determinants (or consensus sequence) for that class. For example, the MODD_PIKK_ 1 class searches for S/TQ motifs phosphorylated by PIKK family members. ELM motif classes are classified according to six broad categories: cleavage (CLV), degradation (DEG), docking (DOC), ligand (LIG), modification (MOD) and targeting (TRG) SLiMs. As of January 2022, the ELM database contains 317 annotated ELM motif classes.

The user provides the amino acid sequence of any protein of interest in FASTA format, and the tool identifies any putative SLiMs present in the primary sequence provided that match any one of the 317 annotated ELM motif classes. Using example of the MODD_PIKK_1 class

highlighted above, any S/TQ sequences present within the protein sequence will be annotated as a putative PIKK phosphorylation site, irrespective of evolutionary conservation or whether it is located within an intrinsically disordered region of the protein. Accordingly, many putative SLiMs are likely to be false positives. Predictive power can be improved using additional filters based on contextual information, such as the cellular compartment of the filter, evolutionary conservation and structural features of the protein sequence. However, these filters are used at the discretion of the user, and they do not function to eliminate SLiMs, only give the user extra contextual information.

5.2.1.2 ScanSite 4

ScanSite (http://scansite.mit.edu) is a database for phosphopeptide SLiMs (Obenauer et al., 2003). ScanSite4 is the latest version and includes 81 SLiMs with consensus sequences matching targets of mammalian S/T- or Y-protein kinases and phosphopeptide binding domains (SH2, SH3, PBZ, 14-3-3 and PTN domains). As with the ELM database, each class has a manually created SLiM pattern defining the key specificity and affinity determinants (or consensus sequence) for that class. These were defined using oriented peptide libraries, which fixed a central S/T- or Y-residue and random libraries of flanking sequences that were incubated with the domain of interest (Yaffe et al., 2001). Subsequent Edman sequencing of the bound peptides generated the amino acid frequency distribution at each position, which was then converted into a sequence recognition profile. This identified the optimal (highest affinity) binding profile for a given phosphopeptide-binding domain in vivo, but it should be noted that biologically relevant SLiM occurrences may not follow this profile.

The user submits the protein/gene identifier for the protein of interest and the tool identifies any putative SLiMs present in the primary sequence provided that match any one of the 81 annotated ScanSite motif classes. The user can set a stringency filter (low, medium,

high), filtering putative SLiMs based on the idealistic experimental sequence recognition profile. Predicting SLiMs based on primary sequence alone, the ScanSite tool is prone to predicting false positives. However, contextual information is also provided, which can improve the predictive power of the SLiM: surface accessibility (a proxy for intrinsically disordered regions, although less accurate), evolutionary conservation of the SLiM, as well as whether the residue has been reported to be phosphorylated in the literature on databases such as PhosphoSitePlus and PhosphoELM. As with ELM, these filters are used at the digression of the user. Again, they do not function to eliminate SLiMs, only to give the user contextual information.

5.2.1.3 Multiple EM for Motif Elicitation (MEME)

The Multiple EM for Motif Elicitation (MEME) resource (https://meme-suite.org/meme/) is a widely used tool for searching for DNA and protein sequence motifs (Bailey et al., 2015). MEME is based on the expectation maximisation (EM) algorithm that looks for shared sequence patterns within the input (DNA or protein) sequence.

For any protein of interest, the user must submit between two and 50,000 protein sequences in the FASTA format. MEME searches for SLiMs ranging from 6 to 50 residues by default, although the user can specify other values between 2 and 300 residues. Each output motif is assigned a E-value, which refers to the probability of finding an equally well-conserved sequence pattern in random sequences. Accordingly, the lower the E-value, the greater the statistical significance of the observed motif and output motifs are ranked according to their E-value. Using the default parameters, the 3 highest ranking motifs (according to their E-value) are given in the output. Whilst the tool takes in account sequence conservation, it does not consider other contextual information, such as the structural features of the sequence. In accordance with this, this method is prone to false positives, as it does not consider whether

the putative SLiM is in a structured domain or not. Domains could be identified and eliminated by incorporating data from additional resources that predict disordered regions, such as IUPred. However, this is not done automatically and would have to be done by the user.

5.2.1.4 SLiMPrint

SLiMPrint is a SLiM discovery tool that searches for clusters of conserved residues in intrinsically disordered regions of proteins (Davey et al., 2012a). The base hyperlink for SLiMPrint is http://bioware.ucd.ie/rest/slimprints. The user provides a UniProt identifier for a protein of interest and a text file with the SLiMPrint result is returned.

SLiMPrint accesses precomputed orthologue alignments automatically created by the GOPHER (Generation of Orthologous Proteins from High-throughput Estimation of Relationships) algorithm. Poorly aligned regions of the alignments are removed, and the intrinsic disorder propensity and relative local conservation statistics are calculated for each residue in the protein. The tool identifies clusters of conserved residues in intrinsically disordered regions of proteins and each output motif is assigned an E-value. As with MEME, output motifs are ranked according to their E-value. As matches are filtered based on their presence in regions of the proteome that are predicted to be disordered and evolutionary conservation, this tool should be less prone to false positives.

5.2.2 Prediction Performance on SLiMs: Case Study on SLiMs in the BTR Complex

In Chapter 3, three SLiMs in BLM helicase were identified and validated (**Fig. 5.1A**). To assess their accuracy in predicting these SLiMs in BLM helicase, each of the four SLiM discovery algorithms outlined above were trialled (**Table 5.1**). With each method, the default settings were used first to evaluate how well these parameters performed.

Table 5.1. Trial of four existing SLiM discovery algorithms in the prediction of SLiMs in BLM					
SLiMs in BLM helicase		**SLiM discovery algorithm**			
SLiM	**Residues**	**MEME**	**ELM**	**ScanSite 4**	**SLiMPrint**
1	150-163	No	No	N/A	Yes (3/94)
2	290-305	No	Yes	No	Yes (1/94)
3	551-565	No	No	N/A	Yes (2/94)

Using the default settings, SLiMPrint was the only SLiM discovery tool able to detect all three of the SLiMs in BLM that were identified in chapter 3. SLiMs 1, 2 and 3 were ranked 3rd, 1st and 2nd respectively. In addition to these SLiMs, SLiMPrint predicted 91 other SLiMs in BLM, with the vast majority likely to be false positives incidences.

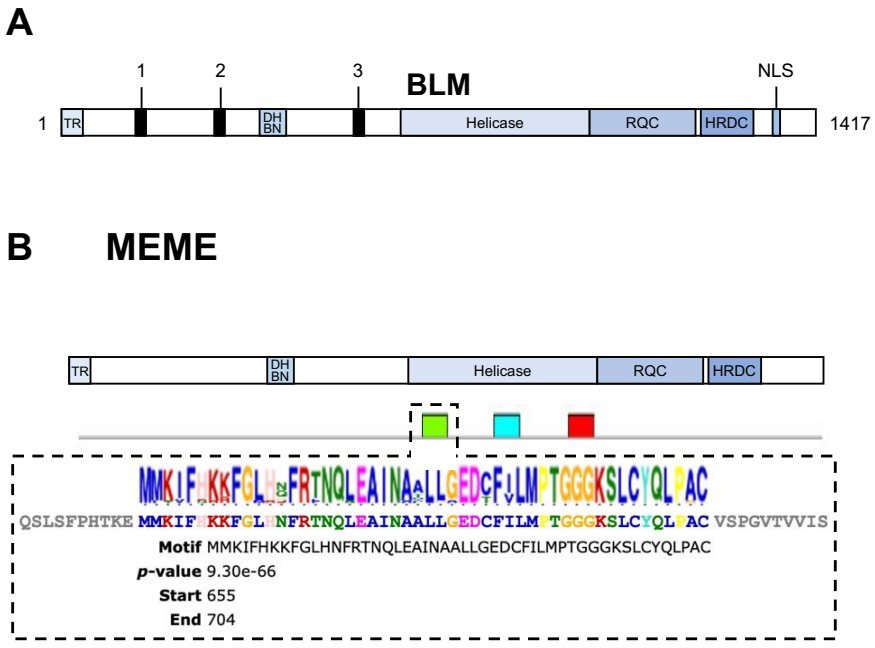


Fig. 5.1. Prediction performance: Output of SLiMs in BLM predicted by the MEME algorithm
A) Three SLiMs in BLM helicase identified and validated in Chapter 3 and 4. B) Output of SLiMs in BLM predicted by MEME using the default parameters. Shown is the prediction of a region in the helicase domain.

MEME was not able to detect any of the SLiMs in BLM using the default settings. The three SLiMs it predicted were found to be in the helicase domain (**Fig 5.1B**). MEME algorithm detects SLiMs based on sequence conservation alone and does not consider regions of intrinsic disorder. This is a flaw with this tool, as it does not consider the structural context of the predicted SLiM. This will result in false positive predictions, as structured regions are highly conserved across evolution. This demonstrates that evolutionary conservation alone cannot predict whether a region of a protein is a SLiM. If the default parameters are modified, it does accurately identify the SLiMs and the structured domains. Furthermore, it is the most time consuming of the tools, as it requires the user to input their own protein sequences (spanning a suitable evolutionary distance).

ELM was only able to detect SLiM 2 (**Figure 5.2**) and unfiltered, the tool predicted 617 SLiMs in BLM from 100 different ELM classes. Using default settings, evolutionary conservation and regions of intrinsic disorder are not considered during the prediction of SLiMs. However, hits can be annotated to highlight which of the predicted hits are outside of structured domains (as annotated according to the SMART tool). This reduced the number of hits down to 287 SLiMs across 40 ELM classes. Evolutionary conservation cannot be applied as a filter, but each predicted SLiM is colour coded with a conservation score. Again, this tool results in the prediction of a large number of SLiMs that are likely to be false positives, as it predicts SLiMs based on sequence alone. As it is only looking for SLiMs that match a specific encoded consensus sequence, it does not consider enough contextual information, such as whether the region is intrinsically disordered or conserved across evolution. It should be noted that even though ELM predicted SLiMs surrounding the region of SLiMs 1 and 3, none of the predicted SLiMs included the acidic residues that are key for the interaction between BLM and the basic N-terminal cleft of RPA1.



Fig. 5.2. Prediction performance: Output of SLiMs in BLM predicted by ELM.
Shown are the 617 predicted SLiMs in BLM from 100 different ELM classes.

Given that ScanSite only detects phosphorylated SLiMs and those interacting with phosphopeptide binding domains, it was not surprising that it could not detect SLiMs 1 and 3 (**Figure 5.3**). However, using their default settings, it also did not predict SLiM 2, despite it being a target for phosphorylation. The user can set a stringency filter (low, medium, high), filtering putative SLiMs based on the idealistic experimental sequence recognition profile. It was only under the lowest stringency filter that SLiM 2 was identified, ranked 165th out of 377 predicted SLiMs (as opposed to the 5 that were predicted with the highest stringency filter). The user can scan orthologues for conservation at that residue but predicted SLiMs are not eliminated according to evolutionary conservation. These things mark flaws of this tool. Whilst it can predict phosphosites that match experimental *in vitro* binding profiles, it only predicts SLiMs of a particular class (phosphorylated SLiMs and those interacting with phosphopeptide binding domains). In addition, given that it scans input sequences using the optimal experimental *in vitro* binding profiles, it is likely to be far too stringent in its higher confidence settings to be biologically relevant. However, lowering the stringency results in an extremely high number predicted SLiMs, most of which are likely to be false positives. Furthermore, again it does not filter hits according to whether the SLiM in located within an intrinsic region of disorder or evolutionary conservation.

This revealed that whilst web-based tools do exist for the prediction of SLiMs, each of the commonly used tools had key flaws. It is interesting to note that the SLiM discovery algorithm that was the most successful in predicting the SLiMs in BLM was SLiMPrints, which is the only one that automatically filters predicted SLiMs based on both evolutionary conservation and the likelihood that it is in a region of intrinsic disorder.

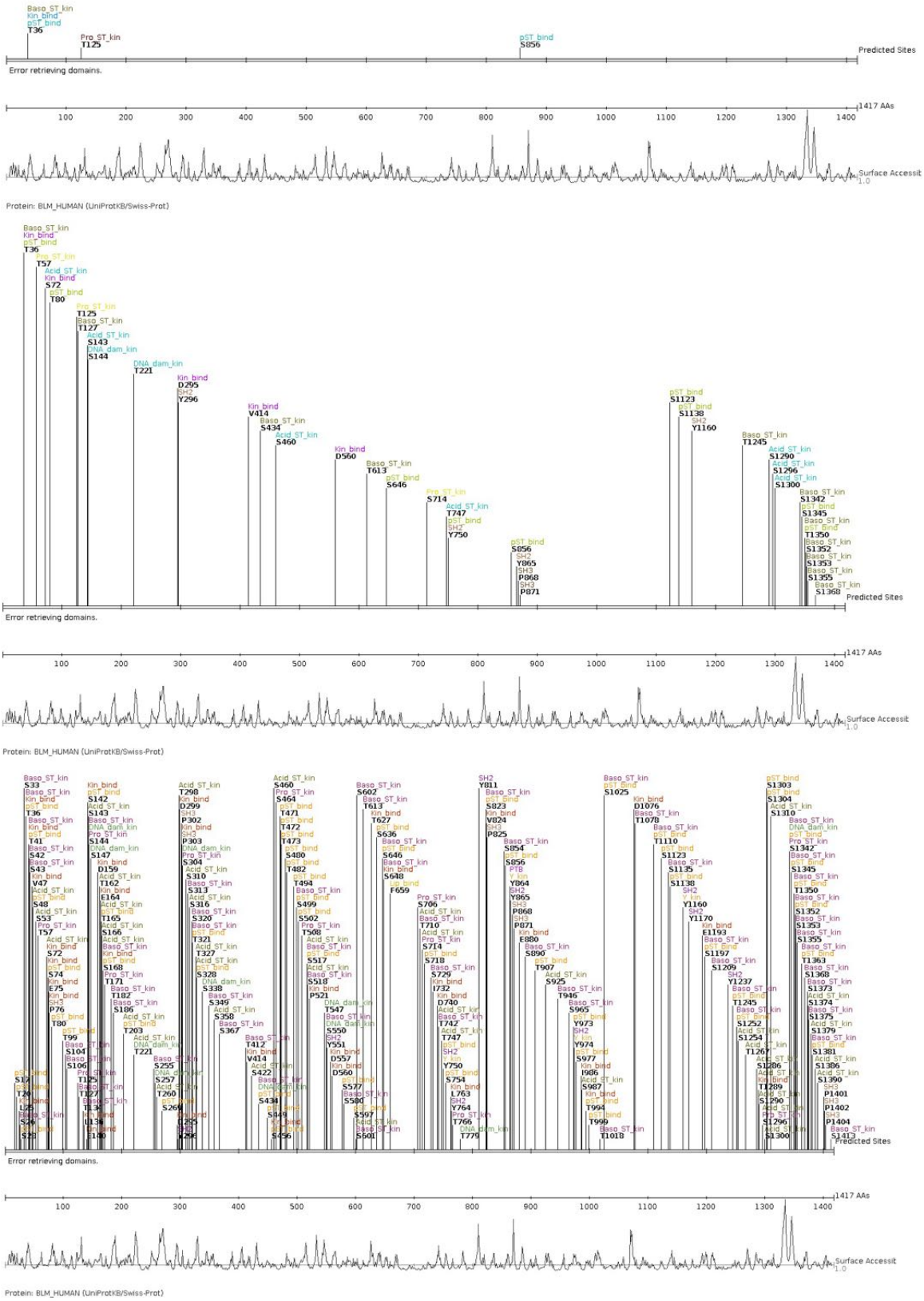


Fig. 5.3. Prediction performance: Output of SLiMs in BLM predicted by ScanSite4.
Shown are the predicted SLiMs in BLM when the stringency filter is varied: low (bottom panel), medium (middle panel) high (top panel).

5.2.3 Aims and Objectives

In Chapter 3, a combined bioinformatics and proteomics approach to identify conserved SLiMs in DSB repair proteins was developed. Although the pipeline outlined was effective, it relied on the time-consuming generation of sequence alignments generated manually (**Fig. 3.2**). As they are labour intensive, if they are being generated on a larger scale, there is a limit to the number of vertebrate sequences that can be acquired and the total number of alignments that can be feasibly generated on a larger scale.

To develop a DNA repair interactome based on SLiMs and overcome the labour-intensive nature of the generation of multiple sequence alignments, an automatic prediction tool was required. However, analysis of existing SLiM discovery algorithms (**Figure 5.1-3**), revealed that these web-based tools had varying success in detecting the validated SLiMs in the BTR complex. Therefore, I aimed to develop a web-based tool to predict SLiMs automatically. This required a method to generate multiple sequence alignments automatically, and then predict SLiMs in these multiple sequence alignments based on the known criteria of SLiMs outlined in chapter 3: short, conserved across a large evolutionary distance and present within regions of intrinsic disorder.

5.3 Results

5.3.1 Generation of Automated Sequence Alignments for Every Protein Encoding Gene in the Human Genome

To develop a web-based tool to predict SLiMs in DDR proteins, the "SLiMCon" tool was generated. First, automated multiple sequence alignments for each protein of interest had to be generated. To do this, the protein sequence for each isoform of the human protein for each

protein of interest was obtained from NCBI (**Fig. 5.4A,B**). The largest isoform for each protein was selected as the default to avoid the use of partial sequences as the human reference sequence (**Fig. 5.4C**). Next, orthologous sequences were acquired from species from representative taxonomic groups, with increasing evolutionary distance from humans. Only orthologues within 10% of the length of the acquired human protein sequence were acquired. Orthologous sequences not fitting this criterion can introduce large insertions or deletions in the multiple sequence alignment, which can occlude the identification of SLiMs in that protein of interest.

With the reference human sequence and orthologues acquired for each protein of interest, an automated sequence alignment was generated using T-Coffee (Notredame et al., 2000). The sequences used for the automated sequence alignment were selected from the acquired orthologues, but 50% of the sequences used were required to be non-mammalian species and include species as evolutionarily distant as fish to ensure sequences span a suitable evolutionary distance. Selection of sequences across a smaller evolutionary distance, may introduce bias and make a region to appear more conserved than it is. With the protein sequences aligned, the multiple sequence alignment can be visualised using JalView (Waterhouse et al., 2009 Fig 5.2E). JalView is alternative free online resource to BoxShade and allows amino acid sequence conservation to be visualised.



Fig. 5.4. Generation of automated sequence alignments for every protein encoding gene in the human genome.
A-B) Automated sequence alignments can be accessed by inputting the gene name of the protein of interest (A), such as BLM (B). C) The protein sequence for each isoform of the human protein for each protein of interest. Orthologous sequences acquired from species from representative taxonomic groups, with increasing evolutionary distance from humans. D) Automated sequence alignment generated using T-Coffee

If the automated sequence alignment was unsuitable for any reason, such as the presence of an insertion in a region of interest, different orthologues could be selected (**Fig 5.4C**) and another alternative alignment generated (**Fig 5.5B**). An example of the sequence alignment automatically generated for BLM can be found in section 8.6. Automated sequence alignments have been generated for every protein coding gene in the human genome, which can be accessed by inputting the gene name of the protein of interest (**Fig 5.4A**).

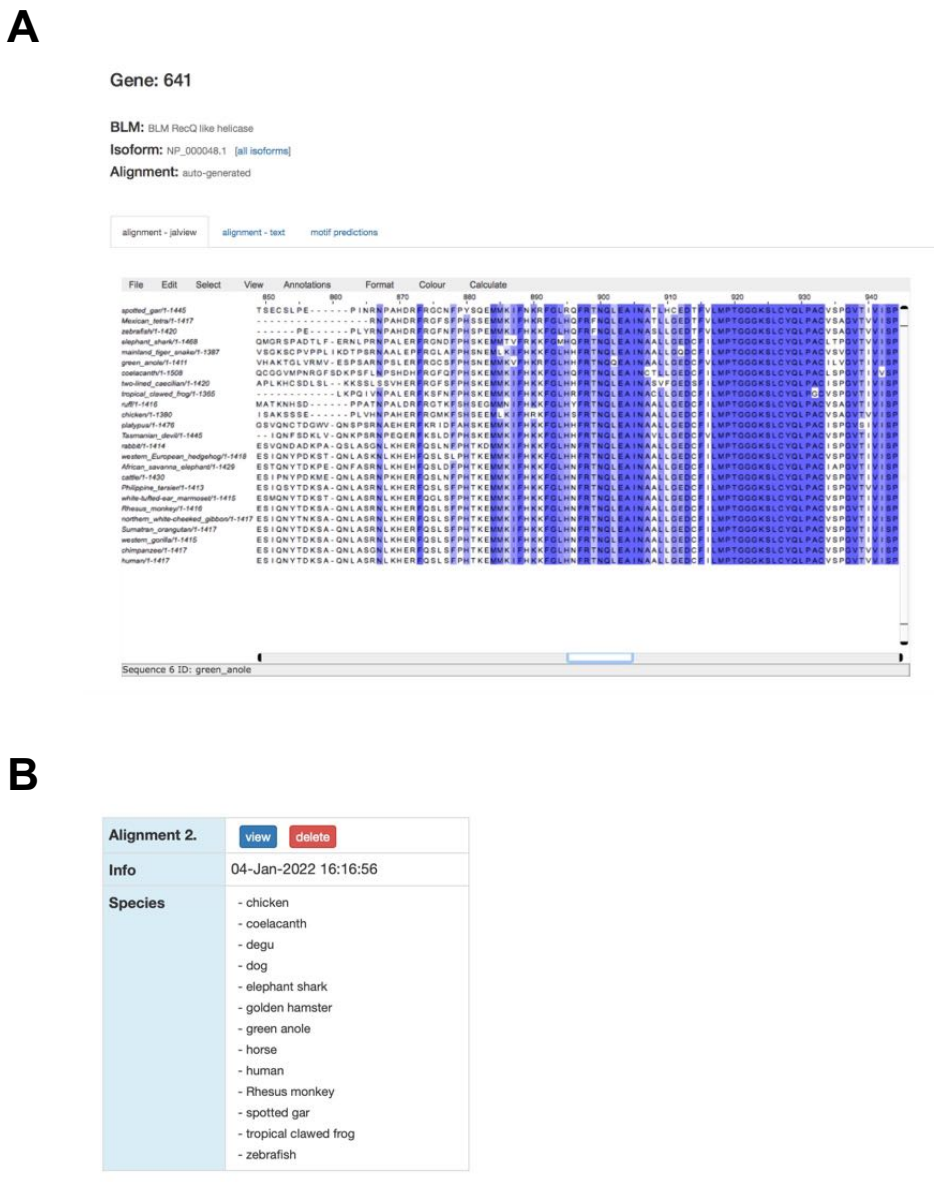


Fig 5.5. Example of a sequence alignment generated by the SLiMCon tool.
A) The automated sequence alignment for BLM, visualized using JalView. B) Alternative sequence alignments can be generated using the SLiMCon tool, using orthologues selected by the user.

5.3.2 Automated Identification of SLiMs

With automated sequence alignments generated, the next step was to generate a tool that could predict SLiMs automatically, such as SLiM 1 in BLM helicase (**Fig 5.6A**). To aid the development of this tool, a way of restricting the prediction of SLiMs to short conserved sequences found within regions of intrinsic disorder was required. To do this, a conservation threshold and the minimum and maximum number of residues had to be defined. To predict whether the SLiM was within a region of predicted disorder, the IUPred score of the region and the length of an unconserved flank length were used.

Using SLiM 1 in BLM as an example (**Fig 5.6A**): the "consensus threshold" would be 90%, the "length" would be a 7 and the "flank length" could be over 30 amino acids, with 0 "conserved residues." It became clear that extra criteria were required: a "number of gaps", a "gap length" and a "maximum total of gap positions". Without these metrics, as demonstrated in Fig 5.3A, if there is a "gap" i.e., a non-conserved residue, the tool would recognise the two halves of the SLiM as two separate SLiMs. The other half of the SLiM would be recognised to be part of the "flank" and both would be eliminated.

To determine what filters should be designated as the default, values for the above parameters were calculated for several SLiMs identified in literature or in Chapter 3 (**Table 5.2**). For each SLiM, the relevant multiple sequence alignment was downloaded from the "SLiMCon" web tool. Then based on these calculated values, SLiM prediction parameters were established (**Fig 5.6B**). Any predicted SLiMs with an average IUPred score below 0.3 or C/N-terminal flanks below 0.4 were eliminated. From calculated IUPred values of published SLiMs, often the IUPred score of the SLiM was lower than the IUPred score of the flanks. This reflects that although SLiMs are found in intrinsic regions of disorder, they may adopt local structure. Whilst these parameters were set as the default, they can be adapted by the user.

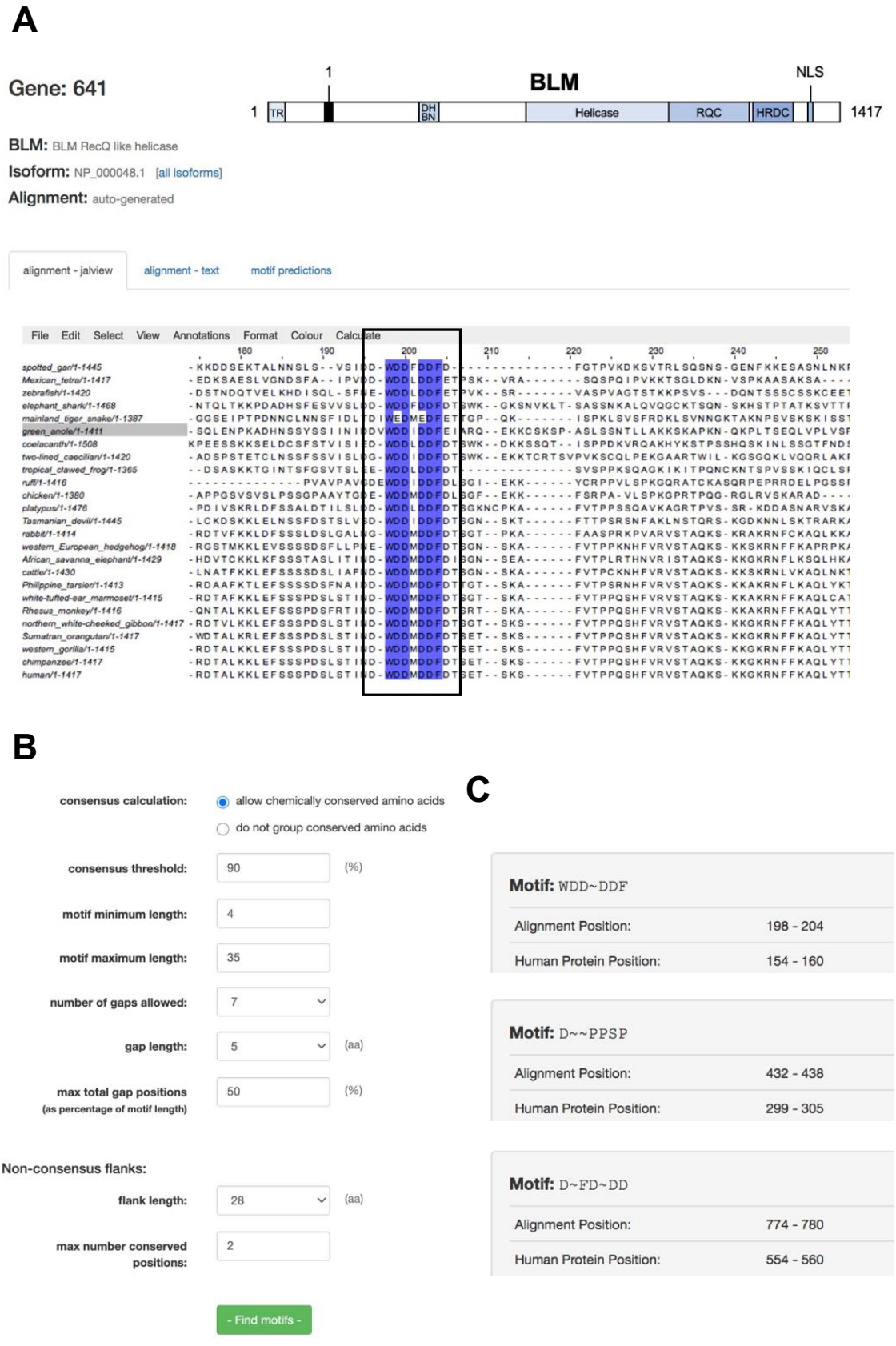


Fig. 5.6. Automated identification of SLiMs.
A) The SLiMCon automated sequence alignment for BLM, with SLiM 1 highlighted. B) Established SLiM prediction parameters based on values calculated for several SLiMs identified in literature or in Chapter 3. C) SLiMs predicted in BLM predicted by the SLiMCon tool. All three SLiMs outlined in Chapter 3 were successfully detected via this tool, with no false positive examples.

Table 5.2. Calculation of values for the development of SLiM prediction parameters

Protein	Residues	Published			No. of Residues	No. of Gaps	Max Gap Length	*No. of Gap Residues*	% Total Gap Positions	Flank Length	No. of Conserved Positions in Flank	IUPred Scores			
		Published	Pubmed hyperlink	Interacting Partner								Motif	N-terminal flank	C-terminal flank	Average of flanks
APLF	182-191	Yes (Shirodakar et al., 2013)	https://pubmed.ncbi.nlm.nih.gov/23689425	Ku70/80	9	1	1	1	10%	30+	1	0.4648	0.51	0.5381	0.52405
AUNIP	340-348	No	-	-	8	0	0	0	0%	30+	3	0.4362	0.4958	1	0.7479
BARD1	140-161	No	-	-	21	3	2	5	23%	25	0	0.3287	0.3247	0.5195	0.4221
BLM	153-161	Yes (Shorrocks et al., 2021)	https://pubmed.ncbi.nlm.nih.gov/33500419/	RPA1	8	0	0	0	0%	30+	0	0.5798	0.5375	0.5223	0.5299
BLM	554-563	Yes (Shorrocks et al., 2021)	https://pubmed.ncbi.nlm.nih.gov/33500419/	RPA1	9	3	2	3	30%	30+	3	0.4416	0.5221	0.4547	0.4884
BLM	299-305	Yes (Blackford et al., 2015)	https://pubmed.ncbi.nlm.nih.gov/25794620/	TOPBP1	6	0	0	0	0%	30+	3	0.5544	0.6082	0.4986	0.5534
BRCA2	1103-1131	No	-	-	28	5	3	8	28%	30+	1	0.3723	0.387	0.3829	0.38495
CtIP	327-331	Yes (Yu & Chen., 2004)	https://pubmed.ncbi.nlm.nih.gov/15485915/	BRCA1	4	0	0	0	0%	30+	2	0.6449	0.7321	0.5181	0.6251
CtIP	488-496	Yes (Schaeper et al., 1998)	https://pubmed.ncbi.nlm.nih.gov/9535825/	CtBP	8	0	0	0	0%	15	1	0.5354	0.5344	0.5561	0.54525
ERCC6L2	380-399	No	-	-	19	0	0	0	0%	30+	1	0.5914	0.6274	0.6363	0.63185
FANCJ	1127-1138	Yes (Gong et al., 2010)	https://pubmed.ncbi.nlm.nih.gov/20159562/	TOPBP1	11	1	1	1	6%	30+	2	0.6712	0.626	0.5449	0.58545
FANCM	1133-1139	No	-	-	6	2	1	2	29%	30+	0	0.3995	0.5435	0.4502	0.49685
FANCM	1224-1236	Yes (Deans et al., 2009)	https://pubmed.ncbi.nlm.nih.gov/20064461/	RMI1/2	12	2	1	2	15%	30+	2	0.251	0.4199	0.4142	0.41705
MDC1	192-199	Yes (Leimbacher et al., 2018)	https://pubmed.ncbi.nlm.nih.gov/30898438/	TOPBP1	7	0	0	0	0%	20	2	0.6873	0.5881	0.7433	0.6657
MDC1	218-222	Yes (various)	https://pubmed.ncbi.nlm.nih.gov/18583988/	NBS1	4	1	2	2	33%	15	1	0.7978	0.7392	0.7096	0.7244
POLQ	1321-1332	No	-	-	11	3	2	5	42%	30+	1	0.2853	0.3713	0.4277	0.3995

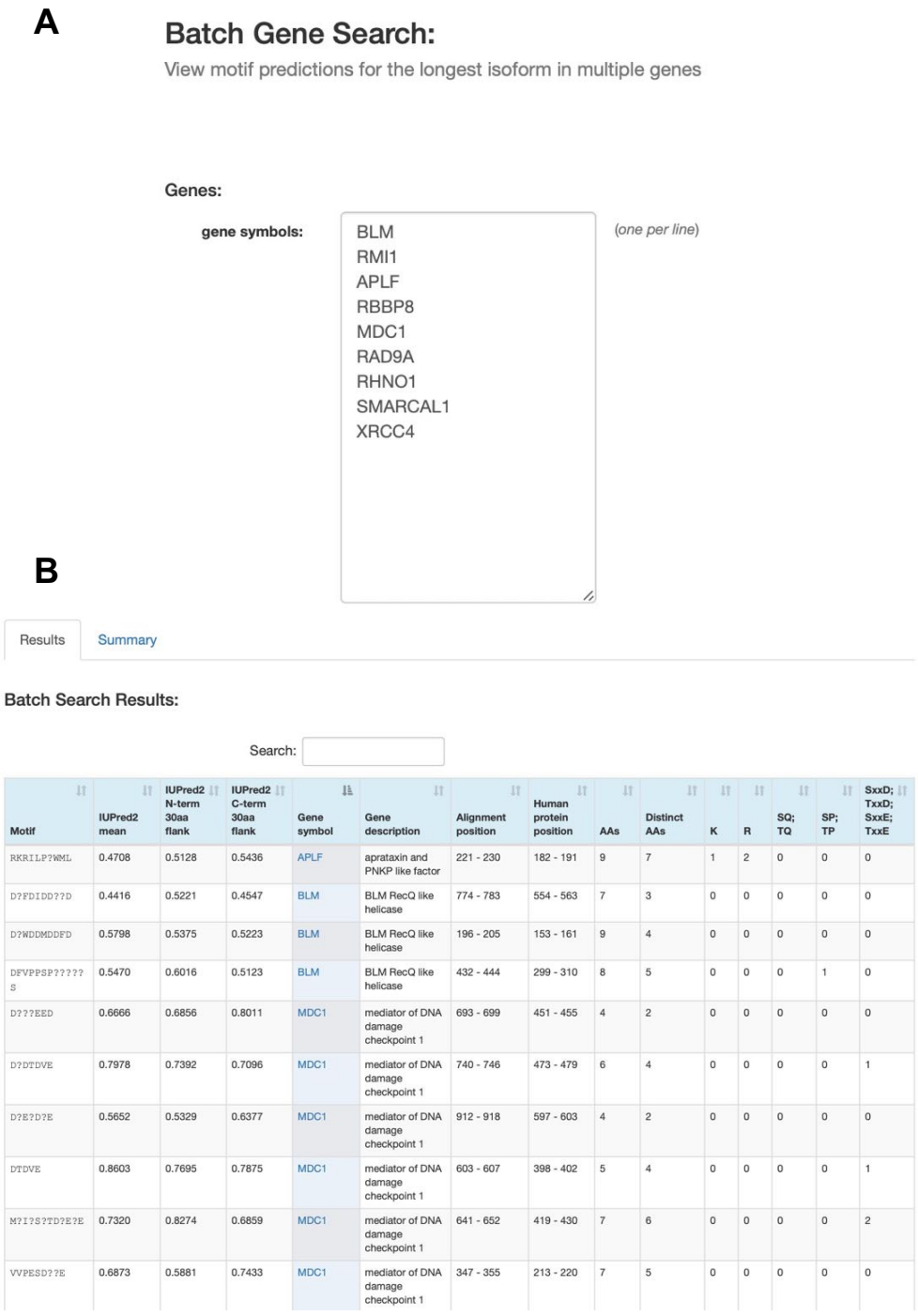
Protein	Residues	Published			No. of Residues	No. of Gaps	Max Gap Length	*No. of Gap Residues*	% Total Gap Positions	Flank Length	No. of Conserved Positions in Flank	IUPred Scores			
		Published	Pubmed hyperlink	Interacting Partner								Motif	N-terminal flank	C-terminal flank	Average of flanks
POLQ	1480-1494	No	-	-	14	3	3	6	40%	30	1	0.3782	0.4174	0.477	0.4472
POLQ	1636-1647	No	-	-	11	3	2	5	42%	30+	0	0.4185	0.5816	0.5218	0.5517
Rhino	198-229	Yes (Day et al., 2018)	https://pubmed.ncbi.nlm.nih.gov/30295604/	TOPBP1	33	4	2	7	24%	30+	0	0.3912	0.5266	1	0.7633
RMI1	241-253	Yes (Shorrocks et al., 2021)	https://pubmed.ncbi.nlm.nih.gov/33500419/	RPA1	12	2	1	2	15%	24	1	0.6139	0.4785	0.6411	0.5598
RMI1	312-324	No	-	-	12	4	1	4	31%	25	1	0.555	0.5992	0.6246	0.6119
RNF168	559-565	BioRiv	https://www.biorxiv.org/content/10.1101/2021.03.17.435897v1	PCNA	6	2	2	3	43%	30+	0	0.6453	0.7624	1	0.8812
SMARCAL1	1-27	Yes (Ciccia et al., 2009)	https://pubmed.ncbi.nlm.nih.gov/19793862/	RPA2	27	1	1	1	4%	30+	2	0.7615	1	0.7791	0.88955
XRCC4	229-236	Yes (Koch et al., 2004; Clements et al., 2004; Cherry et al., 2015)	https://pubmed.ncbi.nlm.nih.gov/15385968/; https://pubmed.ncbi.nlm.nih.gov/15380105/; https://pubmed.ncbi.nlm.nih.gov/26519825/	PNKP; APTX; APLF	7	1	2	2	25%	25	1	0.717	0.6248	0.6055	0.61515

To assess the accuracy of this tool, it was applied to BLM and 19 other proteins. It was able to accurately identify all expected SLiMs in these 20 proteins, confirming its utility. For example, the three SLiMs outlined in chapter 3 in BLM were successfully detected via this tool, with no false positive examples (**Fig 5.6C**). Of note, four different iterations of parameters were trialled, with the parameters outlined being the most accurate in the prediction of putative SLiMs.

5.3.3 Batch search for SLiMs in DDR proteins

SLiMCon allows the user to access the multiple sequence alignment and the identified SLiMs using the gene search function (**Fig 5.4A**). To allow the prediction of SLiMs in a larger range of proteins of interest, a batch gene search function was added (**Fig. 5.7A**). This allows the user to supply the gene names of all proteins of interest as an input. Then, using the default parameters defined above, the batch search function provides an output detailing all the SLiMs identified in each protein of interest. The default parameters can be modified by the user.

For each gene listed all the predicted SLiMs are listed as an output table (**Fig 5.7B**). The output table also contains the IUPred scores of the SLiM and the N/C-terminal flanks, the amino acid position of the SLiM in the SLiMCon sequence alignment and the human protein and whether there is a consensus sequence for PIKKs, CDKs or CK2 within the predicted SLiM. An added feature splits the gene names supplied into those containing SLiMs, those not containing SLiMs and those "not found" by the tool, highlighting that the user must supply an alternate gene name.



Motif	IUPred2 mean	IUPred2 N-term 30aa flank	IUPred2 C-term 30aa flank	Gene symbol	Gene description	Alignment position	Human protein position	AAs	Distinct AAs	K	R	SQ; TQ	SP; TP	SxxD; TxxD; SxxE; TxxE
RKRILP?WML	0.4708	0.5128	0.5436	APLF	aprataxin and PNKP like factor	221 - 230	182 - 191	9	7	1	2	0	0	0
D?FDIDD??D	0.4416	0.5221	0.4547	BLM	BLM RecQ like helicase	774 - 783	554 - 563	7	3	0	0	0	0	0
D?WDDMDDFD	0.5798	0.5375	0.5223	BLM	BLM RecQ like helicase	196 - 205	153 - 161	9	4	0	0	0	0	0
DFVPPSP?????S	0.5470	0.6016	0.5123	BLM	BLM RecQ like helicase	432 - 444	299 - 310	8	5	0	0	0	1	0
D???EED	0.6666	0.6856	0.8011	MDC1	mediator of DNA damage checkpoint 1	693 - 699	451 - 455	4	2	0	0	0	0	0
D?DTDVE	0.7978	0.7392	0.7096	MDC1	mediator of DNA damage checkpoint 1	740 - 746	473 - 479	6	4	0	0	0	0	1
D?E?D?E	0.5652	0.5329	0.6377	MDC1	mediator of DNA damage checkpoint 1	912 - 918	597 - 603	4	2	0	0	0	0	0
DTDVE	0.8603	0.7695	0.7875	MDC1	mediator of DNA damage checkpoint 1	603 - 607	398 - 402	5	4	0	0	0	0	1
M?I?S?TD?E?E	0.7320	0.8274	0.6859	MDC1	mediator of DNA damage checkpoint 1	641 - 652	419 - 430	7	6	0	0	0	0	2
VVPESD??E	0.6873	0.5881	0.7433	MDC1	mediator of DNA damage checkpoint 1	347 - 355	213 - 220	7	5	0	0	0	0	0

Fig. 5.7. Batch search for SLiMs in DDR proteins.
A) Batch search gene function in the SLiMCon tool. This allows the user to supply the gene names of all proteins of interest as an input to identify all SLiMs predicted by the SLiMCon tool. B) Output table showing SLiMs predicted using the batch search gene function. The output table also contains: the average IUPred score of the SLiM as well as the N/C-terminal 30 amino acid flank; the position of the SLiM in the SLiMCon sequence alignment, as well as in the reference human protein, allowing the user to verify the predicted SLiM; the number of amino acids, as well as the number of distinct amino acids; and lastly whether there is a consensus sequence for PIKKs, CDKs or CK2 within the predicted SLiM.

To assess whether the tool is accurate in predicting SLiMs across a larger range of proteins, it was used to identify SLiMs in 120 DDR proteins (**Table 8.7.1**). As in section 3.3.1, these proteins were selected from a curated list of proteins associated with DNA repair, assembled by experts (Pearl et al., 2015). 222 SLiMs were identified in 111 proteins, 9 proteins were not predicted to contain any SLiMs. No supplied gene names were "not found."

Each predicted SLiM was inspected to validate the accuracy of the tool (**Table 5.3**). A manual sequence alignment was generated for each query protein, to assess whether the prediction was correct in an independently derived sequence alignment. SLiMs highlighted in green in the output table below (**Table 5.3**) were determined to be appropriate for synthesis and for use in peptide pulldowns for mass spectrometry. For example, a SLiM in WRN helicase predicted by SLiMCon is shown in **Fig. 5.8A**. Using the independently derived sequence alignment in **Fig. 5.8B**, the prediction was assessed to be correct. SLiMs that have already been synthesised (in Chapter 3) were highlighted in blue. Predicted SLiMs that were considered to be incorrectly assigned as a SLiM ("false positive") were highlighted in red. For example, a SLiM in TOP3A predicted by SLiMCon is shown in **Fig. 5.9A**. Using the independently derived sequence alignment shown in **Fig. 5.9B**, the prediction was assessed to be incorrect. It was incorrectly assigned as a SLiM because the SLiMCon tool cannot distinguish between insertions in the sequence alignments, such as the insertion in the armadillo sequence (highlighted in red in **Fig. 5.9A**), and unconserved residues. An insertion in the alignment was recognised by the tool to be a unconserved flank and as such the region was predicted to be a SLiM. Most of the false positive examples were eliminated automatically by the IUPred threshold. Of the incorrectly assigned SLiMs in **Table 5.3**, all were incorrectly assigned because of an issue with the automated sequence alignment, rather than the default parameters used by the SLiMCon tool. Fourteen out of the fifteen were wrongly assigned because of an insertion in the sequence alignment, as with TOP3A (**Fig. 5.9**).

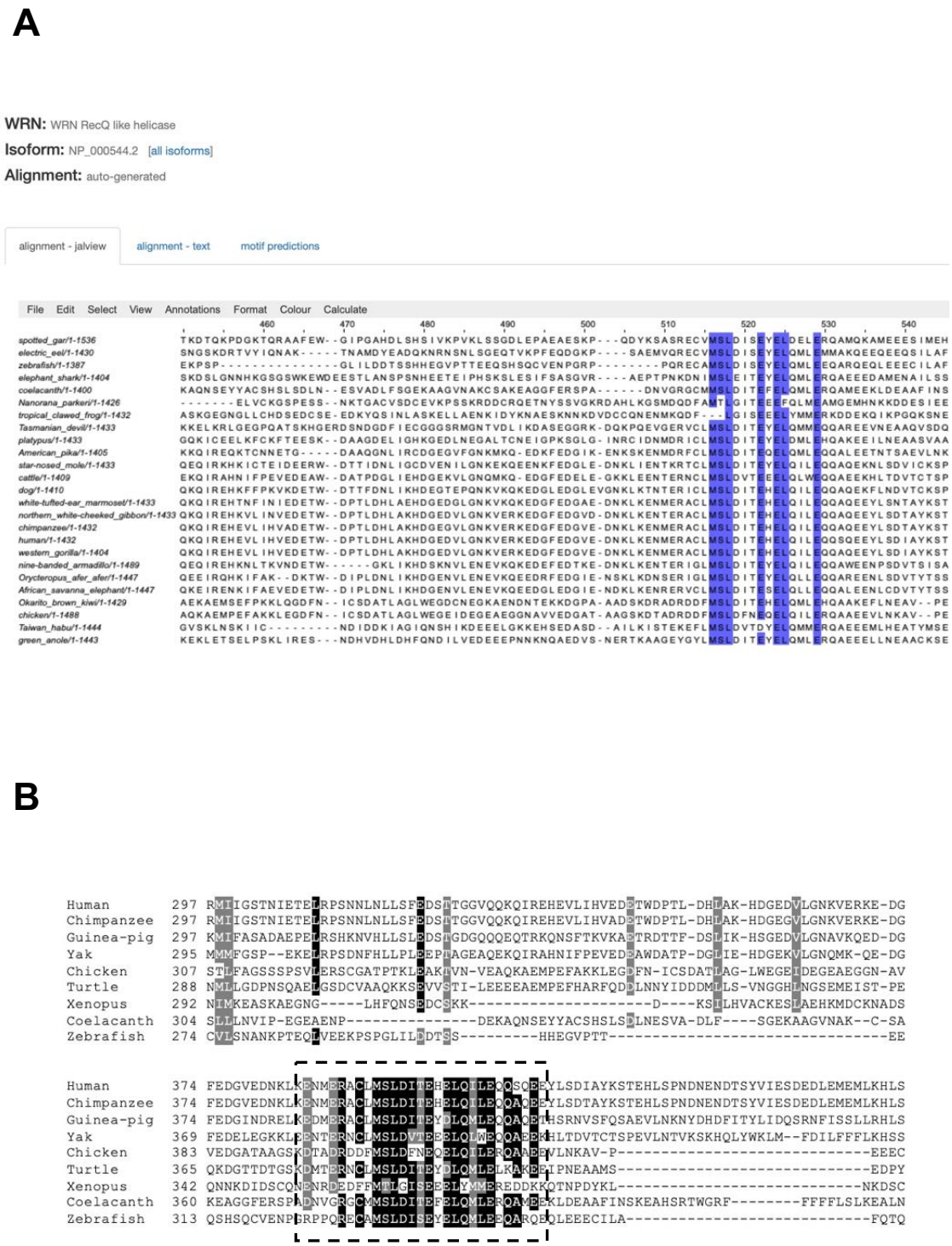


Fig. 5.8. Assessment of the accuracy of the SLiMCon tool: a SLiM predicted in WRN.
A) The SLiMCon automated sequence alignment for WRN, with a predicted SLiM highlighted. B) An independently derived sequence alignment for WRN, demonstrating that the SLiMCon prediction was correct. In both alignments, the putative SLiM is an “island of conservation.

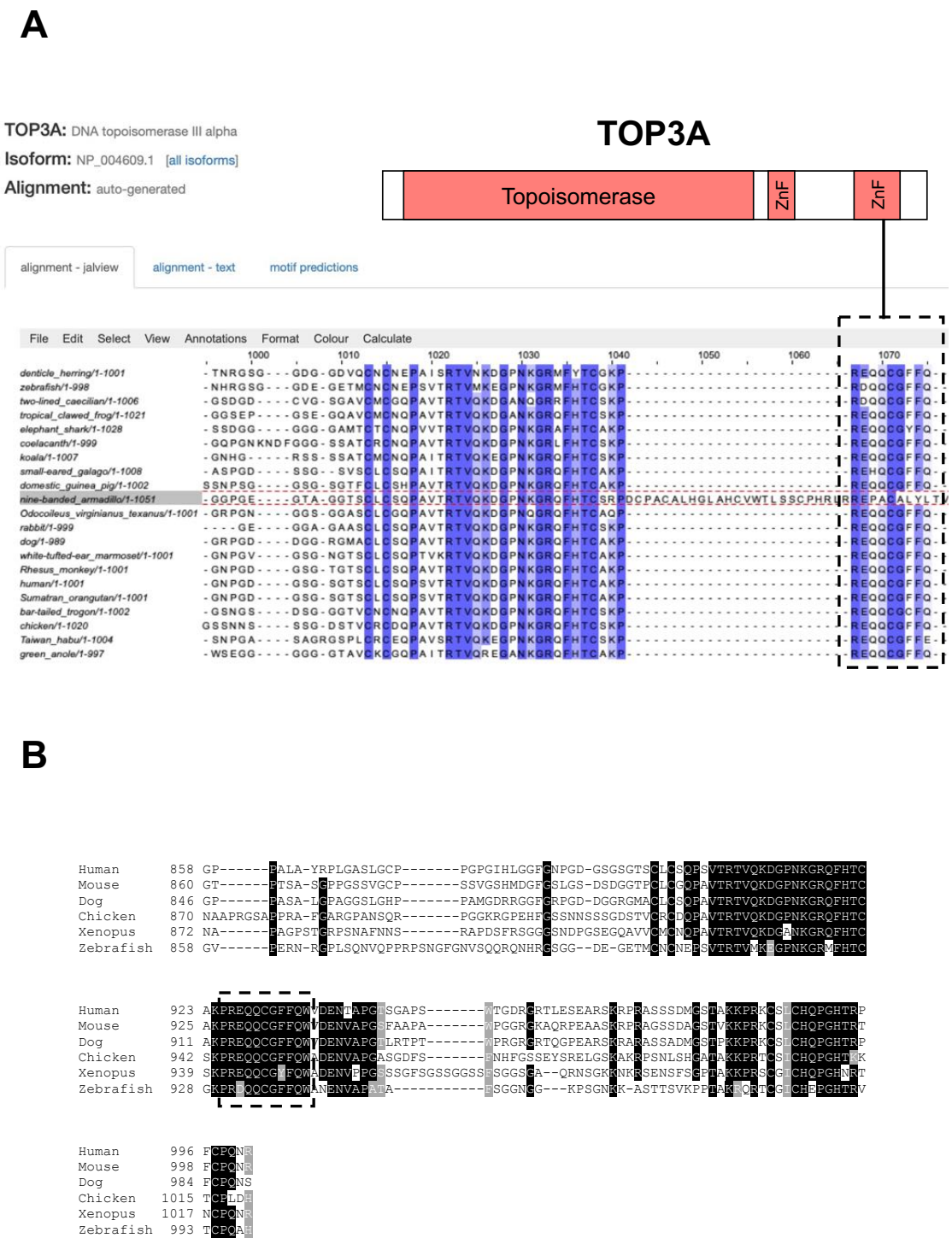


Fig. 5.9. Assessment of the accuracy of the SLiMCon tool: a SLiM predicted in TOP3A.
A) The SLiMCon automated sequence alignment for TOP3A, with a predicted SLiM highlighted. An insertion in the armadillo sequence is highlighted in red. B) An independently derived sequence alignment for TOP3A, demonstrating that the SLiMCon prediction was incorrect. The predicted SLiM (highlighted) is part of a larger conserved region.

It is therefore recommended that predicted SLiMs should be visually inspected in the automated sequence alignment before experimental work is pursued. In addition, for each SLiM predicted by the SLiMCon tool, a literature search was conducted to see if it had already been identified. If so, the Pubmed hyperlink and the proposed interacting partner was included in Table 5.3. Published SLiMs are highlighted in purple in the output below. In summary, of the 222 SLiMs predicted: 64% were categorised to be synthesised in a future mass spectrometry screen outside of this thesis (highlighted green), 6% as proposed false positives (highlighted red), 22% are already published (highlighted purple) and 8% have already been synthesised (highlighted blue) as part of this thesis (**Fig 5.10**). The ability of the tool to identify numerous published examples, as well as having a low false positive discovery rate (6%) highlights the accuracy of the SLiMCon tool, thus confirming its utility.

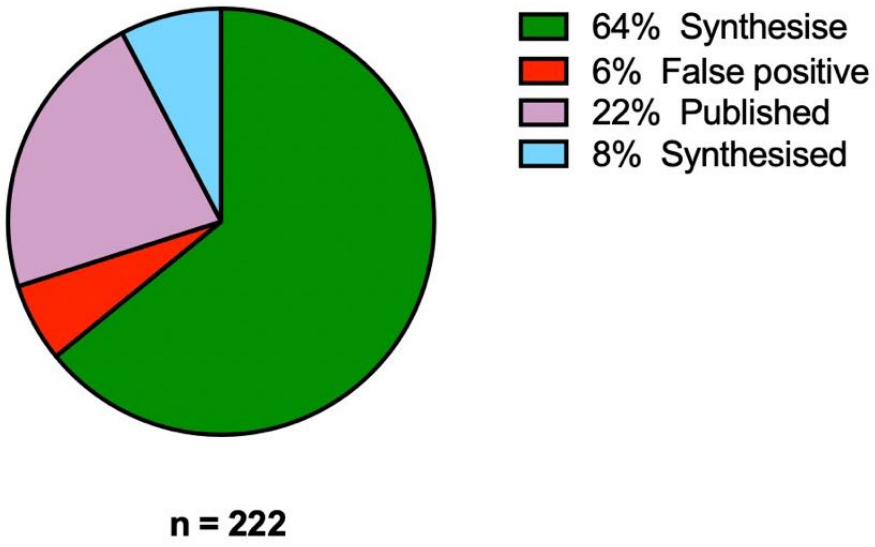


Fig. 5.10. 64% of SLiMs predicted by the SLiMCon tool are previously uncharacterised putative SLiMs.
Of the 222 SLiMs predicted: 64% were categorized to be synthesised in a future mass spectrometry screen (green), 6% as proposed false positives (red), 22% are already published (purple) and 8% have already been synthesised (blue) as part of this thesis

Table 5.3. Predicted SLiMs in 120 DDR proteins using the SLiMCon tool

Motif	Gene	Residues	Published			Synthesised	Alignment problem	Alphafold	Conclusion
				Pubmed hyperlink	Interacting Partner				
RKRILP?WML	APLF	182 - 191	Yes (Shirodakar et al., 2013)	https://pubmed.ncbi.nlm.nih.gov/23689425	Ku70/80	No	No	Local structure in unstructured region	Eliminate
F?TKE?F	ATRIP	247 - 253	Yes (Mordes et al., 2008)	https://pubmed.ncbi.nlm.nih.gov/18519640/	TOPBP1	No	No	Local structure in unstructured region	Eliminate
LT?EELRKRR??YFEK	ATXN3	276 - 291	Yes (Antony et al., 2009)	https://pubmed.ncbi.nlm.nih.gov/19660550/	Proposed NLS	No	No	Local structure in unstructured region	Eliminate
LFT?DS?G??VI?H	AUNIP	287 - 300	No	-	-	No	No	Local structure in unstructured region	Synthesise
LLFT?DS	AUNIP	341 - 347	No	-	-	Yes	No	Local structure in unstructured region	Eliminate
K??IKMWFSPRSKKVR??V?K	BARD1	173 - 193	No	-	-	Yes	No	Local structure in unstructured region	Eliminate
K?VS	BARD1	271 - 274	No	-	-	No	No	Unstructured region	Synthesise
K?L??IN??W	BARD1	242 - 251	No	-	-	No	No	Local structure in unstructured region	Synthesise
D?FDIDD??D	BLM	554 - 563	Yes (Shorrocks et al., 2021)	https://pubmed.ncbi.nlm.nih.gov/33500419/	RPA1	Yes	No	Local structure in unstructured region	Eliminate
D?WDDMDDFD	BLM	153 - 161	Yes (Shorrocks et al., 2021)	https://pubmed.ncbi.nlm.nih.gov/33500419/	RPA1	Yes	No	Local structure in unstructured region	Eliminate
DFVPPSP?????S	BLM	299 - 310	Yes (Blackford et al., 2015)	https://pubmed.ncbi.nlm.nih.gov/25794620/	TOPBP1	Yes	No	Local structure in unstructured region	Eliminate
ED?IFGKTY?R?K???P	BRCA1	457 - 472	No	-	-	Yes	No	Unstructured region	Eliminate
L??SI?KV?EW??R	BRCA1	375 - 388	No	-	-	No	No	Local structure in unstructured region	Synthesise
KRR????L?PEDFIKK	BRCA1	505 - 520	No	-	-	Yes	No	Unstructured region	Eliminate
E?EDSE?D???L???F???KR?SF?L	BRCA1	846 - 871	No	-	-	No	No	Local structure in unstructured region	Synthesise

S?TPD?LL	BRCA1	1147 - 1154	No	-	-	No	No	Local structure in unstructured region	Synthesise
EA?E???A?MED	BRCA2	2226 - 2237	No	-	-	No	No	-	Synthesize
E???V?DEELALINTQ?L	BRCA2	3335 - 3352	No	-	-	Yes	No	-	Eliminate
SP??L	BRCA2	142 - 146	No	-	-	No	Yes (insertion)	-	Synthesize
K??K?FVPPFK	BRCA2	2401 - 2411	Yes (Thorslund et al., 2007)	https://pubmed.ncbi.nlm.nih.gov/17541404/	DMC1	Yes	No	-	Eliminate
D?SIYFTPELY	BRIP1	1147 - 1157	Yes (Gong et al., 2010)	https://pubmed.ncbi.nlm.nih.gov/20159562/	TOPBP1	No	No	Unstructured region	Eliminate
DEDDGFLDLLD???L?ND?E?PS?MASLWTAPLV	CDC25A	213 - 246	No	-	-	No	Yes (insertion to the C-terminus of the SLiM, but borderline)	Local structure in unstructured region	Synthesize
L?R??SSES?D?G??LDSP	CDC25A	71 - 89	No	-	-	No	Yes (Insertion to the C-terminus of the SLiM; structured region)	Unstructured region	Eliminate
PVT?L???M??L??LG???ETP	CDC25B	51 - 70	Yes (Uchida et al., 2004)	https://pubmed.ncbi.nlm.nih.gov/15003534/	Proposed NES	No	Yes	Local structure in unstructured region	Eliminate
CSTP	CDC25C	206 - 209	No	-	-	No	No	Local structure in unstructured region	Synthesize
LYR?PSMP??L?????K	CDC25C	289 - 305	No	-	-	No	No	Local structure in unstructured region	Synthesize
S?KR	CHEK2	571 - 574	Yes (Zannini et al., 2003)	https://www.jbc.org/article/S0021-9258(20)82775-6/fulltext	Proposed NLS	No	No	Unstructured region	Eliminate
EELL?LCSG?F?S	CLSPN	938 - 950	Yes (Kumagai & Dunphy., 2003)	https://www.nature.com/articles/ncb921	CHK1	No	No	Local structure in unstructured region	Eliminate
SD?T?LLF	CLSPN	721 - 728	No	-	-	No	No	Unstructured region	Eliminate
K??RI?????DSDES	CLSPN	99 - 113	No	-	-	No	No	Unstructured region	Synthesise
M?E?LALCSG?F???P	CLSPN	974 - 986	Yes (Kumagai & Dunphy., 2003)	https://www.nature.com/articles/ncb921	CHK1	No	Yes (insertion splits SLiM in two)	Local structure in unstructured region	Eliminate
EE?E???F?LV	CLSPN	994 - 1004	Yes (Kumagai & Dunphy., 2003)	https://www.nature.com/articles/ncb921	CHK1	No	Yes (insertion splits SLiM in two)	Unstructured region	Eliminate
E??E?LL	CLSPN	676 - 682	No	-	-	No	Yes (N-terminal flank unconserved in invertebrates,	Unstructured region	Eliminate

							generating an insertion in the alignment)		
T?YCMNEAELVD?AL?IL?E	CYREN	52 - 71	No	-	-	No	No	Local structure in unstructured region	Synthesize
D??ED?LKYVREIFFS	CYREN	142 - 157	Yes (Hung et al., 2018)	https://pubmed.ncbi.nlm.nih.gov/30017584/	ATM	No	No	Local structure in unstructured region	Eliminate
K?R?LP?WM	CYREN	9 - 17	Yes (Grundy et al., 2016)	https://pubmed.ncbi.nlm.nih.gov/27063109/	Ku70/80	No	No	Local structure in unstructured region	Eliminate
S???ISYSPL?S	DCLRE1A	294 - 305	No	-	-	No	No	Local structure in unstructured region	Synthesize
K?MKQ?DIGVYFGL?P	DCLRE1A	552 - 567	Yes (Yang et al., 2010)	https://pubmed.ncbi.nlm.nih.gov/20385554/	PCNA	No	No	Local structure in unstructured region	Eliminate
KRKA??SL	DCLRE1A	600 - 607	No	-	-	No	No	Unstructured region	Synthesize
EE?IWEYK??RK	DCLRE1A	7 - 18	No	-	-	No	No	Local structure in unstructured region	Synthesize
FVDC?ESN?E?E	DCLRE1C	449 - 460	Yes (Niewolik et al., 2017)	https://pubmed.ncbi.nlm.nih.gov/28082683/	ARTEMIS	No	No	Unstructured region	Eliminate
V??L?S?SEDE????IPL??RL??K	EME1	80 - 103	No	-	-	No	No	Unstructured region	Synthesize
D?E?ELP?F??FL	EME1	16 - 26	No	-	-	No	No	Unstructured region	Synthesize
H?P??K	ERCC1	312 - 317	No	-	-	No	Yes (Insertion to N-terminus of the motif disrupting the sequence alignment)	Structured region	Eliminate
TLTQMV?K	ERCC4	539 - 546	No	-	-	No	No	Local structure in unstructured region	Synthesize
FNL?LE?S	ERCC6L	953 - 960	No	-	-	No	No	Unstructured region	Synthesize
LLGDTSIL??LF	ERCC6L2	1438 - 1449	No	-	-	No	No	Local structure in unstructured region	Synthesize
LTL???GFS	ERCC6L2	788 - 796	No	-	-	Yes	No	Local structure in unstructured region	Eliminate
E??D?I	ERCC6L2	899 - 904	No	-	-	No	No	Unstructured region	Synthesize
S??D	ERCC6L2	876 - 879	No	-	-	No	No	Unstructured region	Synthesize
SDESDDIEI	ERCC6L2	980 - 988	No	-	-	Yes	No	Unstructured region	Eliminate
I??FSSS?D	ERCC6L2	1038 - 1046	No	-	-	No	No	Unstructured region	Synthesize

KF?F?K???S	ETAA1	820 - 829	No	-	-	No	No	Unstructured region	Synthesize
K?S?EEI?RKRQEAL?RR??K	ETAA1	892 - 912	Yes		RPA2	No	No	Local structure in unstructured region	Eliminate
SIW?R?Y	EXO1	369 - 375	No	-	-	No	No	Local structure in unstructured region	Synthesize
LQ?F?R	EXO1	630 - 635	No	-	-	No	No	Local structure in unstructured region	Synthesize
DLSD?E?L	EXO5	19 - 26	No	-	-	No	No	Local structure in unstructured region	Synthesize
D?V??FL	FAAP100	674 - 680	No	-	-	No	No	Local structure in unstructured region	Synthesize
W???LLR	FAAP20	181 - 186	Yes (Leung et al., 2012)	https://pubmed.ncbi.nlm.nih.gov/22396592/	FANCA	No	No	Local structure in unstructured region	Eliminate
L??CPMC???F	FAAP20	285 - 295	No	-	-	No	No	Local structure in unstructured region	Synthesize
KR????LS	FAN1	11 - 18	No	-	-	No	Yes (Insertion to C-terminus of the motif)	Structured region	Eliminate
M??KRR	FANCD2	1 - 6	Yes (Biosvert et al., 2013)	https://pubmed.ncbi.nlm.nih.gov/24278431/	Proposed NLS	No	No	Unstructured region	Eliminate
P?KKKRK	FANCI	1321 - 1327	Yes (Colnaghi et al., 2011)	https://pubmed.ncbi.nlm.nih.gov/20971953/	Proposed NLS	No	No	Unstructured region	Eliminate
SL?LF?D	FANCM	1133 - 1139	No	-	-	Yes	No	Local structure in unstructured region	Eliminate
Q?TL???W	FANCM	5 - 12	Yes (Rohleder et al., 2016)	https://pubmed.ncbi.nlm.nih.gov/26825464/	PCNA	No	No	Local structure in unstructured region	Eliminate
R?HL???DC??L	FBH1	57 - 68	No	-	-	No	No	Local structure in unstructured region	Synthesize
I??L?LS?IDW??TSFS?SP	GEN1	580 - 599	No	-	-	No	No	Local structure in unstructured region	Synthesize
DSPLPL??RLK?R	GEN1	892 - 904	No	-	-	No	No	Local structure in unstructured region	Synthesize
SVC????SS?E	GEN1	794 - 804	No	-	-	No	No	Unstructured region	Synthesize
L?G?L?P?K	HELB	11 - 19	No	-	-	No	No	Unstructured region	Synthesize
P?SQ	HELQ	196 - 199	No	-	-	No	No	Local structure in unstructured region	Synthesize
Q??Y?Q??YQQY????QQW?QY	HNRNPUL1	820 - 836	No	-	-	No	Yes	Local structure in unstructured region	Synthesize

MK??E?K	HNRNPUL1	141 - 146	No	-	-	No	No	Unstructured region	Synthesize
MDVRRLKVNELREEL?RRGLDTRGLKAEL?ERL	HNRNPUL1	1 - 33	Yes (Aravind & Koonin., 2000)	https://pubmed.ncbi.nlm.nih.gov/10694879/	SAP motif; DNA binding	No	No	Local structure in unstructured region	Eliminate
L??LL??L	HROB	630 - 637	No	-	-	No	No	Local structure in unstructured region	Synthesize
G??W	MCM8	29 - 31	No	-	-	No	No	Unstructured region	Synthesize
F?F??K??KL	MCM9	923 - 931	No	-	-	No	No	Unstructured region	Synthesize
YEDYFS???NL	MCPH1	412 - 421	No	-	-	No	No	Local structure in unstructured region	Synthesize
M??RL??M??KREN?S?T	MCPH1	176 - 193	No	-	-	No	No	Local structure in unstructured region	Synthesize
D?E?D?E	MDC1	597 - 603	No	-	-	No	No	Unstructured region	Synthesize
VVPESD??E	MDC1	213 - 220	Yes (Leimbacher et al., 2018)	https://pubmed.ncbi.nlm.nih.gov/30898438/	TOPBP1	Yes	No	Unstructured region	Eliminate
D???EED	MDC1	451 - 455	No	-	-	No	No	Unstructured region	Synthesize
D?DTDVE	MDC1	473 - 479	Yes (various)	https://pubmed.ncbi.nlm.nih.gov/18583988/	NBS1	No	No	Unstructured region	Eliminate
DTDVE	MDC1	398 - 402	Yes (various)	https://pubmed.ncbi.nlm.nih.gov/18583988/	NBS1	No	No	Unstructured region	Eliminate
M?I?S?TD?E?E	MDC1	419 - 430	Yes (various)	https://pubmed.ncbi.nlm.nih.gov/18583988/	NBS1	No	No	Unstructured region	Eliminate
EEG?DVPD????KK	MDM2	360 - 370	No	-	-	No	No	Unstructured region	Synthesize
RW?KYLD	MRNIP	106 - 112	No	-	-	No	No	Local structure in unstructured region	Synthesize
SKW??FL	MRNIP	227 - 233	No	-	-	No	No	Local structure in unstructured region	Synthesize
LF?T?EDF??DL	MRNIP	332 - 343	No	-	-	No	No	Local structure in unstructured region	Synthesize
SLL?R?LVLRT???P?RYSLT??GL?LA??L	MUS81	187 - 216	No	-	-	No	No	Local structure in unstructured region	Synthesize
KK??L	NBN	591 - 595	No	-	-	No	No	Local structure in unstructured region	Synthesize
TMEVSC	NDRG3	383 - 388	No	-	-	No	No	Unstructured region	Synthesize

K?K?KK??GLF	NHEJ1	288 - 298	Yes (Grundy et al., 2016)	https://pubmed.ncbi.nlm.nih.gov/27063109/	Ku70/80	No	No	Local structure in unstructured region	Eliminate
RQRKY	PAGR1	250 - 254	No	-	-	No	No	Unstructured region	Synthesize
I?DF?LPDEDFG?LK??KL	PALB2	611 - 629	No	-	-	No	No	Local structure in unstructured region	Synthesize
P?LG?TP	PALB2	731 - 737	No	-	-	No	No	Local structure in unstructured region	Synthesize
EE??V?PESPLSD	PARG	253 - 265	No	-	-	No	No	Local structure in unstructured region	Synthesize
KR????GESLINPG?K	PAXX	177 - 191	Yes (Blackford et al., 2015)	https://pubmed.ncbi.nlm.nih.gov/25574025/	Ku70/80	No	No	Local structure in unstructured region	Eliminate
L?LLM?NG	PHF8	605 - 612	Yes (Loenarz et al., 2010)	https://pubmed.ncbi.nlm.nih.gov/19843542/	Proposed NLS	No	No	Unstructured region	Eliminate
L???LM	PHF8	556 - 561	Yes (Loenarz et al., 2010)	https://pubmed.ncbi.nlm.nih.gov/19843542/	Proposed NLS	No	No	Unstructured region	Eliminate
Q??GKRPKKGLATAKQRLGRILKI HRNGKLLL	PHF8	1031 - 1060	No	-	-	No	Yes (Insertion to N-terminus of the motif)	Unstructured region	Eliminate
MEFED?DFDEPME	POLA1	253 - 265	No	-	-	No	No	Unstructured region	Synthesize
RK?LITDSY?VVKR	POLD4	3 - 16	No	-	-	No	No	Local structure in unstructured region	Synthesize
SV?DSQL???L??Y?T	POLQ	1436 - 1451	No	-	-	Yes	No	Unstructured region	Eliminate
E??LN?S???LFD?SF	POLQ	1480 - 1494	No	-	-	Yes	No	Unstructured region	Eliminate
S???SE	POLQ	1555 - 1560	No	-	-	No	No	Unstructured region	Synthesize
F?LSP?L??ILD	POLQ	1636 - 1647	No	-	-	Yes	No	Local structure in unstructured region	Eliminate
P?PT	POLQ	1736 - 1738	No	-	-	No	No	Unstructured region	Synthesize
M???YSL?VSV?SDQGGRKYMED VT?IVVEPE	PPM1D	1 - 32	No	-	-	No	No	Local structure in unstructured region	Synthesize
E?LE?YF?E	PRIM2	563 - 571	No	-	-	No	No	Local structure in unstructured region	Synthesize
V??WV?PSF	RAD17	15 - 23	No	-	-	No	No	Local structure in unstructured region	Synthesize
M?LDDKLY?R?LE?AL?LS	RAD51AP 1	88 - 106	No	-	-	No	No	Local structure in unstructured region	Synthesize
SNCSV????L?L??IT	RAD51AP 1	143 - 158	Yes (Liang et al., 2016)	https://pubmed.ncbi.nlm.nih.gov/27239033/	UAF1	No	Yes	Local structure in unstructured region	Eliminate

KSP???LRLGLSR?ARVKPLHP	RAD51AP1	326 - 347	Yes (Modesti et al., 2007)	https://pubmed.ncbi.nlm.nih.gov/17996710/	RAD51	No	No	Unstructured region	Eliminate
SP????K?P?W?PP	RAD51AP1	294 - 307	Yes (Dunlop et al., 2011)	https://pubmed.ncbi.nlm.nih.gov/21903585/	DMC1	No	No	Local structure in unstructured region	Eliminate
D???Q???RKLRQ?QLQ??FR??M	RAD52	252 - 272	No	-	-	No	No	Local structure in unstructured region	Synthesize
R??Q???KKRK?D	RAD52	405 - 416	Yes (Koike, Yutoku & Koike., 2013)	https://pubmed.ncbi.nlm.nih.gov/23639616/	Proposed NLS	No	No	Unstructured region	Eliminate
MRRS?APSQL?G???KK?RFI?P	RAD54B	1 - 23	No	-	-	No	No	Local structure in unstructured region	Synthesize
LDL???KRK??T	RAD54L2	1249 - 1260	Yes (Rouleau et al., 2002)	https://pubmed.ncbi.nlm.nih.gov/12058073/	Proposed NLS	No	No	Unstructured region	Eliminate
DDF??DDID?YMIAMETT	RAD9A	307 - 324	Yes (Xu et al., 2008)	https://pubmed.ncbi.nlm.nih.gov/18936170/	RPA1	No	No	Unstructured region	Eliminate
EV???IE?L?YRE?KRIAIQG?I??I?Y	RADX	534 - 561	No	-	-	No	Yes (insertion to the N-terminus of SLiM)	Structured region	Eliminate
E??IPDSPI???S?S????LRRK?E??HVRY?E	RBBP8	157 - 188	No	-	-	No	No	Local structure in unstructured region	Synthesize
EY?S	RBBP8	735 - 738	No	-	-	No	No	Unstructured region	Synthesize
L??VVAETLG????E??E	RBBP8	253 - 270	No	-	-	No	No	Local structure in unstructured region	Synthesize
SPVFG	RBBP8	327 - 331	Yes (Yu & Chen., 2004)	https://pubmed.ncbi.nlm.nih.gov/15485915/	BRCA1	No	No	Local structure in unstructured region	Eliminate
EILVADT	RBBP8	221 - 227	No	-	-	No	No	Unstructured region	Synthesize
K??K	RECQL	642 - 645	Yes (Miyamoto et al., 1997)	https://pubmed.ncbi.nlm.nih.gov/9334211/	QIP1	No	No	Unstructured region	Eliminate
WG?HLNR	RECQL4	78 - 84	No	-	-	Yes	No	Unstructured region	Eliminate
R???L??SL??RL?SLD??WL?RC	RECQL4	171 - 194	No	-	-	Yes	No	Local structure in unstructured region	Eliminate
YG?KLK?NL	RECQL4	105 - 113	No	-	-	Yes	No	Local structure in unstructured region	Eliminate
MER??DVR??L??WE??F?????RRP	RECQL4	1 - 26	No	-	-	No	Yes (insertion in alignment; 1-49 is a small domain)	Structured region	Eliminate
DV??A??E???LYREYR?LK	RECQL4	30 - 49	No	-	-	No	Yes (insertion in alignment; 1-49 is a small domain)	Structured region	Eliminate

AVELE???FR??K??N?YKA?VLKKV?EI?K	RECQL5	580 - 610	No	-	-	No	Yes (insertion in alignment to N-terminus of SLiM)	Structured region	Eliminate
AS?V	RECQL5	644 - 647	Yes (Schwendener et al., 2010)	https://pubmed.ncbi.nlm.nih.gov/20348101/	RAD51	No	No	Local structure in unstructured region	Synthesize
L??YF	RFWD3	178 - 182	No	-	-	No	No	Unstructured region	Synthesize
PPDI?TPE????K	RHNO1	154 - 166	No	-	-	No	No	Local structure in unstructured region	Synthesize
L??DTPE??YGIKVTWRRR??L??YLR?RGKL	RHNO1	198 - 229	Yes (Day et al., 2018)	https://pubmed.ncbi.nlm.nih.gov/30295604/	TOPBP1	No	No	Local structure in unstructured region	Eliminate
RK????KFP?LTF	RHNO1	84 - 96	No	-	-	No	No	Unstructured region	Synthesize
WV?P?F	RHNO1	56 - 61	No	-	-	No	No	Local structure in unstructured region	Synthesize
R?RR?RR?K?C???C	RIF1	1722 - 1734	No	-	-	No	No	Unstructured region	Synthesize
S???V??TP?????SRRQTFITLEKF??S??R??S	RIF1	1203 - 1236	No	-	-	No	No	Local structure in unstructured region	Synthesize
KE?T	RIF1	1385 - 1388	No	-	-	No	No	Unstructured region	Synthesize
PLLETLED	RIF1	11 - 18	No			No	Yes (insertion in alignment)	Structured region	See notes
L?PSDEELLA?LD	RMI1	240 - 252	Yes (Shorrocks et al., 2021)	https://pubmed.ncbi.nlm.nih.gov/33500419/	RPA1	Yes	No	Local structure in unstructured region	Eliminate
DDF?LE???E?L?L?E	RMI1	312 - 324	Yes (Shorrocks et al., 2021)	https://pubmed.ncbi.nlm.nih.gov/33500419/	N/A	Yes	No	Local structure in unstructured region	Eliminate
PPLKVL??QLR	RMI2	21 - 31	No	-	-	No	Yes (insertion in alignment)	Structured region	Eliminate
R???PDGQCTVTEV	RNF146	346 - 359	No	-	-	No	No	Local structure in unstructured region	Synthesize
D?????KRRK	RNF168	400 - 409	Yes (Pinato et al., 2009)	https://pubmed.ncbi.nlm.nih.gov/19500350/	Proposed NLS	No	No	Unstructured	Eliminate
K????G?I?KYLSP	RNF168	218 - 231	No	-	-	Yes	No	Local structure in unstructured region	Eliminate
Q?SI??M	RNF168	559 - 565	BioRiv	https://www.biorxiv.org/content/10.1101/2021.03.17.435897v1	PCNA	Yes	No	Local structure in unstructured region	Eliminate
M??LS?Q	RNF168	274 - 280	No	-	-	Yes	No	Unstructured region	Eliminate
LD???S	RNF169	522 - 527	No	-	-	No	No	Local structure in unstructured region	Synthesize

QL????HLN?G??HL	RTEL1	1015 - 1029	No	-	-	No	No	Local structure in unstructured region	Synthesize
EE??DSQ?FEFES???VFS?W?D	SETX	1074 - 1095	No	-	-	No	No	Unstructured region	Synthesize
IIISDS	SETX	1014 - 1019	No	-	-	No	No	Unstructured region	Synthesize
KLL??QD????RQ?R	SETX	1341 - 1355	No	-	-	No	No	Unstructured region	Synthesize
DPRL	SETX	2567 - 2570	No	-	-	No	No	Local structure in unstructured region	Synthesize
EE??L?LTQ?DP?DMDL?S	SETX	1492 - 1510	No	-	-	No	No	Unstructured region	Synthesize
K???GLS??II?FLL	SHLD1	196 - 210	No	-	-	No	No	Local structure in unstructured region	Synthesize
MS?RRKRA?P?RV?EE????L?WNMHED?R	SHPRH	1 - 30	No	-	-	No	No	Local structure in unstructured region	Synthesize
KRKR	SLF2	335 - 338	No	-	-	No	No	Unstructured region	Synthesize
F?ELLRSMW??E	SLX4	815 - 826	No	-	-	No	No	Local structure in unstructured region	Synthesize
MDE?D?DF??L	SLX4	44 - 54	No	-	-	No	No	Local structure in unstructured region	Synthesize
D?SWLVP?TP	SLX4	1252 - 1261	Yes (Gritenaite et al., 2014)	https://pubmed.ncbi.nlm.nih.gov/25030699/	TOPBP1	No	No	Unstructured region	Eliminate
E?V?E?EMEEIYEFAATQRKL	SLX4	844 - 864	No	-	-	No	No	Local structure in unstructured region	Synthesize
EVILL???SDEELE	SLX4	1150 - 1162	No	-	-	No	No	Unstructured region	Synthesize
Q?L?DLVDLA?EGL	SLX4	610 - 623	No	-	-	Yes	No	Local structure in unstructured region	Eliminate
V?D?E	SLX4	1395 - 1399	No	-	-	No	No	Unstructured region	Synthesize
R????VL?RMQ?FKR??P??L?H	SLX4	204 - 226	No	-	-	No	No	Local structure in unstructured region	Synthesize
M??FNLDRFRF?K	SMARCAD1	1 - 13	No	-	-	No	No	Local structure in unstructured region	Synthesize
S??KKRR?E??F	SMARCAL1	934 - 944	No	-	-	No	No	Unstructured region	Synthesize
LTEEQ?KKIEENR?KALARRAEK	SMARCAL1	5 - 27	Yes (Ciccia et al., 2009)	https://pubmed.ncbi.nlm.nih.gov/19793862/	RPA2	No	No	Local structure in unstructured region	Eliminate
I?WSSS?SD?S??D	SPIDR	99 - 112	No	-	-	No	No	Unstructured region	Synthesize

V?CPVC???VLE??INEHLD?CL	SPRTN	454 - 476	Yes (Centore et al., 2012)	https://pubmed.ncbi.nlm.nih.gov/22681887/	PCNA	No	No	Local structure in unstructured region	Eliminate
PR?SV?N?K?F??VNG?SP	SPRTN	333 - 350	No	-	-	No	No	Unstructured region	Synthesize
VIPFSGKGYVLG	SPRTN	250 - 261	Yes (Mosbech et al., 2012)	https://pubmed.ncbi.nlm.nih.gov/23042605/	p97	No	No	Local structure in unstructured region	Eliminate
L?RY	SWSAP1	179 - 182	No	-	-	No	No	Structured region	Eliminate
M?Q????GKWTISSS?DE	TDP1	1 - 17	No	-	-	No	Yes (insertion in the C-terminus of the SLiM)	Local structure in unstructured region	Synthesize
VDLT?E	TDP2	85 - 90	No	-	-	No	No	Unstructured region	Synthesize
L?E?LLDYQWDRPEI?SP?K????R	TICRR	104 - 127	No	-	-	No	No	Local structure in unstructured region	Synthesize
R??LS???L??LT?SPLLF	TICRR	1836 - 1854	No	-	-	No	No	Local structure in unstructured region	Synthesize
ELR?RSAKKRR???L?RHKSI?E??Q ?LRQIEIPK	TICRR	846 - 880	No	-	-	No	No	Local structure in unstructured region	Synthesize
TYEVELEMQASGLPKLR?KK	TICRR	1558 - 1577	No	-	-	No	No	Local structure in unstructured region	Synthesize
I?SP??L?FGAV	TICRR	1062 - 1073	No	-	-	No	No	Unstructured region	Synthesize
TP?KS	TICRR	1105 - 1109	No	-	-	No	No	Unstructured region	Synthesize
T?D?RQ??A	TICRR	1528 - 1536	No	-	-	No	No (motif not conserved in human)	Unstructured region	Eliminate
KK???LD?DEE	TIMELESS	1166 - 1176	No	-	-	No	No	Unstructured region	Synthesize
LTEE???RI??NK??AL?RR??K	TIPIN	242 - 264	Yes (Kemp et al., 2010)	https://pubmed.ncbi.nlm.nih.gov/20233725/	RPA2	No	No		Eliminate
L?DL?DYE??EDE?FPP?PPP?S	TIPIN	56 - 78	No	-	-	No	No	Local structure in unstructured region	Synthesize
REQQCGFFQ?????W?DEN??P	TOP3A	926 - 942	Yes (Martin et al., 2018)	https://pubmed.ncbi.nlm.nih.gov/30057030/	ZnF GRF	No	Yes (insertion in alignment)	Structured region	Eliminate
C?C??P?VTRTVQKDG?NKGR?FHTC	TOP3A	897 - 922	Yes (Martin et al., 2018)	https://pubmed.ncbi.nlm.nih.gov/30057030/	ZnF GRF	No	Yes (insertion in alignment)	Structured region	Eliminate
KK?R?C?LCHQPGHTR??CPQ	TOP3A	979 - 999	Yes (Martin et al., 2018)	https://pubmed.ncbi.nlm.nih.gov/30057030/	ZnF CCHC	No	No	Local structure in unstructured region	Eliminate
ADEDLLSQY	TOPBP1	486 - 494	No	-	-	Yes	No	Local structure in unstructured region	Eliminate

VTPLDM?RFQSK?F??VI	TOPBP1	778 - 795	No	-	-	Yes	No	Unstructured region	Eliminate
F?DLW	TP53	19 - 23	Yes (Chang et al., 1995)	https://pubmed.ncbi.nlm.nih.gov/7559631/	TBP	No	No	Local structure in unstructured region	Eliminate
DV??E??V	TP53BP1	959 - 966	No	-	-	No	No	Unstructured region	Synthesize
LRM?LVSP??E?S?????F?L?KP???E	TP53BP1	992 - 1019	No	-	-	No	No	Unstructured region	Synthesize
S??T	TP53BP1	1173 - 1176	No	-	-	No	No	Unstructured region	Synthesize
D???L?LS?S	TP53BP1	516 - 525	BioRiv	https://www.biorxiv.org/content/10.1101/2021.04.26.441429v1	RIF1	No	No	Unstructured region	Eliminate
EV?STQEDLF	TP53BP1	303 - 312	No	-	-	Yes	No	Unstructured region	Eliminate
SQ??FG?L?LS?SQ	TP53BP1	171 - 184	BioRiv	https://www.biorxiv.org/content/10.1101/2021.04.26.441429v1	RIF1	No	No	Unstructured region	Eliminate
DF?LP??R????K??SPRK	TP53BP1	1353 - 1370	No	-	-	No	Yes (insertion in alignment)	Unstructured region	Eliminate
P?T?PRGR?RRGRP??R	TP53BP1	1397 - 1412	No	-	-	No	Yes	Unstructured region	Synthesize
IRN?VL??K	TRAIP	389 - 397	No	-	-	No	No	Local structure in unstructured region	Synthesize
R??I?W?K????QP?I?SKF?SLGKIT	USP1	336 - 362	No	-	-	No	No	Local structure in unstructured region	Synthesize
MDTN	USP4	960 - 963	No	-	-	No	No	Unstructured region	Synthesize
E?MEHQ	USP4	662 - 667	No	-	-	No	Yes	Unstructured region	Synthesize
M??AIS?L	USP45	555 - 562	No	-	-	No	No	Local structure in unstructured region	Synthesize
D?SEK?E????E???DADSE?SE?E???K	USP45	488 - 515	No	-	-	No	No	Unstructured region	Synthesize
E?DSWEEE??GF?GS?SPVKS	WEE1	110 - 127	Yes (Watanabe et al., 1995)	https://pubmed.ncbi.nlm.nih.gov/16085715/	PLK1 & β-TrCP	No	No	Unstructured region	Eliminate
LP?W	WRN	1407 - 1410	Yes (Grundy et al., 2016)	https://pubmed.ncbi.nlm.nih.gov/27063109/	Ku70/80	No	No	Unstructured region	Eliminate
MSLDITE?EL?MLE	WRN	393 - 406	No	-	-	No	No	Local structure in unstructured region	Synthesize
ESDEE?E?EML?????L	WRN	466 - 482	Yes (Yeom, Kim & Park., 2019)	https://pubmed.ncbi.nlm.nih.gov/31570747/	RPA1	No	No	Local structure in unstructured region	Eliminate

Y??STDEE	XRCC4	229 - 236	Yes (Koch et al., 2004; Clements et al., 2004; Cherry et al., 2015)	https://pubmed.ncbi.nlm.nih.gov/15385968/; https://pubmed.ncbi.nlm.nih.gov/15380105/;https://pubmed.ncbi.nlm.nih.gov/26519825/	PNKP; APTX; APLF	No	No	Unstructured region	Eliminate
D??DIAP?RKRR?R	XRCC4	262 - 275	No		-	Yes	No	Unstructured region	Eliminate
LF?EI	XRCC4	332 - 336	No	-	-	Yes	No	Unstructured region	Eliminate
VS??IRS?AQILALL	ZGRF1	245 - 259	No	-	-	No	No	Local structure in unstructured region	Synthesize
P???LP	ZGRF1	1067 - 1072	No	-	-	No	No	Unstructured region	Synthesize
Q?????W?KYQN	ZGRF1	1143 - 1151	No	-	-	No	No	Unstructured region	Synthesize
V?F?L	ZGRF1	530 - 534	No	-	-	No	No	Unstructured region	Synthesize
D??L?EAI?RSLEE	ZNF451	1047 - 1060	No	-	-	No	No	Local structure in unstructured region	Synthesize
W?C??CTY?N???L??CEMCE?P	ZRANB3	625 - 647	Yes (Ciccia et al., 2012)	https://pubmed.ncbi.nlm.nih.gov/22704558/	PCNA	No	No	Local structure in unstructured region	Eliminate
KR?K	ZRANB3	576 - 579	No	-	-	No	No	Unstructured region	Synthesize

5.3.4 Use of AlphaFold to validate SLiMs Predicted by the SLiMCon Tool

At the time of writing this thesis, the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk) was released as an openly accessible, extensive database of high accuracy protein structure predictions (Jumper et al., 2021; Varadi et al., 2022). The initial release of AlphaFold contains over 360,000 predicted structures across 21 model organisms. By supplying the protein or gene name as an input, the user can visualise the predicted structures of the protein of interest. Analysis of the predicted structure of BLM using AlphaFold (**Fig. 5.11**) revealed the expected the structured helicase domain (blue structured region in the centre) and the unstructured N-terminus (orange, unstructured/"string" like region). Interestingly, all three SLiMs outlined in Chapter 3 appeared as locally structured regions within a larger unstructured region (highlighted in **Fig. 5.11**). As an example, SLiM 3 from BLM is highlighted in Fig 5.6A. This reflects the fact that whilst SLiMs are found to be intrinsic regions of disorder, they may adopt local structure and explains why the IUPred score of many SLiMs is lower than the surrounding flanking regions.

Next, this analysis was extended to various SLiMs identified in literature, as well as some examples from Chapter 3 found to have high-confidence candidate interacting proteins (**Table 5.4**). In all cases, each SLiM was predicted to either be locally structured but within a larger unstructured region or be unstructured. No SLiMs were predicted to be part of a larger structured region.

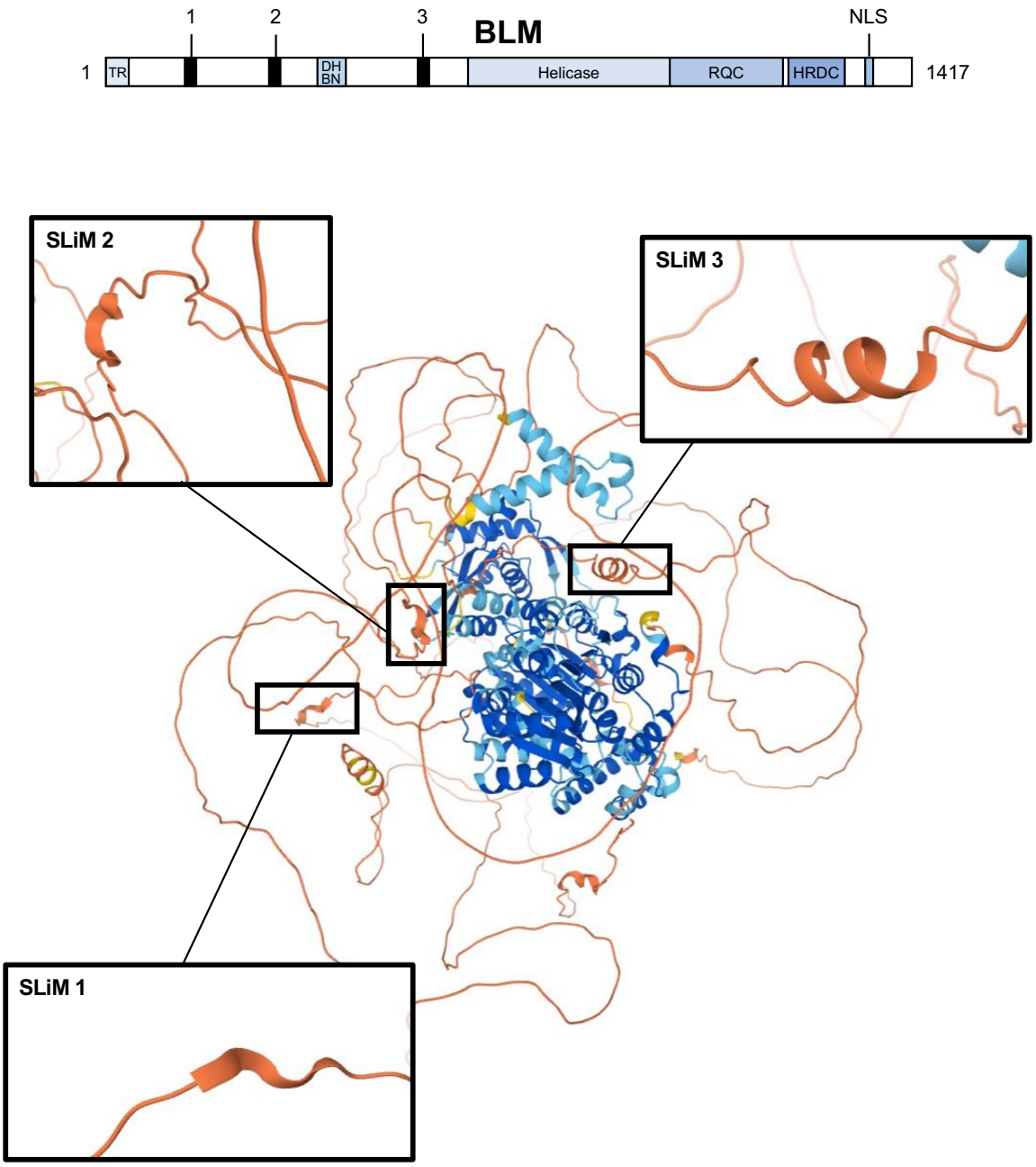


Fig. 5.11 AlphaFold prediction of BLM.
Structured helicase domain of BLM is shown in blue. Unstructured regions of BLM are shown in orange. Zoomed in panels display the three SLiMs outlined in Chapter 3. Each SLiM appeared as a locally structured region within a larger unstructured region (Jumper et al., 2021; Varadi et al., 2022).

Table 5.4. AlphaFold predictions of published and synthesised SLiMs

Gene	Residues	Alphafold
APLF	182 - 191	Local structure in unstructured region
ATRIP	247 - 253	Local structure in unstructured region
AUNIP	341 - 347	Local structure in unstructured region
BARD1	173 - 193	Local structure in unstructured region
BLM	153 - 161	Local structure in unstructured region
BLM	299 - 310	Local structure in unstructured region
BLM	554 - 563	Local structure in unstructured region
BRIP1	1147 - 1157	Unstructured region
CYREN	9 - 17	Local structure in unstructured region
CYREN	142 - 157	Local structure in unstructured region
ERCC6L2	788 - 796	Local structure in unstructured region
ETAA1	892 - 912	Local structure in unstructured region
MDC1	213 - 220	Unstructured region
MDC1	473 - 479	Unstructured region
NHEJ1	288 - 298	Local structure in unstructured region
PAXX	177 - 191	Local structure in unstructured region
RAD9A	307 - 324	Unstructured region
RBBP8	327 - 331	Local structure in unstructured region
RHNO1	198 - 229	Local structure in unstructured region
RMI1	240 - 252	Local structure in unstructured region
RNF168	559 - 565	Local structure in unstructured region
SMARCAL1	5 - 27	Local structure in unstructured region
TIPIN	242 - 264	Local structure in unstructured region
TOPBP1	778 - 795	Unstructured region
TP53	19 - 23	Local structure in unstructured region
TP53BP1	303 - 312	Unstructured region
WRN	1407 - 1410	Unstructured region
XRCC4	229 - 236	Unstructured region

Given these findings, the predicted structure of all SLiMs predicted by the SLiMCon tool was analysed using AlphaFold (**Table 5.3**). Each SLiM highlighted green in Table 5.3,

each SLiM already published (highlighted purple) or those that have already been synthesised (highlighted blue), was predicted to either be locally structured but within a larger unstructured region or be unstructured. No SLiMs were predicted to be part of a larger structured domain. Interestingly, each SLiM designated to be a false positive during the quality control step was predicted to be part of a structured domain. Taken together, these AlphaFold predictions increased confidence in the predictive power of the SLiMCon tool.

In summary, the work in this chapter has developed a tool that can automatically predict SLiMs in any protein of interest in the human proteome with high accuracy. Applying this tool to a wider range of DDR proteins, 222 SLiMs were predicted. 64% of these are novel, previously uncharacterised putative SLiMs that may form key interaction interfaces in the DDR.

5.3 Discussion

In this chapter, I aimed to overcome the limitations of manually identifying SLiMs in the DDR in order to extend the screen and develop a DNA repair interactome based on SLiMs. Analysis of existing SLiM discovery algorithms revealed their limitations; thus, this chapter developed a web-based tool to predict SLiMs in DDR proteins, which combines the criteria used to effect in Chapter 3.

The SLiMCon tool has many advantages to other web-based SLiM discovery tools. It automatically eliminates putative SLiMs based on contextual information, such as evolutionary conservation and its location within an intrinsically disordered region of the protein of interest. Furthermore, the parameters used for the automatic identification of SLiMs were defined according to published examples, as well as putative SLiMs identified in chapter 3. Unlike ELM, ScanSite and SLiMSearch, it is not restricted to the identification of SLiMs according to

specific, pre-defined consensus sequences. In addition, it does not have the limitation of being labour intensive as was found for manual alignments in Chapter 3. This means that it is far more suitable to be applied to a larger range of proteins, which will facilitate the development of an interactome based on SLiMs.

This does not mean that the SLiMCon tool is without flaws. A major flaw is the fact that the tool cannot distinguish between an insertion in the sequence alignment and an unconserved region of the protein. Fourteen out of the fifteen false positive incidences were wrongly assigned because of an insertion in the sequence alignment, which was recognised by the tool to be a unconserved flank. To overcome this, an alternative alignment can be generated that omits the orthologue with the insertion. In the example given in **Fig. 5.9A**, the Armadillo sequence could be omitted. Furthermore, it is impossible to develop default parameters that are inclusive for the identification of every SLiM in the human proteome. For example, a SLiM located in the C-terminus of RAD9 can interact with TOPBP1 (Day et al., 2018). However, within the multiple sequence alignment, this SLiM is located eleven amino acids away from a structured domain. Consequently, using the default parameters it will not be identified. The flank length could be shortened to allow the inclusion of this SLiM, but that would result in the identification of many false positive SLiMs.

To expand on the work in this chapter, the SLiMCon tool could be expanded to include additional characteristics of interest. Currently, it only identifies whether there is a consensus sequence for phosphorylation by PIKKs, CDKs or CK2. However, other ELM classes could be incorporated and therefore identified within the SLiMs the tool predicts, whilst remaining less restricted that the ELM database. Further work building on this chapter would involve the synthesis of the SLiMs predicted by the SLiMCon tool as biotinylated peptides (as outlined in Chapter 3). These could be used as baits in peptide pulldowns in HeLa nuclear extract and mass spectrometry used to build a DNA repair interactome based on SLiMs.

Protein-protein interactions mediated by SLiMs play essential roles in many biological processes, including but not limited to the DDR. The known instances of SLiMs mediate a wide range of important cellular tasks: including cellular signalling, providing specificity for post-translation modifications, targeting proteins to subcellular compartments, acting as recognition sites for proteolytic cleavage, membrane binding, protein folding and cell adhesion (Bhowmick et al., 2015). The SLiMCon tool contains sequences alignments and SLiM predictions for every protein in the human proteome. Given the relevance of SLiMs to a great number of biological processes, the SLiMCon tool is not just useful for members of the DDR field but all biologists hoping to identify novel protein-protein interactions.

Chapter 6: General Discussion

6.1 Summary of Work Presented in this DPhil

6.1.1. Rationale of Project

Like all biological pathways, the DNA damage response is facilitated by a series of protein-protein interactions, but the mechanisms by which DNA repair protein complexes are assembled, disassembled and regulated by post-translational modifications are still poorly understood. Historically, most research has investigated protein-proteins interactions between structured domains, but it has become clear that a large proportion of functional protein-protein interactions are mediated by SLiMs. With the hope of developing a greater understanding of the diverse range of pathologies associated with mutations in DNA repair genes and the development of new therapies, the aim of thesis was to identify novel DNA repair mechanisms using SLiM proteomics. A summary of the three main aspects of this DPhil is given below.

6.1.2. Development of a Combined Bioinformatics and Proteomics Approach to Identify SLiMs in DDR Proteins

Towards the aim described above, a combined bioinformatics and proteomics approach was developed to identify and characterize novel protein-protein interactions in DNA damage response proteins relying on SLiMs. These methods have been previously been used to identify, map and characterize novel individual protein-protein interactions for specific DNA repair proteins (Blackford et al., 2015; Ochi et al., 2015; Polo et al., 2012; Stucki et al., 2005). However, this is the first use of these methods to identify SLiMs within the DDR on a larger scale.

Using this approach 36 putative SLiMs were identified and when peptides corresponding to these SLiMs were used in peptide pulldown experiments for mass spectrometry, 81% (n = 29) had high-confidence candidate interacting proteins. Of these, 55% (n = 16) had DDR proteins in the top ten candidate interacting proteins. Due to time constraints, not all of these candidate interactions could be validated in the context of the full-length protein but SLiM-mediated interactions between MDC1 and TOPBP1, TOPBP1 and CIP2A, AUNIP and CtIP, as well as between the BTR complex and RPA were identified and validated. Further work conducted by the Blackford lab and our collaborators has demonstrated that the interaction between MDC1 and TOPBP1 (Leimbacher et al., 2019), as well as the interaction between TOPBP1 and CIP2A promotes chromosomal stability during mitosis (Adam et al., 2021).

6.1.3. The Bloom Syndrome Complex Senses RPA Coated Single-Stranded DNA to Restart Stalled Replication Forks

The Bloom syndrome helicase BLM interacts with TOP3A, RMI1 and RMI2 to form the BTR complex. The best characterised cellular function of the BTR complex is in dHJ dissolution to produce non crossover HR products (Wu and Hickson, 2003). BLM also promotes DNA end resection, the resolution of ultrafine DNA bridges between sister chromatids in anaphase and the promotes restart of stalled replication forks (Chan et al., 2007; Davies et al., 2007; Gravel et al., 2008). Three SLiMs within the BTR complex were shown to interact cooperatively with RPA. Functional investigation of this interaction demonstrated that RPA-binding is required to promote stable BLM recruitment to sites of DNA replication stress and for fork restart but not promoting dHJ dissolution, DNA end resection or the resolution of ultrafine DNA bridges. These findings suggest a model whereby the BTR complex has an intrinsic ability to sense abundance of RPA-ssDNA at replication forks, regulating BLM recruitment and function in

response to replication stress. This work perfectly fit the aim of this DPhil, as SLiM proteomics was used to identify a novel DNA repair mechanism.

6.1.4. Development of a Web-Based Tool to Identify SLiMs

To overcome the limitations of manually identifying SLiMs, web-based tools that predict SLiMs automatically can be used. However, analysis of pre-existing SLiM discovery algorithms revealed they had varying success in detecting the validated SLiMs in the BTR complex. Accordingly, the SLiMCon web-based tool was developed, which has generated multiple sequence alignments for every protein in the human proteome automatically and can predict putative SLiMs within these generated sequence alignments. The SLiMCon tool was applied to 120 DDR proteins and within these proteins, 222 SLiMs were predicted. 64% of these are novel, previously uncharacterised putative SLiMs that may form key interaction interfaces in the DDR. This work forms the foundation for the development of DNA repair interactome based on SLiMs. Significantly, given the relevance of SLiMs to a great number of biological processes, the SLiMCon tool is not just useful for members of the DDR field but all biologists hoping to identify novel protein-protein interactions.

6.2 Future Work

The experiments described below are larger scale future work. More specific future directions can be found within the discussion of each chapter. In Chapter 3, 16 candidate SLiM based interactions between DDR proteins were identified. Due to time constraints, not all of these candidate interactions could be validated and investigated further. Future work by the Blackford lab will involve the validation of these SLiM-based interactions in the context of the full-length protein. As with the validation experiments in Chapter 3, this will require expression of epitope-tagged proteins, containing mutations in the SLiM within the context of the full-

length protein, to determine whether such mutations disrupt binding in pulldowns from cell extracts. Once the SLiM-based interaction has been validated, it will be crucial to establish whether cells expressing the mutant protein display any phenotypic defects. Phenotypic assays can then be used to identify novel DNA repair mechanisms.

In Chapter 5, using the SLiMCon tool, 222 putative SLiMs were predicted. 64% of these are novel putative SLiMs that may form key interaction interfaces in the DDR. These SLiMs will be synthesised as biotinylated peptides and used as baits in HeLa nuclear extract in peptide pulldowns for mass spectrometry (as in Chapter 3). Identification of the interacting partners of these SLiMs will facilitate the generation of a DNA repair interactome based on SLiMs. As well as the work in this DPhil, this will generate a greater understanding of the diverse range of pathologies associated with mutations in DNA repair genes and form a foundation for the development of new therapies.

6.3 Final Remarks

In the face of various forms of genotoxic stress the DDR is essential for maintaining genome integrity (Jackson and Bartek, 2009). The importance of the DDR in both promoting various physiological, cellular processes and circumventing human disease is demonstrated by the broad range of pathologies associated with mutations in DNA repair genes, such as cancer, immunodeficiency, infertility, premature ageing and microcephaly. The DDR has also become an attractive and promising strategic target for novel cancer therapy, as the exploitation of these pathways can provide a platform to develop novel anti-cancer drugs that can act alone or as chemo- and/or radiosensitisers to enhance the therapeutic response of current conventional DNA-damaging therapies. Due to their smaller interaction interface, interactions based on SLiMs are inherently druggable and molecular compounds have shown promising results for targeted treatment strategies (Uyar et al., 2014). For example, idasanutlin (a Nutlin-derivative)

that competes for binding to the p53-binding site on MDM2, has entered clinical trials for acute myeloid leukaemia (Montesinos et al., 2020). Furthermore, recent work has shown that disruption of the interaction between TOPBP1 and CIP2A is toxic in *BRCA*-deficient tumours, suggesting it represents an attractive synthetic lethal therapeutic target for *BRCA1*- and *BRCA2*-mutated cancers (Adam et al., 2021). Therefore, as well as providing key insight into DNA repair mechanisms, it is my hope that protein-protein interactions identified here and in other studies of this kind will provide the foundation for the development of new anti-cancer treatments, translating this work from the lab bench to bedside.

Chapter 7: References

Adam, S., Rossi, S.E., Moatti, N., Zompit, M. de M., Xue, Y., Ng, T.F., Álvarez-Quilón, A., Desjardins, J., Bhaskaran, V., Martino, G., et al. (2021). The CIP2A–TOPBP1 axis safeguards chromosome stability and is a synthetic lethal target for BRCA-mutated cancer. Nature Cancer *2*, 1357–1371.

Agrawal, A., and Schatz, D.G. (1997). RAG1 and RAG2 Form a Stable Postcleavage Synaptic Complex with DNA Containing Signal Ends in V(D)J Recombination. Cell *89*, 43–53. https://doi.org/10.1016/s0092-8674(00)80181-6.

Aguzzi, A., and Altmeyer, M. (2016). Phase Separation: Linking Cellular Compartmentalization to Disease. Trends in Cell Biology *26*, 547–558.

Ahnesorg, P., Smith, P., and Jackson, S.P. (2006). XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. Cell *124*, 301–313. https://doi.org/10.1016/j.cell.2005.12.031.

Ali, I.U., Lidereau, R., Theillet, C., and Callahan, R. (1987). Reduction to Homozygosity of Genes on Chromosome 11 in Human Breast Neoplasia. Science *238*, 185–188. https://doi.org/10.1126/science.3659909.

Allen, G., Buxton, R.B., Wong, E.C., and Courchesne, E. (1997). Attentional Activation of the Cerebellum Independent of Motor Involvement. Science *275*, 1940–1943. https://doi.org/10.1126/science.275.5308.1940.

Allikmets, R., Schriml, L.M., Hutchinson, A., Romano-Spica, V., and Dean, M. (1998). A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance. Cancer Research *58*, 5337–5339. .

Amé, J.C., Rolli, V., Schreiber, V., Niedergang, C., Apiou, F., Decker, P., Muller, S., Höger, T., Murcia, J. Ḿenissier-de, and Murcia, G. de (1999). PARP-2, A novel mammalian DNA damage-dependent poly(ADP-ribose) polymerase. Journal of Biological Chemistry *274*, 17860–17868. .

Anand, R., Ranjha, L., Cannavo, E., and Cejka, P. (2016). Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. Mol Cell *64*, 940–950. https://doi.org/10.1016/j.molcel.2016.10.017.

Anand, R.P., Tsaponina, O., Greenwell, P.W., Lee, C.-S., Du, W., Petes, T.D., and Haber, J.E. (2014). Chromosome rearrangements via template switching between diverged repeated sequences. Genes \& Development *28*, 2394–2406.

Aranda-Anzaldo, A. (2012). The post-mitotic state in neurons correlates with a stable nuclear higher-order structure. Communicative \& Integrative Biology *5*, 134–139.

Arcas, A., Fernandez-Capetillo, O., Cases, I., and Rojas, A.M. (2014). Emergence and evolutionary analysis of the human DDR network: implications in comparative genomics and downstream analyses. Molecular Biology and Evolution *31*, 940–961.

Arnoult, N., Correia, A., Ma, J., Merlo, A., Garcia-Gomez, S., Maric, M., Tognetti, M., Benner, C.W., Boulton, S.J., Saghatelian, A., et al. (2017). Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. Nature *549*, 548–552.

Atamna, H., Cheung, I., and Ames, B.N. (2000). A method for detecting abasic sites in living cells: Age-dependent changes in base excision repair. Proc National Acad Sci *97*, 686–691. https://doi.org/10.1073/pnas.97.2.686.

Audebert, M., Salles, B., and Calsou, P. (2004). Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. J. Biol. Chem. *279*, 55117–55126.

Bae, K., Kim, H., Bae, S., Kang, H., Brill, S., and Seo, Y. (2003). Bimodal interaction between replication-protein A and Dna2 is critical for Dna2 function both in vivo and in vitro. Nucleic Acids Res *31*, 3006–3015. https://doi.org/10.1093/nar/gkg422.

Bae, S.-H., Bae, K.-H., Kim, J.-A., and Seo, Y.-S. (2001). RPA governs endonuclease switching during processing of Okazaki fragments in eukaryotes. Nature *412*, 456–461. https://doi.org/10.1038/35086609.

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J.W., and Elledge, S.J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell *86*, 263–274. .

Bailey, S.M., Cornforth, M.N., Ullrich, R.L., and Goodwin, E.H. (2004). Dysfunctional mammalian telomeres join with DNA double-strand breaks. Dna Repair *3*, 349–357. https://doi.org/10.1016/j.dnarep.2003.11.007.

Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Research *43*, W39–W49.

Balakrishnan, L., and Bambara, R.A. (2013). Okazaki fragment metabolism. Cold Spring Harbor Perspectives in Biology *5*, a010173.

Bao, S., Wu, Q., McLendon, R.E., Hao, Y., Shi, Q., Hjelmeland, A.B., Dewhirst, M.W., Bigner, D.D., and Rich, J.N. (2006). Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. Nature *444*, 756–760.

Baris, H.N., Kedar, I., Halpern, G.J., Shohat, T., Magal, N., Ludman, M.D., and Shohat, M. (2007). Prevalence of breast and colorectal cancer in Ashkenazi Jewish carriers of Fanconi anemia and Bloom syndrome. Israel Medical Assoc J Imaj *9*, 847–850. .

Barnieh, F.M., Loadman, P.M., and Falconer, R.A. (2021). Progress towards a clinically-successful ATR inhibitor for cancer therapy. Current Research in Pharmacology and Drug Discovery *2*, 100017.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR Provides Aquired Resistance Against Viruses in Prokaryotes. Science *315*, 1709–1712.

Bartkova, J., Hořejší, Z., Koed, K., Krämer, A., Tort, F., Zieger, K., Guldberg, P., Sehested, M., Nesland, J.M., Lukas, C., et al. (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature *434*, 864–870. https://doi.org/10.1038/nature03482.

Bartkova, J., Rezaei, N., Liontos, M., Karakaidos, P., Kletsas, D., Issaeva, N., Vassiliou, L.-V.F., Kolettas, E., Niforou, K., Zoumpourlis, V.C., et al. (2006). Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. Nature *444*, 633–637. https://doi.org/10.1038/nature05268.

Bass, T.E., Luzwick, J.W., Kavanaugh, G., Carroll, C., Dungrawala, H., Glick, G.G., Feldkamp, M.D., Putney, R., Chazin, W.J., and Cortez, D. (2016). ETAA1 acts at stalled

Baumann, P., Benson, F.E., and West, S.C. (1996). Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions in vitro. Cell *87*, 757–766.

Beck, D.B., Oda, H., Shen, S.S., and Reinberg, D. (2012). PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. Genes \& Development *26*, 325–337.

Becker, J.R., Clifford, G., Bonnet, C., Groth, A., Wilson, M.D., and Chapman, J.R. (2021). BARD1 reads H2A lysine 15 ubiquitination to direct homologous recombination. Nature *596*, 433–437.

Beeche, M., Bodnar, A.G., Wahl, G.M., Tlsty, T.D., et al. (1999). Telomerase expression in human somatic cells does not induce changes associated with a transformed phenotype. Nature Genetics *21*, 111–114.

Bekker-Jensen, S., Fugger, K., Danielsen, J.R., Gromova, I., Sehested, M., Celis, J., Bartek, J., Lukas, J., and Mailand, N. (2007). Human Xip1 (C2orf13) Is a Novel Regulator of Cellular Responses to DNA Strand Breaks*. J Biol Chem *282*, 19638–19643. https://doi.org/10.1074/jbc.c700060200.

Benedict, W.F., Murphree, A.L., Banerjee, A., Spina, C.A., Sparkes, M.C., and Sparkes, R.S. (1983). Patient with 13 Chromosome Deletion: Evidence That the Retinoblastoma Gene Is a Recessive Cancer Gene. Science *219*, 973–975. https://doi.org/10.1126/science.6336308.

Benjamin, R.C., and Gill, D.M. (1980). Poly(ADP-ribose) synthesis in vitro programmed by damaged DNA. A comparison of DNA molecules containing different types of strand breaks. Journal of Biological Chemistry *255*, 10502–10508. .

Bennardo, N., Cheng, A., Huang, N., and Stark, J.M. (2008). Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genetics *4*, e1000110.

Bentley, N.J., Holtzman, D.A., Flaggs, G., Keegan, K.S., DeMaggio, A., Ford, J.C., Hoekstra, M., and Carr, A.M. (1996). The Schizosaccharomyces pombe rad3 checkpoint gene. The EMBO Journal *15*, 6641–6651. .

Benz, C., Ali, M., Krystkowiak, I., Simonetti, L., Sayadi, A., Mihalic, F., Kliche, J., Andersson, E., Jemth, P., Davey, N.E., et al. (2022). Proteome-scale mapping of binding sites in the unstructured regions of the human proteome. Mol Syst Biol *18*, e10584.

Berti, M., Ray Chaudhuri, A., Thangavel, S., Gomathinayagam, S., Kenig, S., Vujanovic, M., Odreman, F., Glatter, T., Graziano, S., Mendoza-Maldonado, R., et al. (2013). Human RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition. Nature Structural \& Molecular Biology *20*, 347–354.

Bertolin, A.P., Hoffmann, J.-S., and Gottifredi, V. (2020). Under-Replicated DNA: The Byproduct of Large Genomes? Cancers (Basel) *12*, 2764.

Bertolin, A.P., Hoffmann, J.-S., and Gottifredi, V. (2020). Under-Replicated DNA: The Byproduct of Large Genomes? Cancers *12*, 2764. https://doi.org/10.3390/cancers12102764.

Bétermier, M., Bertrand, P., and Lopez, B.S. (2014). Is non-homologous end-joining really an inherently error-prone process? PLoS Genetics *10*, e1004086.

Bétous, R., Couch, F.B., Mason, A.C., Eichman, B.F., Manosas, M., and Cortez, D. (2013). Substrate-selective repair and restart of replication forks by DNA translocases. CellReports *3*, 1958–1969.

Bétous, R., Mason, A.C., Rambo, R.P., Bansbach, C.E., Badu-Nkansah, A., Sirbu, B.M., Eichman, B.F., and Cortez, D. (2012). SMARCAL1 catalyzes fork regression and Holliday junction migration to maintain genome stability during DNA replication. Genes \& Development *26*, 151–162.

Bhargava, R., Onyango, D.O., and Stark, J.M. (2016). Regulation of Single-Strand Annealing and its Role in Genome Maintenance. Trends in Genetics : TIG *32*, 566–575.

Bhat, K.P., and Cortez, D. (2018). RPA and RAD51: fork reversal, fork protection, and genome stability. Nature Structural \& Molecular Biology *25*, 446–453.

Bhowmick, P., Guharoy, M., and Tompa, P. (2015). Bioinformatics Approaches for Predicting Disordered Protein Motifs. Adv Exp Med Biol *870*, 291–318.

Bhowmick, R., Minocherhomji, S., and Hickson, I.D. (2016). RAD52 Facilitates Mitotic DNA Synthesis Following Replication Stress. Molecular Cell *64*, 1117–1126.

Bizard, A.H., and Hickson, I.D. (2014). The Dissolution of Double Holliday Junctions. Cold Spring Harb Perspect Biol *6*, 1–14.

Bjorn Schwanhausser., Dorothea Busse., Na Li., Gunnar Dittmar., Johannes Schuchhardt., Jana Wolf., Wei Chen., and Matthias Selbach (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Blackford, A.N., and Jackson, S.P. (2017). ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. Molecular Cell *66*, 801–817.

Blackford, A.N., and Stucki, M. (2020). How Cells Respond to DNA Breaks in Mitosis. Trends in Biochemical Sciences *45*, 321–331.

Blackford, A.N., Nieminuszczy, J., Schwab, R.A., Galanty, Y., Jackson, S.P., and Niedzwiedz, W. (2015). TopBP1 Interacts with BLM to Maintain Genome Stability but Is Dispensable for Preventing BLM Degradation. Molecular Cell *57*, 1133–1141.

Blastyák, A., Hajdu, I., Unk, I., and Haracska, L. (2010). Role of double-stranded DNA translocase activity of human HLTF in replication of damaged DNA. MCB *30*, 684–693.

Bloom, D. (1954). Congenital Telangiectatic Erythema Resembling Lupus Erythematosus In Dwarfs. AMA Am J Dis Child *88*, 754–758.

Bochkareva, E., Kaustov, L., Ayed, A., Yi, G.-S., Lu, Y., Pineda-Lucena, A., Liao, J.C.C., Okorokov, A.L., Milner, J., Arrowsmith, C.H., et al. (2005). Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. Proceedings of the National Academy of Sciences *102*, 15412–15417.

Bochman, M.L., Bell, S.P., and Schwacha, A. (2008). Subunit Organization of Mcm2-7 and the Unequal Role of Active Sites in ATP Hydrolysis and Viability. Mol Cell Biol *28*, 5865–5873. https://doi.org/10.1128/mcb.00161-08.

Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. Nature Reviews Genetics *13*, 770–780.

BODER, E., and SEDGWICK, R.P. (1958). Ataxia-telangiectasia; a familial syndrome of progressive cerebellar ataxia, oculocutaneous telangiectasia and frequent pulmonary infection. Pediatrics *21*, 526–554. .

Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., C, C., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. (1998). Extension of Life-Span by Introduction of Telomerase into Normal Human Cells. Science *279*, 349–352.

Boerkoel, C.F., Takashima, H., John, J., Yan, J., Stankiewicz, P., Rosenbarker, L., André, J.-L., Bogdanovic, R., Burguet, A., Cockfield, S., et al. (2002). Mutant chromatin remodeling protein SMARCAL1 causes Schimke immuno-osseous dysplasia. Nat Genet *30*, 215–220. https://doi.org/10.1038/ng821.

Boersma, V., Moatti, N., Segura-Bayona, S., Peuscher, M.H., Torre, J.V.D., Wevers, B.A., Orthwein, A., Durocher, D., and Jacobs, J.J.L. (2015). MAD2L2 controls DNA repair at telomeres and DNA breaks by inhibiting 5' end resection. Nature *521*, 537–540.

Bononi, A., Goto, K., Ak, G., Yoshikawa, Y., Emi, M., Pastorino, S., Carparelli, L., Ferro, A., Nasu, M., Kim, J.-H., et al. (2020). Heterozygous germline BLM mutations increase susceptibility to asbestos and mesothelioma. Proc National Acad Sci *117*, 33466–33473. https://doi.org/10.1073/pnas.2019652117.

Boom, J. van den, Wolf, M., Weimann, L., Schulze, N., Li, F., Kaschani, F., Riemer, A., Zierhut, C., Kaiser, M., Iliakis, G., et al. (2016). VCP/p97 Extracts Sterically Trapped Ku70/80 Rings from DNA in Double-Strand Break Repair. Molecular Cell *64*, 189–198. .

Botuyan, M.V., Lee, J., Ward, I.M., Kim, J.-E., Thompson, J.R., Chen, J., and Mer, G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell *127*, 1361–1373.

Brack, C., Hirama, M., Lenhard-Schuller, R., and Tonegawa, S. (1978). A complete immunoglobulin gene is created by somatic recombination. Cell *15*, 1–14. https://doi.org/10.1016/0092-8674(78)90078-8.

Brambati, A., Barry, R.M., and Sfeir, A. (2020). DNA polymerase theta (Polθ) - an error-prone polymerase necessary for genome stability. Current Opinion in Genetics \& Development *60*, 119–126.

Branzei, D., Sollier, J., Liberi, G., Zhao, X., Maeda, D., Seki, M., Enomoto, T., Ohta, K., and Foiani, M. (2006). Ubc9- and mms21-mediated sumoylation counteracts recombinogenic events at damaged replication forks. Cell *127*, 509–522.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a Cancer Journal for Clinicians *68*, 394–424.

Brookes, P., and Lawley, P.D. (1961). The reaction of mono- and di-functional alkylating agents with nucleic acids. The Biochemical Journal *80*, 496–503. .

Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature *434*, 913–917. .

Buck, D., Malivert, L., Chasseval, R. de, Barraud, A., Fondanèche, M.-C., Sanal, O., Plebani, A., Stéphan, J.-L., Hufnagel, M., Deist, F. le, et al. (2006). Cernunnos, a novel

nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. Cell *124*, 287–299. https://doi.org/10.1016/j.cell.2005.12.030.

Buck, D., Malivert, L., Chasseval, R. de, Barraud, A., Fondanèche, M.-C., Sanal, O., Plebani, A., Stéphan, J.-L., Hufnagel, M., Deist, F.L., et al. (2006). Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. Cell *124*, 287–299.

Bunting, S.F., CallEn, E., Wong, N., Chen, H.-T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., et al. (2010). 53BP1 Inhibits Homologous Recombination in Brca1-Deficient Cells by Blocking Resection of DNA Breaks. Cell *141*, 243–254. https://doi.org/10.1016/j.cell.2010.03.012.

Burg, M. van der, Veelen, L.R. van, Verkaik, N.S., Wiegant, W.W., Hartwig, N.G., Barendregt, B.H., Brugmans, L., Raams, A., Jaspers, N.G.J., Zdzienicka, M.Z., et al. (2006). A new type of radiosensitive T–B–NK+ severe combined immunodeficiency caused by a LIG4 mutation. J Clin Invest *116*, 137–145. https://doi.org/10.1172/jci26121.

Burgers, P.M.J., and Kunkel, T.A. (2016). Eukaryotic DNA Replication Fork. Annu Rev Biochem *86*, 1–22. https://doi.org/10.1146/annurev-biochem-061516-044709.

Bussen, W., Raynard, S., Busygina, V., Singh, A.K., and Sung, P. (2007). Holliday junction processing activity of the BLM-Topo IIIalpha-BLAP75 complex. J. Biol. Chem. *282*, 31484–31492.

Byrnes, J.J., Downey, K.M., and So, A.G. (1973). Bone marrow cytoplasmic deoxyribonucleic acid polymerase. Variation of pH and ionic environment as a possible control mechanism. Biochemistry-Us *12*, 4378–4384. https://doi.org/10.1021/bi00746a013.

Byrnes, J.J., Downey, K.M., Black, V.L., and So, A.G. (1976). A new mammalian DNA polymerase with 3’ to 5’ exonuclease activity: DNA polymerase δ. Biochemistry-Us *15*, 2817–2823. https://doi.org/10.1021/bi00658a018.

Byun, T.S., Pacek, M., Yee, M., Walter, J.C., and Cimprich, K.A. (2005a). Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. Genes \& Development *19*, 1040–1052. .

Bzymek, M., Thayer, N.H., Oh, S.D., Kleckner, N., and Hunter, N. (2010). Double Holliday junctions are intermediates of DNA break repair. Nature *464*, 937–941.

Cadet, J., Davies, K.J.A., Medeiros, M.H., Mascio, P.D., and Wagner, J.R. (2017). Formation and repair of oxidatively generated damage in cellular DNA. Free Radical Biology \& Medicine *107*, 13–34.

Cannavo, E., and Cejka, P. (2014). Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. Nature *514*, 122–125.

Cannavo, E., and Cejka, P. (2014). Sae2 promotes dsDNA endonuclease activity within Mre11–Rad50–Xrs2 to resect DNA breaks. Nature *514*, 122–125. https://doi.org/10.1038/nature13771.

Cardoso, M.C., Leonhardt, H., and Nadal-Ginard, B. (1993). Reversal of terminal differentiation and control of DNA replication: cyclin A and Cdk2 specifically localize at subnuclear sites of DNA replication. Cell *74*, 979–992. .

Carter, T., Vancurová, I., Sun, I., Lou, W., and DeLeon, S. (1990). A DNA-activated protein kinase from HeLa cell nuclei. Molecular and Cellular Biology *10*, 6460–6471.

Carter, T., Vancurová, I., Sun, I., Lou, W., and DeLeon, S. (1990). A DNA-activated protein kinase from HeLa cell nuclei. Molecular and Cellular Biology *10*, 6460–6471. .

Cavenee, W.K., Dryja, T.P., Phillips, R.A., Benedict, W.F., Godbout, R., Gallie, B.L., Murphree, A.L., Strong, L.C., and White, R.L. (1983). Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. Nature *305*, 779–784. https://doi.org/10.1038/305779a0.

Cejka, P., and Symington, L.S. (2021). DNA End Resection: Mechanism and Control. Annu Rev Genet *55*, 285–307. https://doi.org/10.1146/annurev-genet-071719-020312.

Cejka, P., Plank, J.L., Bachrati, C.Z., Hickson, I.D., and Kowalczykowski, S.C. (2010). Rmi1 stimulates decatenation of double Holliday junctions during dissolution by Sgs1-Top3. Nature Structural \& Molecular Biology *17*, 1377–1382.

Celeste, A., Petersen, S., Romanienko, P.J., Fernandez-Capetillo, O., Chen, H.-T., Sedelnikova, O.A., Reina-San-Martin, B., Coppola, V., Meffre, E., Difilippantonio, M.J., et al. (2002). Genomic instability in mice lacking histone H2AX. Science *296*, 922–927.

Cescutti, R., Negrini, S., Kohzaki, M., and Halazonetis, T.D. (2010). TopBP1 functions with 53BP1 in the G1 DNA damage checkpoint. The EMBO Journal *29*, 3723–3732.

CHAMBON, P., WEILL, J.D., and MANDEL, P. (1963). Nicotinamide mononucleotide activation of new DNA-dependent polyadenylic acid synthesizing nuclear enzyme. Biochemical and Biophysical Research Communications *11*, 39–43. .

Chan, K.-L., North, P.S., and Hickson, I.D. (2007). BLM is required for faithful chromosome segregation and its localization defines a class of ultrafine anaphase bridges. The EMBO Journal *26*, 3397–3409.

Chan, Y.W., and West, S.C. (2014). Spatial control of the GEN1 Holliday junction resolvase ensures genome stability. Nature Communications *5*, 4844–4811.

Chang, H.H.Y., Pannunzio, N.R., Adachi, N., and Lieber, M.R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nature Reviews Molecular Cell Biology *18*, 495–506.

Chao, H.X., Poovey, C.E., Privette, A.A., Grant, G.D., Chao, H.Y., Cook, J.G., and Purvis, J.E. (2017). Orchestration of DNA Damage Checkpoint Dynamics across the Human Cell Cycle. Cell Syst *5*, 445-459.e5. https://doi.org/10.1016/j.cels.2017.09.015.

Chapman, J.R., and Jackson, S.P. (2008). Phospho-dependent interactions between NBS1 and MDC1 mediate chromatin retention of the MRN complex at sites of DNA damage. EMBO Reports *9*, 795–801.

Chapman, J.R., Barral, P., Vannier, J.-B., Borel, V., Steger, M., Tomas-Loba, A., Sartori, A.A., Adams, I.R., Batista, F.D., and Boulton, S.J. (2013). RIF1 Is Essential for 53BP1-Dependent Nonhomologous End Joining and Suppression of DNA Double-Strand Break Resection. Molecular Cell *49*, 858–871.

Chapman, J.R., Taylor, M.R.G., and Boulton, S.J. (2012). Playing the end game: DNA double-strand break repair pathway choice. Molecular Cell *47*, 497–510.

Chatterjee, N., and Walker, G.C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. Environmental and Molecular Mutagenesis *58*, 235–263.

Chatzou, M., Magis, C., Chang, J.-M., Kemena, C., Bussotti, G., Erb, I., and Notredame, C. (2016). Multiple sequence alignment modeling: methods and applications. Briefings in Bioinformatics *17*, 1009–1023.

Chen, H., You, M.J., Jiang, Y., Wang, W., and Li, L. (2011). RMI1 attenuates tumor development and is essential for early embryonic survival. Molecular Carcinogenesis *50*, 80–88.

Chen, J., Le, S., Basu, A., Chazin, W.J., and Yan, J. (2015). Mechanochemical regulations of RPA's binding to ssDNA. Sci Rep-Uk *5*, 9296. https://doi.org/10.1038/srep09296.

Chester, N., Kuo, F., Kozak, C., O'Hara, C.D., and Leder, P. (1998). Stage-specific apoptosis, developmental delay, and embryonic lethality in mice homozygous for a targeted disruption in the murine Bloom's syndrome gene. Genes \& Development *12*, 3382–3393.

Chiang, T.-W.W. (2016). CRISPR-Cas9D10A nickase-based genotypic and phenotypic screening to enhance genome editing. Nature Publishing Group 1–17.

Cho, Y.S., Challa, S., Moquin, D., Genga, R., Ray, T.D., Guildford, M., and Chan, F.K.-M. (2009). Phosphorylation-driven assembly of the RIP1-RIP3 complex regulates programmed necrosis and virus-induced inflammation. Cell *137*, 1112–1123. .

Chu, W.K., and Hickson, I.D. (2009). RecQ helicases: multifunctional genome caretakers. Nature Reviews Cancer *9*, 644–654.

Chun, H.H., and Gatti, R.A. (2004). Ataxia–telangiectasia, an evolving phenotype. Dna Repair *3*, 1187–1196. https://doi.org/10.1016/j.dnarep.2004.04.010.

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. Molecular Cell *40*, 179–204.

Ciccia, A., Bredemeyer, A.L., Sowa, M.E., Terret, M.-E., Jallepalli, P.V., Harper, J.W., and Elledge, S.J. (2009). The SIOD disorder protein SMARCAL1 is an RPA-interacting protein involved in replication fork restart. Genes \& Development *23*, 2415–2425.

Ciccia, A., Nimonkar, A.V., Hu, Y., Hajdu, I., Achar, Y.J., Izhar, L., Petit, S.A., Adamson, B., Yoon, J.C., Kowalczykowski, S.C., et al. (2012). Polyubiquitinated PCNA recruits the ZRANB3 translocase to maintain genomic integrity after replication stress. Molecular Cell *47*, 396–409.

Cimprich, K.A., and Cortez, D. (2008). ATR: an essential regulator of genome integrity. Nature Reviews Molecular Cell Biology *9*, 616–627.

Cimprich, K.A., Shin, T.B., Keith, C.T., and Schreiber, S.L. (1996). cDNA cloning and gene mapping of a candidate human cell cycle checkpoint protein. Proceedings of the National Academy of Sciences *93*, 2850–2855. .

Cleary, S.P., Zhang, W., Nicola, N.D., Aronson, M., Aube, J., Steinman, A., Haddad, R., Redston, M., Gallinger, S., Narod, S.A., et al. (2003). Heterozygosity for the BLM(Ash) mutation and cancer risk. Cancer Res *63*, 1769–1771. .

CLEAVER, J.E. (1968). Defective Repair Replication of DNA in Xeroderma Pigmentosum. Nature *218*, 652–656. https://doi.org/10.1038/218652a0.

Clements, P.M., Breslin, C., Deeks, E.D., Byrd, P.J., Ju, L., Bieganowski, P., Brenner, C., Moreira, M.-C., Taylor, A.M.R., and Caldecott, K.W. (2004). The ataxia–oculomotor apraxia 1 gene product has a role distinct from ATM and interacts with the DNA strand break repair proteins XRCC1 and XRCC4. Dna Repair *3*, 1493–1502. https://doi.org/10.1016/j.dnarep.2004.06.017.

Cobb, J.A., Bjergbaek, L., Shimada, K., Frei, C., and Gasser, S.M. (2003). DNA polymerase stabilization at stalled replication forks requires Mec1 and the RecQ helicase Sgs1. Embo J *22*, 4325–4336. https://doi.org/10.1093/emboj/cdg391.

Cobb, J.A., Schleker, T., Rojas, V., Bjergbaek, L., Tercero, J.A., and Gasser, S.M. (2005). Replisome instability, fork collapse, and gross chromosomal rearrangements arise synergistically from Mec1 kinase and RecQ helicase mutations. Gene Dev *19*, 3055–3069. https://doi.org/10.1101/gad.361805.

Cole, S.P., Bhardwaj, G., Gerlach, J.H., Mackie, J.E., Grant, C.E., Almquist, K.C., Stewart, A.J., Kurz, E.U., Duncan, A.M., and Deeley, R.G. (1992). Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. Science *258*, 1650–1654. .

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science *339*, 819–823.

Cortez, D. (2015). Preventing replication fork collapse to maintain genome integrity. DNA Repair *32*, 149–157.

Cortez, D. (2019). Replication-Coupled DNA Repair. Molecular Cell *74*, 866–876.

Costa, A., Ilves, I., Tamberg, N., Petojevic, T., Nogales, E., Botchan, M.R., and Berger, J.M. (2011). The structural basis for MCM2–7 helicase activation by GINS and Cdc45. Nat Struct Mol Biol *18*, 471–477. https://doi.org/10.1038/nsmb.2004.

Costantino, L., Sotiriou, S.K., Rantala, J.K., Magin, S., Mladenov, E., Helleday, T., Haber, J.E., Iliakis, G., Kallioniemi, O.P., and Halazonetis, T.D. (2014). Break-induced replication repair of damaged forks induces genomic duplications in human cells. Science *343*, 88–91.

Couch, F.B., Bansbach, C.E., Driscoll, R., Luzwick, J.W., Glick, G.G., Bétous, R., Carroll, C.M., Jung, S.Y., Qin, J., Cimprich, K.A., et al. (2013). ATR phosphorylates SMARCAL1 to prevent replication fork collapse. Gene Dev *27*, 1610–1623. https://doi.org/10.1101/gad.214080.113.

Countryman, P.I., and Heddle, J.A. (1976). The production of micronuclei from chromosome aberrations in irradiated cultures of human lymphocytes. Mutat Res Fundam Mol Mech Mutagen *41*, 321–331. https://doi.org/10.1016/0027-5107(76)90105-6.

Crickard, J.B., Xue, C., Wang, W., Kwon, Y., Sung, P., and Greene, E.C. (2019). The RecQ helicase Sgs1 drives ATP-dependent disruption of Rad51 filaments. Nucleic Acids Research *47*, 4694–4706.

D'Angiolella, V., Donato, V., Forrester, F.M., Jeong, Y.-T., Pellacani, C., Kudo, Y., Saraf, A., Florens, L., Washburn, M.P., and Pagano, M. (2012). Cyclin F-mediated degradation of ribonucleotide reductase M2 controls genome integrity and DNA repair. Cell *149*, 1023–1034.

Dahai, Y., Sanyuan, S., Hong, L., Di, Z., and Chong, Z. (2013). A Relationship Between Replication Protein A and Occurrence and Prognosis of Esophageal Carcinoma. Cell Biochem Biophys *67*, 175–180. https://doi.org/10.1007/s12013-013-9530-y.

Dai, Y., Kysela, B., Hanakahi, L.A., Manolis, K., Riballo, E., Stumm, M., Harville, T.O., West, S.C., Oettinger, M.A., and Jeggo, P.A. (2003). Nonhomologous end joining and V(D)J recombination require an additional factor. Proc National Acad Sci *100*, 2462–2467. https://doi.org/10.1073/pnas.0437964100.

Dalgaard, J.Z. (2012). Causes and consequences of ribonucleotide incorporation into nuclear DNA. Trends in Genetics : TIG *28*, 592–597.

Datta, A., and Jinks-Robertson, S. (1995). Association of increased spontaneous mutation rates with high levels of transcription in yeast. Science *268*, 1616–1619. .

Davey, M.J., Indiani, C., and O'Donnell, M. (2003). Reconstitution of the Mcm2-7p Heterohexamer, Subunit Arrangement, and ATP Site Architecture*. J Biol Chem *278*, 4491–4499. https://doi.org/10.1074/jbc.m210511200.

Davey, N.E., Cowan, J.L., Shields, D.C., Gibson, T.J., Coldwell, M.J., and Edwards, R.J. (2012a). SLiMPrints: conservation-based discovery of functional motif fingerprints in intrinsically disordered protein regions. Nucleic Acids Research *40*, 10628–10641.

Davey, N.E., Trave, G., and Gibson, T.J. (2011). How viruses hijack cell regulation. Trends in Biochemical Sciences *36*, 159–169.

Davey, N.E., Van Roey, K., Weatheritt, R.J., Toedt, G., Uyar, B., Altenberg, B., Budd, A., Diella, F., Dinkel, H., and Gibson, T.J. (2012b). Attributes of short linear motifs. Mol Biosyst *8*, 268–281.

Davies, S.L., North, P.S., and Hickson, I.D. (2007). Role for BLM in replication-fork restart and suppression of origin firing after replicative stress. Nature Structural \& Molecular Biology *14*, 677–679.

Davies, S.L., North, P.S., Dart, A., Lakin, N.D., and Hickson, I.D. (2004). Phosphorylation of the Bloom's syndrome helicase and its role in recovery from S-phase arrest. Molecular and Cellular Biology *24*, 1279–1291.

Davis, M.M., Calame, K., Early, P.W., Livant, D.L., Joho, R., Weissman, I.L., and Hood, L. (1980). An immunoglobulin heavy-chain gene is formed by at least two recombinational events. Nature *283*, 733–739. https://doi.org/10.1038/283733a0.

Day, M., Rappas, M., Ptasinska, K., Boos, D., Oliver, A.W., and Pearl, L.H. (2018). BRCT domains of the DNA damage checkpoint proteins TOPBP1/Rad4 display distinct specificities for phosphopeptide ligands. eLife *7*.

Deegan, T.D., and Diffley, J.F.X. (2016). MCM: one ring to rule them all. Curr Opin Struct Biol *37*, 145–151. https://doi.org/10.1016/j.sbi.2016.01.014.

Deegan, T.D., and Diffley, J.F.X. (2016). MCM: one ring to rule them all. Current Opinion in Structural Biology *37*, 145–151.

Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., Chabes, A., and Malkova, A. (2011). Break-induced replication is highly inaccurate. PLoS Biol *9*, e1000594.

Delacroix, S., Wagner, J.M., Kobayashi, M., Yamamoto, K.-I., and Karnitz, L.M. (2007). The Rad9-Hus1-Rad1 (9-1-1) clamp activates checkpoint signaling via TopBP1. Genes & Development *21*, 1472–1477. https://doi.org/10.1101/gad.1547007.

Deng, L., Wu, R.A., Sonneville, R., Kochenova, O.V., Labib, K., Pellman, D., and Walter, J.C. (2019). Mitotic CDK Promotes Replisome Disassembly, Fork Breakage, and Complex DNA Rearrangements. Molecular Cell *73*, 915–929.e916.

Denko, N.C., Giaccia, A.J., Stringer, J.R., and Stambrook, P.J. (1994). The human Ha-ras oncogene induces genomic instability in murine fibroblasts within one cell cycle. Proc National Acad Sci *91*, 5124–5128. https://doi.org/10.1073/pnas.91.11.5124.

Dephoure, N., Zhou, C., Villén, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., and Gygi, S.P. (2008). A quantitative atlas of mitotic phosphorylation. Proceedings of the National Academy of Sciences of the United States of America *105*, 10762–10767.

Deshpande, R.A., Lee, J.-H., Arora, S., and Paull, T.T. (2016). Nbs1 Converts the Human Mre11/Rad50 Nuclease Complex into an Endo/Exonuclease Machine Specific for Protein-DNA Adducts. Mol Cell *64*, 593–606. https://doi.org/10.1016/j.molcel.2016.10.010.

Dev, H., Chiang, T.-W.W., Lescale, C., Krijger, I., Martin, A.G., Pilger, D., Coates, J., Sczaniecka-Clift, M., Wei, W., Ostermaier, M., et al. (2018). Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. Nature Cell Biology 1–19.

Dewar, J.M., and Walter, J.C. (2017). Mechanisms of DNA replication termination. Nature Reviews Molecular Cell Biology *18*, 507–516.

Dewar, J.M., Low, E., Mann, M., Räschle, M., and Walter, J.C. (2017). CRL2Lrr1 promotes unloading of the vertebrate replisome from chromatin during replication termination. Genes \& Development *31*, 275–290.

Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N.P., Trave, G., and Gibson, T.J. (2008). Understanding eukaryotic linear motifs and their role in cell signaling and regulation. Front Biosci *13*, 6580–6603.

Ding, Z., and Kihara, D. (2019). Computational identification of protein-protein interactions in model plant proteomes. Scientific Reports *9*, 8740–8713.

Doherty, K.M., Sommers, J.A., Gray, M.D., Lee, J.W., Kobbe, C. von, Thoma, N.H., Kureekattil, R.P., Kenny, M.K., and Brosh, R.M. (2005). Physical and functional mapping of the replication protein a interaction domain of the werner and bloom syndrome helicases. J. Biol. Chem. *280*, 29494–29505.

Doil, C., Mailand, N., Bekker-Jensen, S., Menard, P., Larsen, D.H., Pepperkok, R., Ellenberg, J., Panier, S., Durocher, D., Bartek, J., et al. (2009). RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. Cell *136*, 435–446.

Donnianni, R.A., and Symington, L.S. (2013). Break-induced replication occurs by conservative DNA synthesis. Proceedings of the National Academy of Sciences of the United States of America *110*, 13475–13480.

Doyle, L.A., Yang, W., Abruzzo, L.V., Krogmann, T., Gao, Y., Rishi, A.K., and Ross, D.D. (1998). A multidrug resistance transporter from human MCF-7 breast cancer cells. Proceedings of the National Academy of Sciences *95*, 15665–15670. .

Draetta, G., Piwnica-Worms, H., Morrison, D., Druker, B., Roberts, T., and Beach, D. (1988). Human cdc2 protein kinase is a major cell-cycle regulated tyrosine kinase substrate. Nature *336*, 738–744. .

Dray, E., Etchin, J., Wiese, C., Saro, D., Williams, G.J., Hammel, M., Yu, X., Galkin, V.E., Liu, D., Tsai, M.-S., et al. (2010). Enhancement of RAD51 recombinase activity by the tumor suppressor PALB2. Nature Structural \& Molecular Biology *17*, 1255–1259.

Dulić, V., Lees, E., and Reed, S.I. (1992). Association of human cyclin E with a periodic G1-S phase protein kinase. Science *257*, 1958–1961. .

Dunphy, W.G., and Kumagai, A. (1991). The cdc25 protein contains an intrinsic phosphatase activity. Cell *67*, 189–196. https://doi.org/10.1016/0092-8674(91)90582-j.

Easton, D.F., Ford, D., and Bishop, D.T. (1995). Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. Am J Hum Genet *56*, 265–271. .

Eddins, M.J., Carlile, C.M., Gomez, K.M., Pickart, C.M., and Wolberger, C. (2006). Mms2-Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. Nature Structural \& Molecular Biology *13*, 915–920.

El-Deiry, W.S., Tokino, T., Velculescu, V.E., Levy, D.B., Parsons, R., Trent, J.M., Lin, D., Mercer, W.E., Kinzler, K.W., and Vogelstein, B. (1993). WAF1, a potential mediator of p53 tumor suppression. Cell *75*, 817–825. https://doi.org/10.1016/0092-8674(93)90500-p.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2018). The Pfam protein families database in 2019. Nucleic Acids Research *28*, 405–406.

El-Khamisy, S.F., Masutani, M., Suzuki, H., and Caldecott, K.W. (2003). A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage. Nucleic Acids Research *31*, 5526–5533. .

Eladad, S., Ye, T.-Z., Hu, P., Leversha, M., Beresten, S., Matunis, M.J., and Ellis, N.A. (2005). Intra-nuclear trafficking of the BLM helicase to DNA damage-induced foci is regulated by SUMO modification. Human Molecular Genetics *14*, 1351–1365.

Elango, R., Osia, B., Harcy, V., Malc, E., Mieczkowski, P.A., Roberts, S.A., and Malkova, A. (2019). Repair of base damage within break-induced replication intermediates promotes kataegis associated with chromosome rearrangements. Nucleic Acids Research *47*, 9666–9684.

Ellis, N.A., Groden, J., Ye, T.Z., Straughen, J., Lennon, D.J., Ciocci, S., Proytcheva, M., and German, J. (1995). The Bloom's syndrome gene product is homologous to RecQ helicases. Cell *83*, 655–666.

Ellison, M.J., and Childs, J.D. (1981). Pyrimidine dimers induced in Escherichia coli DNA by ultraviolet radiation present in sunlight. Photochemistry and Photobiology *34*, 465–469. .

Ellison, V., and Stillman, B. (2003). Biochemical Characterization of DNA Damage Checkpoint Complexes: Clamp Loader and Clamp Complexes with Specificity for 5′ Recessed DNA. Plos Biol *1*, e33. https://doi.org/10.1371/journal.pbio.0000033.

Epum, E.A., and Haber, J.E. (2021). DNA replication: the recombination connection. Trends Cell Biol *32*, 45–57. https://doi.org/10.1016/j.tcb.2021.07.005.

Epum, E.A., and Haber, J.E. (2021). DNA replication: the recombination connection. Trends in Cell Biology.

Erdős, G., Pajkos, M., and Dosztányi, Z. (2021). IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. Nucleic Acids Research *49*, W297–W303.

Escribano-Díaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J.T.F., Tkáč, J., Cook, M.A., Rosebrock, A.P., Munro, M., Canny, M.D., et al. (2013). A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. Molecular Cell *49*, 872–883.

Evans, T., Rosenthal, E.T., Youngblom, J., Distel, D., and Hunt, T. (1983). Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. Cell *33*, 389–396. .

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., and Speck, C. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc National Acad Sci *106*, 20240–20245. https://doi.org/10.1073/pnas.0911500106.

Fackenthal, J.D., and Olopade, O.I. (2007). Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. Nature Publishing Group *7*, 937–948.

Falck, J., Coates, J., and Jackson, S.P. (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. Nature *434*, 605–611.

Falck, J., Mailand, N., Syljuåsen, R.G., Bartek, J., and Lukas, J. (2001). The ATM–Chk2–Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. Nature *410*, 842–847. https://doi.org/10.1038/35071124.

Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N.J., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., et al. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature *434*, 917–921. .

Fekairi, S., Scaglione, S., Chahwan, C., Taylor, E.R., Tissier, A., Coulon, S., Dong, M.-Q., Ruse, C., Yates, J.R., Russell, P., et al. (2009). Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. Cell *138*, 78–89.

Feng, L., and Chen, J. (2012). The E3 ligase RNF8 regulates KU80 removal and NHEJ repair. Nature Structural \& Molecular Biology *19*, 201–206. .

Feng, S., Zhao, Y., Xu, Y., Ning, S., Huo, W., Hou, M., Gao, G., Ji, J., Guo, R., and Xu, D. (2016). Ewing Tumor-associated Antigen 1 Interacts with Replication Protein A to Promote Restart of Stalled Replication Forks. Journal of Biological Chemistry *291*, 21956–21962.

Feng, W., and Jasin, M. (2017). BRCA2 suppresses replication stress-induced mitotic and G1 abnormalities through homologous recombination. Nature Communications *8*, 525–515.

Ferro, A.M., and Olivera, B.M. (1982). Poly(ADP-ribosylation) in vitro. Reaction parameters and enzyme mechanism. Journal of Biological Chemistry *257*, 7808–7813. .

Fichtinger-Schepman, A.M., Veer, J.L. van der, Hartog, J.H. den, Lohman, P.H., and Reedijk, J. (1985). Adducts of the antitumor drug cis-diamminedichloroplatinum(II) with DNA: formation, identification, and quantitation. Biochemistry *24*, 707–713. .

Findlay, S., Heath, J., Luo, V.M., Malina, A., Morin, T., Coulombe, Y., Djerir, B., Li, Z., Samiei, A., Cheyou, E.S., et al. (2018). SHLD2/FAM35A co-operates with REV7 to coordinate DNA double-strand break repair pathway choice. The EMBO Journal *37*, e100158–20.

Fishel, R., Lescoe, M.K., Rao, M.R.S., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., and Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell *75*, 1027–1038. https://doi.org/10.1016/0092-8674(93)90546-3.

Flach, J., Bakker, S.T., Mohrin, M., Conroy, P.C., Pietras, E.M., Reynaud, D., Alvarez, S., Diolaiti, M.E., Ugarte, F., Forsberg, E.C., et al. (2014). Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. Nature *512*, 198–202. https://doi.org/10.1038/nature13619.

Flanagan, M., and Cunniff, C.M. (2019). Bloom Syndrome. Gene Reviews 1–19.

Forment, J.V., Walker, R.V., and Jackson, S.P. (2012). A high-throughput, flow cytometry-based method to quantify DNA-end resection in mammalian cells. Cytom Part A *81A*, 922–928. https://doi.org/10.1002/cyto.a.22155.

Fradet-Turcotte, A., Canny, M.D., Escribano-Díaz, C., Orthwein, A., Leung, C.C.Y., Huang, H., Landry, M.-C., Kitevski-LeBlanc, J., Noordermeer, S.M., Sicheri, F., et al. (2013). 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. Nature *499*, 50–54.

Freifelder, D. (1965). Mechanism of inactivation of coliphage T7 by x-rays. Proceedings of the National Academy of Sciences *54*, 128–134. .

Frey, J.D., Salibian, A.A., Schnabel, F.R., Choi, M., and Karp, N.S. (2017). Non-BRCA1/2 Breast Cancer Susceptibility Genes: A New Frontier with Clinical Consequences for Plastic Surgeons. Plast Reconstr Surg Glob Open *5*, e1564.

Fujimura, S., Hasegawa, S., Shimizu, Y., and Sugimura, T. (1967). Polymerization of the adenosine 5'-diphosphate-ribose moiety of nicotinamide-adenine dinucleotide by nuclear enzyme. I. Enzymatic reactions. Biochimica et Biophysica Acta *145*, 247–259. .

Fukao, T., Chen, P., Ren, J., Kaneko, H., Zhang, G.X., Kondo, M., Yamamoto, K.-I., Furuichi, Y., Takeda, S., Kondo, N., et al. (2004). Disruption of the BLM gene in ATM-null DT40 cells does not exacerbate either phenotype. Oncogene *23*, 1498–1506.

Galaktionov, K., and Beach, D. (1991). Specific activation of cdc25 tyrosine phosphatases by B-type cyclins: Evidence for multiple roles of mitotic cyclins. Cell *67*, 1181–1194. https://doi.org/10.1016/0092-8674(91)90294-9.

Gambus, A., Khoudoli, G.A., Jones, R.C., and Blow, J.J. (2011). MCM2-7 Form Double Hexamers at Licensed Origins in Xenopus Egg Extract*. J Biological Chem *286*, 11855–11864. https://doi.org/10.1074/jbc.m110.199521.

Gao, S., Feng, S., Ning, S., Liu, J., Zhao, H., Xu, Y., Shang, J., Li, K., Li, Q., Guo, R., et al. (2018). An OB-fold complex controls the repair pathways for DNA double-strand breaks. Nature Communications 1–10.

Garaycoechea, J.I., Crossan, G.P., Langevin, F., Mulderrig, L., Louzada, S., Yang, F., Guilbaud, G., Park, N., Roerink, S., Nik-Zainal, S., et al. (2018). Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. Nature *553*, 171–177.

Garcia, V., Phelps, S.E.L., Gray, S., and Neale, M.J. (2011). Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. Nature *479*, 241–244. https://doi.org/10.1038/nature10515.

Gari, K., Décaillet, C., Delannoy, M., Wu, L., and Constantinou, A. (2008a). Remodeling of DNA replication structures by the branch point translocase FANCM. Proceedings of the National Academy of Sciences of the United States of America *105*, 16107–16112.

Gari, K., Décaillet, C., Stasiak, A.Z., Stasiak, A., and Constantinou, A. (2008b). The Fanconi anemia protein FANCM can promote branch migration of Holliday junctions and replication forks. Molecular Cell *29*, 141–148.

Garinis, G.A. (2008). Nucleotide excision repair deficiencies and the somatotropic axis in aging. Hormones (Athens, Greece) *7*, 9–16.

Gennery, A.R. (2006). Primary immunodeficiency syndromes associated with defective DNA double-strand break repair. British Medical Bulletin *77-78*, 71–85.

Genschel, J., Bazemore, L.R., and Modrich, P. (2002). Human Exonuclease I Is Required for 5′ and 3′ Mismatch Repair∗. J Biol Chem *277*, 13302–13311. https://doi.org/10.1074/jbc.m111854200.

Gérard, C., and Goldbeter, A. (2009). Temporal self-organization of the cyclin/Cdk network driving the mammalian cell cycle. Proceedings of the National Academy of Sciences of the United States of America *106*, 21643–21648. .

German J. (1993). Bloom syndrome: a mendelian prototype of somatic mutational disease. Medicine, *72*(6), 393–406.

Ghezraoui, H., Oliveira, C., Becker, J.R., Bilham, K., Moralli, D., Anzilotti, C., Fischer, R., Deobagkar-Lele, M., Sanchiz-Calvo, M., Fueyo-Marcos, E., et al. (2018). 53BP1 cooperation with the REV7–shieldin complex underpins DNA structure-specific NHEJ. Nature *560*, 1–25.

Gibson, T.J., Dinkel, H., Van Roey, K., and Diella, F. (2015). Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. Cell Commun Signal *13*, 42–15.

Gillespie, P.J., Li, A., and Blow, J.J. (2001). Reconstitution of licensed replication origins on Xenopus sperm nuclei using purified proteins. Bmc Biochem *2*, 15–15. https://doi.org/10.1186/1471-2091-2-15.

Gisselsson, D., Pettersson, L., Höglund, M., Heidenblad, M., Gorunova, L., Wiegant, J., Mertens, F., Cin, P.D., Mitelman, F., and Mandahl, N. (2000). Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. Proc National Acad Sci *97*, 5357–5362. https://doi.org/10.1073/pnas.090013497.

Giunta, S., Belotserkovskaya, R., and Jackson, S.P. (2010). DNA damage signaling in response to double-strand breaks during mitosis. J Cell Biol *190*, 197–207. https://doi.org/10.1083/jcb.200911156.

Giunta, S., Belotserkovskaya, R., and Jackson, S.P. (2010). DNA damage signaling in response to double-strand breaks during mitosis. The Journal of Cell Biology *190*, 197–207.

Givalos, N., Gakiopoulou, H., Skliri, M., Bousboukea, K., Konstantinidou, A.E., Korkolopoulou, P., Lelouda, M., Kouraklis, G., Patsouris, E., and Karatzas, G. (2007). Replication protein A is an independent prognostic indicator with potential therapeutic implications in colon cancer. Modern Pathol *20*, 159–166. https://doi.org/10.1038/modpathol.3800719.

Gong, Z., Kim, J.-E., Leung, C.C.Y., Glover, J.N.M., and Chen, J. (2010). BACH1/FANCJ acts with TopBP1 and participates early in DNA replication checkpoint control. Molecular Cell *37*, 438–446.

Goss, K.H., Risinger, M.A., Kordich, J.J., Sanz, M.M., Straughen, J.E., Slovek, L.E., Capobianco, A.J., German, J., Boivin, G.P., and Groden, J. (2002). Enhanced Tumor Formation in Mice Heterozygous for Blm Mutation. Science *297*, 2051–2053. https://doi.org/10.1126/science.1074340.

Goulian, M., Richards, S.H., Heard, C.J., and Bigsby, B.M. (1990). Discontinuous DNA synthesis by purified mammalian proteins. J Biological Chem *265*, 18461–18471. .

Gradia, S., Subramanian, D., Wilson, T., Acharya, S., Makhov, A., Griffith, J., and Fishel, R. (1999). hMSH2–hMSH6 Forms a Hydrolysis-Independent Sliding Clamp on Mismatched DNA. Mol Cell *3*, 255–261. https://doi.org/10.1016/s1097-2765(00)80316-0.

Graham, T.G.W., Carney, S.M., Walter, J.C., and Loparo, J.J. (2018). A single XLF dimer bridges DNA ends during nonhomologous end joining. Nature Structural \& Molecular Biology 1–14.

Graham, T.G.W., Walter, J.C., and Loparo, J.J. (2016). Two-Stage Synapsis of DNA Ends during Non-homologous End Joining. Molecular Cell *61*, 850–858.

Gravel, S., Chapman, J.R., Magill, C., and Jackson, S.P. (2008). DNA helicases Sgs1 and BLM promote DNA double-strand break resection. Genes \& Development *22*, 2767–2772.

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T.E., Mann, M., and Lieber, M.R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature *388*, 492–495.

Green, B.M., Finn, K.J., and Li, J.J. (2010). Loss of DNA replication control is a potent inducer of gene amplification. Science *329*, 943–946.

Greider, C.W., and Blackburn, E.H. (1987). The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. Cell *51*, 887–898. .

Groden, J., Nakamura, Y., and German, J. (1990). Molecular evidence that homologous recombination occurs in proliferating human somatic cells. Proceedings of the National Academy of Sciences *87*, 4315–4319.

Gueven, N., Becherel, O.J., Kijas, A.W., Chen, P., Howe, O., Rudolph, J.H., Gatti, R., Date, H., Onodera, O., Taucher-Scholz, G., et al. (2004). Aprataxin, a novel protein that protects against genotoxic stress. Hum Mol Genet *13*, 1081–1093. https://doi.org/10.1093/hmg/ddh122.

Guilliam, T.A., Brissett, N.C., Ehlinger, A., Keen, B.A., Kolesar, P., Taylor, E.M., Bailey, L.J., Lindsay, H.D., Chazin, W.J., and Doherty, A.J. (2017). Molecular basis for PrimPol recruitment to replication forks by RPA. Nature Communications *8*, 15222–15214.

Gundry, R.L., White, M.Y., Murray, C.I., Kane, L.A., Fu, Q., Stanley, B.A., and Eyk, J.E.V. (2009). Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. Current Protocols in Molecular Biology *Chapter 10*, Unit10.25–10.25.23.

Guo, Z., Kumagai, A., Wang, S.X., and Dunphy, W.G. (2000). Requirement for Atr in phosphorylation of Chk1 and cell cycle regulation in response to DNA replication blocks

and UV-damaged DNA in Xenopus egg extracts. Gene Dev *14*, 2745–2756. https://doi.org/10.1101/gad.842500.

Gupta, R., Somyajit, K., Narita, T., Maskey, E., Stanlie, A., Kremer, M., Typas, D., Lammers, M., Mailand, N., Nussenzweig, A., et al. (2018). DNA Repair Network Analysis Reveals Shieldin as a Key Regulator of NHEJ and PARP Inhibitor Sensitivity. Cell 1–41.

Gupta, R., Somyajit, K., Narita, T., Maskey, E., Stanlie, A., Kremer, M., Typas, D., Lammers, M., Mailand, N., Nussenzweig, A., et al. (2018). DNA Repair Network Analysis Reveals Shieldin as a Key Regulator of NHEJ and PARP Inhibitor Sensitivity. Cell 1–41. https://doi.org/10.1016/j.cell.2018.03.050.

Haahr, P., Hoffmann, S., Tollenaere, M.A.X., Ho, T., Toledo, L.I., Mann, M., Bekker-Jensen, S., Räschle, M., and Mailand, N. (2016). Activation of the ATR kinase by the RPA-binding protein ETAA1. Nature Cell Biology *18*, 1196–1207.

Halazonetis, T.D., Gorgoulis, V.G., and Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. Science *319*, 1352–1355.

Hambarde, S., Tsai, C.-L., Pandita, R.K., Bacolla, A., Maitra, A., Charaka, V., Hunt, C.R., Kumar, R., Limbo, O., Meur, R.L., et al. (2021). EXO5-DNA structure and BLM interactions direct DNA resection critical for ATR-dependent replication restart. Molecular Cell 1–28.

Hanada, K., Budzowska, M., Davies, S.L., Drunen, E.V., Onizawa, H., Beverloo, H.B., Maas, A., Essers, J., Hickson, I.D., and Kanaar, R. (2007). The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. Nature Structural \& Molecular Biology *14*, 1096–1104.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hanzlikova, H., Kalasova, I., Demin, A.A., Pennicott, L.E., Cihlarova, Z., and Caldecott, K.W. (2018). The Importance of Poly(ADP-Ribose) Polymerase as a Sensor of Unligated Okazaki Fragments during DNA Replication. Mol Cell *71*, 319-331.e3. https://doi.org/10.1016/j.molcel.2018.06.004.

Haring, S.J., Mason, A.C., Binz, S.K., and Wold, M.S. (2008). Cellular functions of human RPA1. Multiple roles of domains in replication, repair, and checkpoints. J. Biol. Chem. *283*, 19095–19111.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458–460. https://doi.org/10.1038/345458a0.

HARMAN, D. (1956). Aging: a theory based on free radical and radiation chemistry. Journal of Gerontology *11*, 298–300. .

Harper, J.W., Adami, G.R., Wei, N., Keyomarsi, K., and Elledge, S.J. (1993). The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. Cell *75*, 805–816. https://doi.org/10.1016/0092-8674(93)90499-g.

Harper, J.W., and Elledge, S.J. (2007). The DNA Damage Response: Ten Years After. Molecular Cell *28*, 739–745.

Harrington, J.J., and Lieber, M.R. (1995). DNA Structural Elements Required for FEN-1 Binding (∗). J Biol Chem *270*, 4503–4508. https://doi.org/10.1074/jbc.270.9.4503.

Hashimoto, Y., Chaudhuri, A.R., Lopes, M., and Costanzo, V. (2010). Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis. Nature Structural \& Molecular Biology *17*, 1305–1311.

Haupt, Y., Maya, R., Kazaz, A., and Oren, M. (1997). Mdm2 promotes the rapid degradation of p53. Nature *387*, 296–299.

Hekmat-Nejad, M., You, Z., Yee, M., Newport, J.W., and Cimprich, K.A. (2000). Xenopus ATR is a replication-dependent chromatin-binding protein required for the DNA replication checkpoint. Curr Biol *10*, 1565–1573. https://doi.org/10.1016/s0960-9822(00)00855-1.

Heller, R.C., Kang, S., Lam, W.M., Chen, S., Chan, C.S., and Bell, S.P. (2011). Eukaryotic Origin-Dependent DNA Replication In Vitro Reveals Sequential Action of DDK and S-CDK Kinases. Cell *146*, 80–91. https://doi.org/10.1016/j.cell.2011.06.012.

Helmrich, A., Ballarino, M., Nudler, E., and Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. Nature Structural \& Molecular Biology *20*, 412–418.

Henle, E.S., and Linn, S. (1997). Formation, prevention, and repair of DNA damage by iron/hydrogen peroxide. Journal of Biological Chemistry *272*, 19095–19098.

Hershko, A., Heller, H., Elias, S., and Ciechanover, A. (1983). Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. Journal of Biological Chemistry *258*, 8206–8214. .

Heyer, W.-D. (2015). Regulation of recombination and genomic maintenance. Cold Spring Harbor Perspectives in Biology *7*, a016501.

Higgs, M.R., Reynolds, J.J., Winczura, A., Blackford, A.N., Borel, V., Miller, E.S., Zlatanou, A., Nieminuszczy, J., Ryan, E.L., Davies, N.J., et al. (2015). BOD1L Is Required to Suppress Deleterious Resection of Stressed Replication Forks. Molecular Cell *59*, 462–477.

Hirao, A., Kong, Y.-Y., Matsuoka, S., Wakeham, A., Ruland, J., Yoshida, H., Liu, D., Elledge, S.J., and Mak, T.W. (2000). DNA Damage-Induced Activation of p53 by the Checkpoint Kinase Chk2. Science *287*, 1824–1827. https://doi.org/10.1126/science.287.5459.1824.

Hodson, C., Low, J.K.K., van Twest, S., Jones, S.E., Swuec, P., Murphy, V., Tsukada, K., Fawkes, M., Bythell-Douglas, R., Davies, A., et al. (2022). Mechanism of Bloom syndrome complex assembly required for double Holliday junction dissolution and genome stability. Proceedings of the National Academy of Sciences of the United States of America *119*.

Hoeijmakers, J.H.J. (2009). DNA damage, aging, and cancer. N Engl J Med *361*, 1475–1485.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature *496*, 498–503.

Hozumi, N., and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc National Acad Sci *73*, 3628–3632. https://doi.org/10.1073/pnas.73.10.3628.

Hsiang, Y.H., Hertzberg, R., Hecht, S., and Liu, L.F. (1985). Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. Journal of Biological Chemistry *260*, 14873–14878. .

Hu, Q., Botuyan, M.V., Zhao, D., Cui, G., Mer, E., and Mer, G. (2021). Mechanisms of BRCA1-BARD1 nucleosome recognition and ubiquitylation. Nature *596*, 438–443.

Hu, Y., Raynard, S., Sehorn, M.G., Lu, X., Bussen, W., Zheng, L., Stark, J.M., Barnes, E.L., Chi, P., Janscak, P., et al. (2007). RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. Genes \& Development *21*, 3073–3084.

Hudson, D.F., Amor, D.J., Boys, A., Butler, K., Williams, L., Zhang, T., and Kalitsis, P. (2016). Loss of RMI2 Increases Genome Instability and Causes a Bloom-Like Syndrome. PLoS Genetics *12*, e1006483.

Huen, M.S.Y., Grant, R., Manke, I., Minn, K., Yu, X., Yaffe, M.B., and Chen, J. (2007). RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. Cell *131*, 901–914.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Molecular Cell *12*, 711–721. .

Huertas, P., and Jackson, S.P. (2009). Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. J. Biol. Chem. *284*, 9558–9565.

Hustedt, N., Álvarez-Quilón, A., McEwan, A., Yuan, J.Y., Cho, T., Koob, L., Hart, T., and Durocher, D. (2019). A consensus set of genetic vulnerabilities to ATR inhibition. Open Biology *9*, 190156.

Hustedt, N., and Durocher, D. (2017). The control of DNA repair by the cell cycle. Nature Cell Biology *19*, 1–9.

Hwang, T., Parker, S.S., Hill, S.M., Grant, R.A., Ilunga, M.W., Sivaraman, V., Mouneimne, G., and Keating, A.E. (2022). Native proline-rich motifs exploit sequence context to target actin-remodeling Ena/VASP protein ENAH. eLife *11*.

Hwang, T., Parker, S.S., Hill, S.M., Ilunga, M.W., Grant, R.A., Mouneimne, G., and Keating, A.E. (2021). A distributed residue network permits conformational binding specificity in a conserved family of actin remodelers. eLife *10*.

IKEGAMI, S., TAGUCHI, T., OHASHI, M., OGURO, M., NAGANO, H., and MANO, Y. (1978). Aphidicolin prevents mitotic cell division by interfering with the activity of DNA polymerase-α. Nature *275*, 458–460. https://doi.org/10.1038/275458a0.

Iles, N., Rulten, S., El-Khamisy, S.F., and Caldecott, K.W. (2007). APLF (C2orf13) Is a Novel Human Protein Involved in the Cellular Response to Chromosomal DNA Strand Breaks. Mol Cell Biol *27*, 3793–3803. https://doi.org/10.1128/mcb.02269-06.

Imamura, O., Fujita, K., Shimamoto, A., Tanabe, H., Takeda, S., Furuichi, Y., and Matsumoto, T. (2001). Bloom helicase is involved in DNA surveillance in early S phase in vertebrate cells. Oncogene *20*, 1143–1151.

Imlay, J.A., and Linn, S. (1988). DNA damage and oxygen radical toxicity. Science *240*, 1302–1309. .

Ip, S.C.Y., Rass, U., Blanco, M.G., Flynn, H.R., Skehel, J.M., and West, S.C. (2008). Identification of Holliday junction resolvases from humans and yeast. Nature *456*, 357–361.

Ishimi, Y., Claude, A., Bullock, P., and Hurwitz, J. (1988). Complete enzymatic synthesis of DNA containing the SV40 origin of replication. J Biological Chem *263*, 19723–19733. .

Ismail, I.H., Gagné, J.-P., Genois, M.-M., Strickfaden, H., McDonald, D., Xu, Z., Poirier, G.G., Masson, J.-Y., and Hendzel, M.J. (2015). The RNF138 E3 ligase displaces Ku to promote DNA end resection and regulate DNA repair pathway choice. Nature Cell Biology *17*, 1446–1457. .

IYER, V.N., and SZYBALSKI, W. (1963). A MOLECULAR MECHANISM OF MITOMYCIN ACTION: LINKING OF COMPLEMENTARY DNA STRANDS. Proceedings of the National Academy of Sciences *50*, 355–362. .

IYER, V.N., and SZYBALSKI, W. (1964). MITOMYCINS AND PORFIROMYCIN: CHEMICAL MECHANISM OF ACTIVATION AND CROSS-LINKING OF DNA. Science *145*, 55–58. .

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071–1078.

Jackson, S.P., MacDonald, J.J., Lees-Miller, S., and Tjian, R. (1990). GC box binding induces phosphorylation of Sp1 by a DNA-dependent protein kinase. Cell *63*, 155–165. .

Jasin, M., and Rothstein, R. (2013). Repair of Strand Breaks by Homologous Recombination. Cold Spring Harbor Perspectives in Biology *5*, a012740–a012740.

Jazayeri, A., Balestrini, A., Garner, E., Haber, J.E., and Costanzo, V. (2008). Mre11–Rad50–Nbs1-dependent processing of DNA breaks generates oligonucleotides that stimulate ATM activity. Embo J *27*, 1953–1962. https://doi.org/10.1038/emboj.2008.128.

Jazayeri, A., Falck, J., Lukas, C., Bartek, J., Smith, G.C.M., Lukas, J., and Jackson, S.P. (2006). ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. Nature Cell Biology *8*, 37–45. .

Jensen, R.B., Carreira, A., and Kowalczykowski, S.C. (2010). Purified human BRCA2 stimulates RAD51-mediated recombination. Nature *467*, 678–683.

Jiang, X., Jimenez, G., Chang, E., Frollkis, M., Kusler, B., Sage, M.,

Jones, J.F., Ritenbaugh, C.K., Spence, M.A., and Hayward, A. (1991). Severe combined immunodeficiency among the Navajo. I. Characterization of phenotypes, epidemiology, and population genetics. Hum Biol *63*, 669–682. .

Jones, R.M., and Petermann, E. (2012). Replication fork dynamics and the DNA damage response. Biochem J *443*, 13–26.

Jowsey, P., Morrice, N.A., Hastie, C.J., McLauchlan, H., Toth, R., and Rouse, J. (2007). Characterisation of the sites of DNA damage-induced 53BP1 phosphorylation catalysed by ATM and ATR. DNA Repair *6*, 1536–1544.

Jr., R.M.B., Li, J.-L., Kenny, M.K., Karow, J.K., Cooper, M.P., Kureekattil, R.P., Hickson, I.D., and Bohr, V.A. (2000). Replication Protein A Physically Interacts with the Bloom's Syndrome Protein and Stimulates Its Helicase Activity. J. Biol. Chem. *275*, 23500–23508.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.

Jungmichel, S., Clapperton, J.A., Lloyd, J., Hari, F.J., Spycher, C., Pavic, L., Li, J., Haire, L.F., Bonalli, M., Larsen, D.H., et al. (2012). The molecular basis of ATM-dependent dimerization of the Mdc1 DNA damage checkpoint mediator. Nucleic Acids Research *40*, 3913–3928.

Junttila, M.R., Puustinen, P., Niemelä, M., Ahola, R., Arnold, H., Böttzauw, T., Ala-aho, R., Nielsen, C., Ivaska, J., Taya, Y., et al. (2007). CIP2A inhibits PP2A in human malignancies. Cell *130*, 51–62.

Kadyrov, F.A., Dzantiev, L., Constantin, N., and Modrich, P. (2006). Endonucleolytic Function of MutLα in Human Mismatch Repair. Cell *126*, 297–308. https://doi.org/10.1016/j.cell.2006.05.039.

Kanaar, R., and Wyman, C. (2008). DNA repair by the MRN complex: break it to make it. Cell *135*, 14–16.

Kang, D., Lee, S., Ryu, K.-S., Cheong, H.-K., Kim, E.-H., and Park, C.-J. (2018). Interaction of replication protein A with two acidic peptides from human Bloom syndrome protein. FEBS Letters *592*, 547–558.

Kaplan, H.S. (1966). DNA-strand scission and loss of viability after x irradiation of normal and sensitized bacterial cells. Proceedings of the National Academy of Sciences *55*, 1442–1446. .

Karlseder, J., Broccoli, D., Dai, Y., Hardy, S., and Lange, T. de (1999). p53- and ATM-Dependent Apoptosis Induced by Telomeres Lacking TRF2. Science *283*, 1321–1325. https://doi.org/10.1126/science.283.5406.1321.

Karmakar, P., Seki, M., Kanamori, M., Hashiguchi, K., Ohtsuki, M., Murata, E., Inoue, E., Tada, S., Lan, L., Yasui, A., et al. (2006). BLM is an early responder to DNA double-strand breaks. Biochem Biophys Res Commun *348*, 62–69.

Kastan, M.B., and Bartek, J. (2004). Cell-cycle checkpoints and cancer. Nature *432*, 316–323. https://doi.org/10.1038/nature03097.

Kato, H., and Sandberg, A.A. (1968). Chromosome pulverization in human cells with micronuclei. J Natl Cancer I *40*, 165–179. .

Katyal, S., and McKinnon, P.J. (2008). DNA strand breaks, neurodegeneration and aging in the brain. Mechanisms of Ageing and Development *129*, 483–491.

Kelland, L. (2007). The resurgence of platinum-based cancer chemotherapy. Nature Reviews Cancer *7*, 573–584.

Kelliher, M.A., Grimm, S., Ishida, Y., Kuo, F., Stanger, B.Z., and Leder, P. (1998). The death domain kinase RIP mediates the TNF-induced NF-kappaB signal. Immunity *8*, 297–303. .

Kennedy, J.A., Daughdrill, G.W., and Schmidt, K.H. (2013). A transient α-helical molecular recognition element in the disordered N-terminus of the Sgs1 helicase is critical for chromosome stability and binding of Top3/Rmi1. Nucleic Acids Research *41*, 10215–10227.

Kim, C., Paulus, B.F., and Wold, M.S. (1994). Interactions of human replication protein A with oligonucleotides. Biochemistry *33*, 14197–14206.

Kim, H., and D'Andrea, A.D. (2012). Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. Genes \& Development *26*, 1393–1408.

King, R.W., Peters, J.M., Tugendreich, S., Rolfe, M., Hieter, P., and Kirschner, M.W. (1995). A 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. Cell *81*, 279–288. .

Kitao, S., Shimamoto, A., Goto, M., Miller, R.W., Smithson, W.A., Lindor, N.M., and Furuichi, Y. (1999). Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome. Nature Genetics *22*, 82–84.

Klein Douwel, D., Boonen, R.A.C.M., Long, D.T., Szypowska, A.A., Räschle, M., Walter, J.C., and Knipscheer, P. (2014). XPF-ERCC1 Acts in Unhooking DNA Interstrand Crosslinks in Cooperation with FANCD2 and FANCP/SLX4. Mol Cell *54*, 460–471. https://doi.org/10.1016/j.molcel.2014.03.015.

Kliszczak, M., Sedlackova, H., Pitchai, G.P., Streicher, W.W., Krejci, L., and Hickson, I.D. (2015). Interaction of RECQ4 and MCM10 is important for efficient DNA replication origin firing in human cells. Oncotarget *6*, 40464–40479.

Koch, C.A., Agyei, R., Galicia, S., Metalnikov, P., O'Donnell, P., Starostine, A., Weinfeld, M., and Durocher, D. (2004). Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. Embo J *23*, 3874–3885. https://doi.org/10.1038/sj.emboj.7600375.

Koff, A., Giordano, A., Desai, D., Yamashita, K., Harper, J.W., Elledge, S., Nishimoto, T., Morgan, D.O., Franza, B.R., and Roberts, J.M. (1992). Formation and activation of a cyclin E-cdk2 complex during the G1 phase of the human cell cycle. Science *257*, 1689–1694. .

Kolas, N.K., Chapman, J.R., Nakada, S., Ylanko, J., Chahwan, R., Sweeney, F.D., Panier, S., Mendez, M., Wildenhain, J., Thomson, T.M., et al. (2007). Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. Science *318*, 1637–1640.

Koning, A.P.J. de, Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. *7*, e1002384.

Kowalczykowski, S.C. (2015). An Overview of the Molecular Mechanisms of Recombinational DNA Repair. Cold Spring Harbor Perspectives in Biology *7*, a016410.

Krokan, H.E., and Bjørås, M. (2013). Base Excision Repair. Csh Perspect Biol *5*, a012583. https://doi.org/10.1101/cshperspect.a012583.

Kulyyassov, A., Fresnais, M., and Longuespée, R. (2021). Targeted liquid chromatography-tandem mass spectrometry analysis of proteins: Basic principles, applications, and perspectives. Proteomics *21*, e2100153.

Kumagai, A., and Dunphy, W.G. (2000). Claspin, a Novel Protein Required for the Activation of Chk1 during a DNA Replication Checkpoint Response in Xenopus Egg Extracts. Mol Cell *6*, 839–849. https://doi.org/10.1016/s1097-2765(05)00092-4.

Kumagai, A., Lee, J., Yoo, H.Y., and Dunphy, W.G. (2006). TopBP1 activates the ATR-ATRIP complex. Cell *124*, 943–955.

Kumagai, A., Shevchenko, A., Shevchenko, A., and Dunphy, W.G. (2010). Treslin collaborates with TopBP1 in triggering the initiation of DNA replication. Cell *140*, 349–359.

Kumar, M., Michael, S., Alvarado-Valverde, J., Mészáros, B., Sámano-Sánchez, H., Zeke, A., Dobson, L., Lazar, T., Örd, M., Nagpal, A., et al. (2021). The Eukaryotic Linear Motif resource: 2022 release. Nucleic Acids Research.

Kunkel, T.A., and Erie, D.A. (2015). Eukaryotic Mismatch Repair in Relation to DNA Replication. Annu Rev Genet *49*, 1–23. https://doi.org/10.1146/annurev-genet-112414-054722.

Kunkel, T.A., Sabatino, R.D., and Bambara, R.A. (1987). Exonucleolytic proofreading by calf thymus DNA polymerase delta. Proc National Acad Sci *84*, 4865–4869. https://doi.org/10.1073/pnas.84.14.4865.

Kurup, G. (2010). CyberKnife: A new paradigm in radiotherapy. Journal of Medical Physics *35*, 63–64.

Lähdesmäki, A., Taylor, A.M.R., Chrzanowska, K.H., and Pan-Hammarström, Q. (2004). Delineation of the Role of the Mre11 Complex in Class Switch Recombination*. J Biol Chem *279*, 16479–16487. https://doi.org/10.1074/jbc.m312796200.

Laine, A., Nagelli, S.G., Farrington, C., Butt, U., Cvrljevic, A.N., Vainonen, J.P., Feringa, F.M., Grönroos, T.J., Gautam, P., Khan, S., et al. (2021a). CIP2A Interacts with TopBP1 and Drives Basal-Like Breast Cancer Tumorigenesis. Cancer Res *81*, 4319–4331. https://doi.org/10.1158/0008-5472.can-20-3651.

Laine, A., Nagelli, S.G., Farrington, C., Butt, U., Cvrljevic, A.N., Vainonen, J.P., Feringa, F.M., Grönroos, T.J., Gautam, P., Khan, S., et al. (2021b). CIP2A Interacts with TopBP1 and Drives Basal-Like Breast Cancer Tumorigenesis. Cancer Research *81*, 4319–4331. .

Langerak, P., Mejia-Ramirez, E., Limbo, O., and Russell, P. (2011). Release of Ku and MRN from DNA ends by Mre11 nuclease activity and Ctp1 is required for homologous recombination repair of double-strand breaks. PLoS Genetics *7*, e1002271.

Lasko, D., Cavenee, W., and Nordenskjold, M. (1991). Loss of Constitutional Heterozygosity in Human Cancer. Annu Rev Genet *25*, 281–314. https://doi.org/10.1146/annurev.ge.25.120191.001433.

Latre, L., Tusell, L., Martin, M., Miró, R., Egozcue, J., Blasco, M.A., and Genescà, A. (2003). Shortened telomeres join to DNA breaks interfering with their correct repair. Exp Cell Res *287*, 282–288. https://doi.org/10.1016/s0014-4827(03)00134-4.

Le, S., Moore, J.K., Haber, J.E., and Greider, C.W. (1999). RAD50 and RAD51 define two pathways that collaborate to maintain telomeres in the absence of telomerase. Genetics *152*, 143–152.

Le, S., Moore, J.K., Haber, J.E., and Greider, C.W. (1999). RAD50 and RAD51 Define Two Pathways That Collaborate to Maintain Telomeres in the Absence of Telomerase. Genetics *152*, 143–152. https://doi.org/10.1093/genetics/152.1.143.

Lee, D.-H., Acharya, S.S., Kwon, M., Drane, P., Guan, Y., Adelmant, G., Kalev, P., Shah, J., Pellman, D., Marto, J.A., et al. (2014). Dephosphorylation enables the recruitment of 53BP1 to double-strand DNA breaks. Molecular Cell *54*, 512–525.

Lee, D.-H., Acharya, S.S., Kwon, M., Drane, P., Guan, Y., Adelmant, G., Kalev, P., Shah, J., Pellman, D., Marto, J.A., et al. (2014). Dephosphorylation enables the recruitment of 53BP1 to double-strand DNA breaks. Molecular Cell *54*, 512–525. https://doi.org/10.1016/j.molcel.2014.03.020.

Lee, J.-H., and Paull, T.T. (2004). Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. Science *304*, 93–96.

Lee, J., and Dunphy, W.G. (2010). Rad17 Plays a Central Role in Establishment of the Interaction between TopBP1 and the Rad9-Hus1-Rad1 Complex at Stalled Replication Forks. Mol Biol Cell *21*, 926–935. https://doi.org/10.1091/mbc.e09-11-0958.

Lee, J., Kumagai, A., and Dunphy, W.G. (2007). The Rad9-Hus1-Rad1 checkpoint clamp regulates interaction of TopBP1 with ATR. J. Biol. Chem. *282*, 28036–28044. https://doi.org/10.1074/jbc.m704635200.

Lee, M.G., and Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. Nature *327*, 31–35.

Lee, M.G., and Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. Nature *327*, 31–35. .

Lee, S.H., Pan, Z.Q., Kwong, A.D., Burgers, P.M., and Hurwitz, J. (1991). Synthesis of DNA by DNA polymerase epsilon in vitro. J Biological Chem *266*, 22707–22717. .

Lees-Miller, S.P., Chen, Y.R., and Anderson, C.W. (1990). Human cells contain a DNA-activated protein kinase that phosphorylates simian virus 40 T antigen, mouse p53, and the human Ku autoantigen. Molecular and Cellular Biology *10*, 6472–6481.

Lees-Miller, S.P., Chen, Y.R., and Anderson, C.W. (1990). Human cells contain a DNA-activated protein kinase that phosphorylates simian virus 40 T antigen, mouse p53, and the human Ku autoantigen. Molecular and Cellular Biology *10*, 6472–6481. .

Leimbacher, P.-A., Jones, S.E., Shorrocks, A.-M.K., Zompit, M. de M., Day, M., Blaauwendraad, J., Bundschuh, D., Bonham, S., Fischer, R., Fink, D., et al. (2019). MDC1 Interacts with TOPBP1 to Maintain Chromosomal Stability during Mitosis. Molecular Cell *74*, 571–583.e578.

Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B.-K., and Gao, N. (2015). Structure of the eukaryotic MCM complex at 3.8 Å. Nature *524*, 186–191. https://doi.org/10.1038/nature14685.

Li, S., Wang, H., Jehi, S., Li, J., Liu, S., Wang, Z., Truong, L., Chiba, T., Wang, Z., and Wu, X. (2021). PIF1 helicase promotes break-induced replication in mammalian cells. The EMBO Journal *40*, e104509.

Li, W., and Wang, J.C. (1998). Mammalian DNA topoisomerase IIIalpha is essential in early embryogenesis. Proceedings of the National Academy of Sciences *95*, 1010–1013.

Li, Z., Otevrel, T., Gao, Y., Cheng, H.L., Seed, B., Stamato, T.D., Taccioli, G.E., and Alt, F.W. (1995). The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. Cell *83*, 1079–1089.

Liang, L., Deng, L., Nguyen, S.C., Zhao, X., Maulion, C.D., Shao, C., and Tischfield, J.A. (2008). Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. Nucleic Acids Res. *36*, 3297–3310.

Liberi, G., Maffioletti, G., Lucca, C., Chiolo, I., Baryshnikova, A., Cotta-Ramusino, C., Lopes, M., Pellicioli, A., Haber, J.E., and Foiani, M. (2005). Rad51-dependent DNA structures accumulate at damaged replication forks in sgs1 mutants defective in the yeast ortholog of BLM RecQ helicase. Genes \& Development *19*, 339–350.

Lieber, M.R., Ma, Y., Pannicke, U., and Schwarz, K. (2004). The mechanism of vertebrate nonhomologous DNA end joining and its role in V(D)J recombination. Dna Repair *3*, 817–826. https://doi.org/10.1016/j.dnarep.2004.03.015.

Lim, W.A., Richards, F.M., and Fox, R.O. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. Nature *372*, 375–379.

Limoli, C.L., Giedzinski, E., Bonner, W.M., and Cleaver, J.E. (2002). UV-induced replication arrest in the xeroderma pigmentosum variant leads to DNA double-strand breaks, gamma -H2AX formation, and Mre11 relocalization. Proceedings of the National Academy of Sciences *99*, 233–238. .

Lindahl, T. (1974). An N-Glycosidase from Escherichia coli That Releases Free Uracil from DNA Containing Deaminated Cytosine Residues. Proc National Acad Sci *71*, 3649–3653. https://doi.org/10.1073/pnas.71.9.3649.

Liu, D., O'Connor, M.S., Qin, J., and Songyang, Z. (2004). Telosome, a Mammalian Telomere-associated Complex Formed by Multiple Telomeric Proteins*. J Biol Chem *279*, 51338–51342. https://doi.org/10.1074/jbc.m409293200.

Liu, E., Lee, A.Y.-L., Chiba, T., Olson, E., Sun, P., and Wu, X. (2007). The ATR-mediated S phase checkpoint prevents rereplication in mammalian cells when licensing control is disrupted. J Cell Biology *179*, 643–657. https://doi.org/10.1083/jcb.200704138.

Liu, J., Doty, T., Gibson, B., and Heyer, W.-D. (2010). Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. Nature Structural \& Molecular Biology *17*, 1260–1262.

Liu, J., Hanne, J., Britton, B.M., Bennett, J., Kim, D., Lee, J.-B., and Fishel, R. (2016). Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. Nature *539*, 583–587. https://doi.org/10.1038/nature20562.

Liu, J., Luo, S., Zhao, H., Liao, J., Li, J., Yang, C., Xu, B., Stern, D.F., Xu, X., and Ye, K. (2012). Structural mechanism of the phosphorylation-dependent dimerization of the MDC1 forkhead-associated domain. Nucleic Acids Research *40*, 3898–3912.

Liu, L., Yan, Z., Osia, B.A., Twarowski, J., Sun, L., Kramara, J., Lee, R.S., Kumar, S., Elango, R., Li, H., et al. (2021). Tracking break-induced replication shows that it stalls at roadblocks. Nature *590*, 655–659.

Liu, Q., Guntuku, S., Cui, X.S., Matsuoka, S., Cortez, D., Tamai, K., Luo, G., Carattini-Rivera, S., DeMayo, F., Bradley, A., et al. (2000). Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. Gene Dev *14*, 1448–1459. https://doi.org/10.1101/gad.14.12.1448.

Liu, T., Wan, L., Wu, Y., Chen, J., and Huang, J. (2011). hSWS1·SWSAP1 Is an Evolutionarily Conserved Complex Required for Efficient Homologous Recombination Repair*. J Biol Chem *286*, 41758–41766. https://doi.org/10.1074/jbc.m111.271080.

Longley, M.J., Pierce, A.J., and Modrich, P. (1997). DNA Polymerase δ Is Required for Human Mismatch Repair in Vitro *. J Biol Chem *272*, 10917–10921. https://doi.org/10.1074/jbc.272.16.10917.

Lopes, M., Cotta-Ramusino, C., Pellicioli, A., Liberi, G., Plevani, P., Muzi-Falconi, M., Newlon, C.S., and Foiani, M. (2001). The DNA replication checkpoint response stabilizes stalled replication forks. Nature *412*, 557–561. https://doi.org/10.1038/35087613.

Lopez-Martinez, D., Liang, C.-C., and Cohn, M.A. (2016). Cellular response to DNA interstrand crosslinks: the Fanconi anemia pathway. Cell Mol Life Sci *73*, 3097–3114. https://doi.org/10.1007/s00018-016-2218-x.

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194–1217.

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194–1217. https://doi.org/10.1016/j.cell.2013.05.039.

Lord, C.J., and Ashworth, A. (2017). PARP inhibitors: Synthetic lethality in the clinic. Science *355*, 1152–1158.

Lou, J., Chen, H., Han, J., He, H., Huen, M.S.Y., Feng, X.-H., Liu, T., and Huang, J. (2017). AUNIP/C1orf135 directs DNA double-strand breaks towards the homologous recombination repair pathway. Nature Communications *8*, 985–914.

Lou, Z., Minter-Dykhouse, K., Franco, S., Gostissa, M., Rivera, M.A., Celeste, A., Manis, J.P., Deursen, J.V., Nussenzweig, A., Paull, T.T., et al. (2006). MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. Molecular Cell *21*, 187–200.

Low, E., Chistol, G., Zaher, M.S., Kochenova, O.V., and Walter, J.C. (2020). The DNA replication fork suppresses CMG unloading from chromatin before termination. Genes \& Development *34*, 1534–1545.

Lu, T., Pan, Y., Kao, S.-Y., Li, C., Kohane, I., Chan, J., and Yankner, B.A. (2004). Gene regulation and DNA damage in the ageing human brain. Nature *429*, 883–891. .

Lujan, S.A., Williams, J.S., and Kunkel, T.A. (2016). DNA Polymerases Divide the Labor of Genome Replication. Trends in Cell Biology *26*, 640–654.

Lukas, C., Savic, V., Bekker-Jensen, S., Doil, C., Neumann, B., Pedersen, R.S., Grøfte, M., Chan, K.-L., Hickson, I.D., Bartek, J., et al. (2011). 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. Nature Cell Biology *13*, 243–253.

Lundblad, V., and Blackburn, E.H. (1993). An alternative pathway for yeast telomere maintenance rescues est1− senescence. Cell *73*, 347–360. https://doi.org/10.1016/0092-8674(93)90234-h.

Luo, G., Santoro, I.M., McDaniel, L.D., Nishijima, I., Mills, M., Youssoufian, H., Vogel, H., Schultz, R.A., and Bradley, A. (2000). Cancer predisposition caused by elevated mitotic recombination in Bloom mice. Nature Genetics *26*, 424–429.

Lydeard, J.R., Jain, S., Yamaguchi, M., and Haber, J.E. (2007). Break-induced replication and telomerase-independent telomere maintenance require Pol32. Nature *448*, 820–823.

Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M.R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell *108*, 781–794.

Machwe, A., Xiao, L., Groden, J., and Orren, D.K. (2006). The Werner and Bloom syndrome proteins catalyze regression of a model replication fork. Biochemistry *45*, 13939–13946.

Macias, M.J., Hyvönen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996). Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. Nature *382*, 646–649.

Macrae, C.J., McCulloch, R.D., Ylanko, J., Durocher, D., and Koch, C.A. (2008). APLF (C2orf13) facilitates nonhomologous end-joining and undergoes ATM-dependent hyperphosphorylation following ionizing radiation. Dna Repair *7*, 292–302. https://doi.org/10.1016/j.dnarep.2007.10.008.

Mailand, N., Bekker-Jensen, S., Bartek, J., and Lukas, J. (2006). Destruction of Claspin by SCFβTrCP Restrains Chk1 Activation and Facilitates Recovery from Genotoxic Stress. Mol Cell *23*, 307–318. https://doi.org/10.1016/j.molcel.2006.06.016.

Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C., and Lukas, J. (2007). RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. Cell *131*, 887–900.

Makharashvili, N., Tubbs, A.T., Yang, S.-H., Wang, H., Barton, O., Zhou, Y., Deshpande, R.A., Lee, J.-H., Lobrich, M., Sleckman, B.P., et al. (2014). Catalytic and Noncatalytic Roles of the CtIP Endonuclease in Double-Strand Break End Resection. Mol Cell *54*, 1022–1033. https://doi.org/10.1016/j.molcel.2014.04.011.

Mali, P., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013a). RNA-Guided Human Genome Engineering via Cas9. Science *339*, 823–826.

Mali, P., Esvelt, K.M., and Church, G.M. (2013b). Cas9 as a versatile tool for engineering biology. Nature Methods *10*, 957–963.

Mamely, I., Vugt, M.A. van, Smits, V.A., Semple, J.I., Lemmens, B., Perrakis, A., Medema, R.H., and Freire, R. (2006). Polo-like Kinase-1 Controls Proteasome-Dependent Degradation of Claspin during Checkpoint Recovery. Curr Biol *16*, 1950–1955. https://doi.org/10.1016/j.cub.2006.08.026.

Mandavilli, B.S., and Rao, K.S. (1996). Neurons in the cerebral cortex are most susceptible to DNA-damage in aging rat brain. Iubmb Life *40*, 507–514. https://doi.org/10.1080/15216549600201073.

Manthei, K.A., and Keck, J.L. (2013). The BLM dissolvasome in DNA replication and repair. Cellular and Molecular Life Sciences : CMLS *70*, 4067–4084.

Maric, M., Maculins, T., Piccoli, G.D., and Labib, K. (2014). Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. Science *346*, 1253596.

Martin, C.-A., Sarlós, K., Logan, C.V., Thakur, R.S., Parry, D.A., Bizard, A.H., Leitch, A., Cleal, L., Ali, N.S., Al-Owain, M.A., et al. (2018). Mutations in TOP3A Cause a Bloom Syndrome-like Disorder. American Journal of Human Genetics *103*, 221–231.

Mason, J.M., Dusad, K., Wright, W.D., Grubb, J., Budke, B., Heyer, W.-D., Connell, P.P., Weichselbaum, R.R., and Bishop, D.K. (2015). RAD54 family translocases counter genotoxic effects of RAD51 in human tumor cells. Nucleic Acids Res *43*, 3180–3196. https://doi.org/10.1093/nar/gkv175.

Matsushime, H., Ewen, M.E., Strom, D.K., Kato, J.Y., Hanks, S.K., Roussel, M.F., and Sherr, C.J. (1992). Identification and properties of an atypical catalytic subunit (p34PSK-J3/cdk4) for mammalian D type G1 cyclins. Cell *71*, 323–334. .

Mattiroli, F., Vissers, J.H.A., Dijk, W.J.V., Ikpa, P., Citterio, E., Vermeulen, W., Marteijn, J.A., and Sixma, T.K. (2012). RNF168 ubiquitinates K13-15 on H2A/H2AX to drive DNA damage signaling. Cell *150*, 1182–1195.

McClintock, B. (1941). THE STABILITY OF BROKEN ENDS OF CHROMOSOMES IN ZEA MAYS. Genetics *26*, 234–282. https://doi.org/10.1093/genetics/26.2.234.

McElhinny, S.A.N., Gordenin, D.A., Stith, C.M., Burgers, P.M.J., and Kunkel, T.A. (2008). Division of Labor at the Eukaryotic Replication Fork. Mol Cell *30*, 137–144. https://doi.org/10.1016/j.molcel.2008.02.022.

McElhinny, S.A.N., Kumar, D., Clark, A.B., Watt, D.L., Watts, B.E., Lundström, E.-B., Johansson, E., Chabes, A., and Kunkel, T.A. (2010). Genome instability due to ribonucleotide incorporation into DNA. Nat Chem Biol *6*, 774–781.

McGrath, R.A., and Williams, R.W. (1966). Reconstruction in vivo of irradiated Escherichia coli deoxyribonucleic acid; the rejoining of broken pieces. Nature *212*, 534–535. .

McKusick, V.A., Mahloudji, M., Abbott, M.H., Lindenberg, R., and Kepas, D. (1967). Seckel's Bird-Headed Dwarfism*. New Engl J Medicine *277*, 279–286. https://doi.org/10.1056/nejm196708102770602.

McMurray, C.T. (2010). Mechanisms of trinucleotide repeat instability during human development. Nature Reviews Genetics *11*, 786–799.

Meetei, A.R., Sechi, S., Wallisch, M., Yang, D., Young, M.K., Joenje, H., Hoatlin, M.E., and Wang, W. (2003). A Multiprotein Nuclear Complex Connects Fanconi Anemia and Bloom Syndrome. MCB *23*, 3417–3426.

Meetei, A.R., Winter, J.P. de, Medhurst, A.L., Wallisch, M., Waisfisz, Q., Vrugt, H.J. van de, Oostra, A.B., Yan, Z., Ling, C., Bishop, C.E., et al. (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. Nat Genet *35*, 165–170. https://doi.org/10.1038/ng1241.

Melander, F., Bekker-Jensen, S., Falck, J., Bartek, J., Mailand, N., and Lukas, J. (2008). Phosphorylation of SDT repeats in the MDC1 N terminus triggers retention of NBS1 at the DNA damage-modified chromatin. The Journal of Cell Biology *181*, 213–226.

Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.-P., St-Denis, N.A., Li, T., Miteva, Y.V., Hauri, S., Sardiu, M.E., Low, T.Y., et al. (2013). The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. Nature Methods *10*, 730–736.

Mendez-Dorantes, C., Bhargava, R., and Stark, J.M. (2018). Repeat-mediated deletions can be induced by a chromosomal break far from a repeat, but multiple pathways suppress such rearrangements. Genes \& Development *32*, 524–536.

Mendez-Dorantes, C., Tsai, L.J., Jahanshir, E., Lopezcolorado, F.W., and Stark, J.M. (2020). BLM has Contrary Effects on Repeat-Mediated Deletions, based on the Distance of DNA DSBs to a Repeat and Repeat Divergence. CellReports *30*, 1342–1357.e1344.

Mengwasser, K.E., Adeyemi, R.O., Leng, Y., Choi, M.Y., Clairmont, C., D'Andrea, A.D., and Elledge, S.J. (2019). Genetic Screens Reveal FEN1 and APEX2 as BRCA2 Synthetic Lethal Targets. Molecular Cell *73*, 885–899.e886.

Mészáros, B., Sámano-Sánchez, H., Alvarado-Valverde, J., Čalyševa, J., Martínez-Pérez, E., Alves, R., Shields, D.C., Kumar, M., Rippmann, F., Chemes, L.B., et al. (2021). Short linear motif candidates in the cell entry system used by SARS-CoV-2 and their potential therapeutic implications. Sci Signal *14*.

Mészáros, B., Simon, I., and Dosztányi, Z. (2009). Prediction of protein binding regions in disordered proteins. PLoS Comput Biol *5*, e1000376.

Meyerson, M., and Harlow, E. (1994). Identification of G1 kinase activity for cdk6, a novel cyclin D partner. Molecular and Cellular Biology *14*, 2077–2086. .

Micco, R.D., Fumagalli, M., Cicalese, A., Piccinin, S., Gasparini, P., Luise, C., Schurra, C., Garre', M., Nuciforo, P.G., Bensimon, A., et al. (2006). Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. Nature *444*, 638–642. https://doi.org/10.1038/nature05327.

Mikhailov, A., Cole, R.W., and Rieder, C.L. (2002). DNA Damage during Mitosis in Human Cells Delays the Metaphase/Anaphase Transition via the Spindle-Assembly Checkpoint. Curr Biol *12*, 1797–1806. https://doi.org/10.1016/s0960-9822(02)01226-5.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., et al. (1994). A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1. Science *266*, 66–71. https://doi.org/10.1126/science.7545954.

Mimori, T., and Hardin, J.A. (1986). Mechanism of interaction between Ku protein and DNA. J. Biol. Chem. *261*, 10375–10379.

Mirman, Z., Lottersberger, F., Takai, H., Kibe, T., Gong, Y., Takai, K., Bianchi, A., Zimmermann, M., Durocher, D., and Lange, T. (2018). 53BP1–RIF1–shieldin counteracts DSB resection through CST- and Polα-dependent fill-in. Nature 1–20.

Miyabe, I., Kunkel, T.A., and Carr, A.M. (2011). The Major Roles of DNA Polymerases Epsilon and Delta at the Eukaryotic Replication Fork Are Evolutionarily Conserved. Plos Genet *7*, e1002407. https://doi.org/10.1371/journal.pgen.1002407.

Miyaki, M., Konishi, M., Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Yasuno, M., Igari, T., Koike, M., Chiba, M., and Mori, T. (1997). Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. Nat Genet *17*, 271–272. https://doi.org/10.1038/ng1197-271.

Modesti, M., Budzowska, M., Baldeyron, C., Demmers, J.A.A., Ghirlando, R., and Kanaar, R. (2007). RAD51AP1 is a structure-specific DNA binding protein that stimulates joint molecule formation during RAD51-mediated homologous recombination. Molecular Cell *28*, 468–481.

Mol, C.D., Izumi, T., Mitra, S., and Tainer, J.A. (2000). DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination. Nature *403*, 451–456. https://doi.org/10.1038/35000249.

Mönnich, M., Hess, I., Wiest, W., Bachrati, C., Hickson, I.D., Schorpp, M., and Boehm, T. (2010). Developing T lymphocytes are uniquely sensitive to a lack of topoisomerase III alpha. European Journal of Immunology *40*, 2379–2384.

Montesinos, P., Beckermann, B.M., Catalani, O., Esteve, J., Gamel, K., Konopleva, M.Y., Martinelli, G., Monnet, A., Papayannidis, C., Park, A., et al. (2020). MIRROS: a randomized, placebo-controlled, Phase III trial of cytarabine ± idasanutlin in relapsed or refractory acute myeloid leukemia. Future Oncol *16*, 807–815.

Morales, M.E., Kaul, T., and Deininger, P. (2018). Long-Distance Relationships: Suppression of Repeat-Mediated Deletions. Trends Genet *34*, 572–574.

Mordes, D.A., Glick, G.G., Zhao, R., and Cortez, D. (2008). TopBP1 activates ATR through ATRIP and a PIKK regulatory domain. Genes \& Development *22*, 1478–1489.

Moreno, S.P., Bailey, R., Campion, N., Herron, S., and Gambus, A. (2014). Polyubiquitylation drives replisome disassembly at the termination of DNA replication. Science *346*, 477–481.

Morrison, A., and Sugino, A. (1994). The 3′ → 5′ exonucleases of both DNA polymerases δ and ε participate in correcting errors of DNA replication in Saccharomyces cerevisiae. Mol Gen Genetics Mgg *242*, 289–296. https://doi.org/10.1007/bf00280418.

Moshous, D., Callebaut, I., Chasseval, R. de, Corneo, B., Cavazzana-Calvo, M., Deist, F.L., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., et al. (2001). Artemis, a Novel DNA Double-Strand Break Repair/V(D)J Recombination Protein, Is Mutated in Human Severe Combined Immune Deficiency. Cell *105*, 177–186. https://doi.org/10.1016/s0092-8674(01)00309-9.

Motycka, T.A., Bessho, T., Post, S.M., Sung, P., and Tomkinson, A.E. (2004). Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. Journal of Biological Chemistry *279*, 13634–13639.

Moynahan, M.E., and Jasin, M. (1997). Loss of heterozygosity induced by a chromosomal double-strand break. Proc National Acad Sci *94*, 8988–8993. https://doi.org/10.1073/pnas.94.17.8988.

Munoz, I.M., Jowsey, P.A., Toth, R., and Rouse, J. (2007). Phospho-epitope binding by the BRCT domains of hPTIP controls multiple aspects of the cellular response to DNA damage. Nucleic Acids Research *35*, 5312–5322.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. Cell *102*, 553–563. https://doi.org/10.1016/s0092-8674(00)00078-7.

Muramatsu, S., Hirai, K., Tak, Y.-S., Kamimura, Y., and Araki, H. (2010). CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol ε, and GINS in budding yeast. Gene Dev *24*, 602–612. https://doi.org/10.1101/gad.1883410.

Myers, J.S., and Cortez, D. (2006). Rapid activation of ATR by ionizing radiation requires ATM and Mre11. Journal of Biological Chemistry *281*, 9346–9350. .

Nakamura, A.J., Rao, V.A., Pommier, Y., and Bonner, W.M. (2014). The complexity of phosphorylated H2AX foci formation and DNA repair assembly at DNA double-strand breaks. Cell Cycle *9*, 389–397. https://doi.org/10.4161/cc.9.2.10475.

Nakamura, K., Kogame, T., Oshiumi, H., Shinohara, A., Sumitomo, Y., Agama, K., Pommier, Y., Tsutsui, K.M., Tsutsui, K., Hartsuiker, E., et al. (2010). Collaborative action of Brca1 and CtIP in elimination of covalent modifications from double-strand breaks to facilitate subsequent break repair. PLoS Genetics *6*, e1000828.

Nakamura, K., Saredi, G., Becker, J.R., Foster, B.M., Nguyen, N.V., Beyer, T.E., Cesa, L.C., Faull, P.A., Lukauskas, S., Frimurer, T., et al. (2019). H4K20me0 recognition by BRCA1–BARD1 directs homologous recombination to sister chromatids. Nature Cell Biology *21*, 1–14.

Nakano, K., and Vousden, K.H. (2001). PUMA, a Novel Proapoptotic Gene, Is Induced by p53. Mol Cell *7*, 683–694. https://doi.org/10.1016/s1097-2765(01)00214-3.

Nakazawa, Y., Sasaki, K., Mitsutake, N., Matsuse, M., Shimada, M., Nardo, T., Takahashi, Y., Ohyama, K., Ito, K., Mishima, H., et al. (2012). Mutations in UVSSA cause UV-sensitive syndrome and impair RNA polymerase IIo processing in transcription-coupled nucleotide-excision repair. Nat Genet *44*, 586–592. https://doi.org/10.1038/ng.2229.

Nassif, N., Penney, J., Pal, S., Engels, W.R., and Gloor, G.B. (1994). Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Molecular and Cellular Biology *14*, 1613–1625.

Navarra, P., and Preziosi, P. (1999). Hydroxyurea: new insights on an old drug. Crit Rev Oncol Hematol *29*, 249–255.

Neale, M.J., and Keeney, S. (2006). Clarifying the mechanics of DNA strand exchange in meiotic recombination. Nature *442*, 153–158.

Neduva, V., and Russell, R.B. (2005). Linear motifs: evolutionary interaction switches. FEBS Letters *579*, 3342–3345.

Neelsen, K.J., and Lopes, M. (2015). Replication fork reversal in eukaryotes: from dead end to dynamic response. Nature Reviews Molecular Cell Biology *16*, 207–220.

Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability--an evolving hallmark of cancer. Nature Reviews Molecular Cell Biology *11*, 220–228.

Neill, C.A., and Dingwall, M.M. (1950). A Syndrome Resembling Progeria: A Review of Two Cases. Arch Dis Child *25*, 213. https://doi.org/10.1136/adc.25.123.213.

Nelson, G., Buhmann, M., and Zglinicki, T. von (2009). DNA damage foci in mitosis are devoid of 53BP1. Cell Cycle *8*, 3379–3383. https://doi.org/10.4161/cc.8.20.9857.

Nethanel, T., Reisfeld, S., Dinter-Gottlieb, G., and Kaufmann, G. (1988). An Okazaki piece of simian virus 40 may be synthesized by ligation of shorter precursor chains. J Virol *62*, 2867–2873. https://doi.org/10.1128/jvi.62.8.2867-2873.1988.

Nicolaides, N.C., Papadopoulos, N., Liu, B., Weit, Y.-F., Carter, K.C., Ruben, S.M., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., et al. (1994). Mutations of two P/WS homologues in hereditary nonpolyposis colon cancer. Nature *371*, 75–80. https://doi.org/10.1038/371075a0.

Nieminuszczy, J., Schwab, R.A., and Niedzwiedz, W. (2016). The DNA fibre technique - tracking helicases at work. Methods (San Diego, Calif.) *108*, 92–98.

Nimonkar, A.V., and Kowalczykowski, S.C. (2009). Second-end DNA capture in double-strand break repair: how to catch a DNA by its tail. Cell Cycle *8*, 1816–1817.

Nimonkar, A.V., Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J.L., Wyman, C., Modrich, P., and Kowalczykowski, S.C. (2011). BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. Genes \& Development *25*, 350–362.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods *6*, 917–922. https://doi.org/10.1038/nmeth.1401.

Nishizuka, Y., Ueda, K., Nakazawa, K., and Hayaishi, O. (1967). Studies on the polymer of adenosine diphosphate ribose. I. Enzymic formation from nicotinamide adenine dinuclotide in mammalian nuclei. Journal of Biological Chemistry *242*, 3164–3171. .

Noordermeer, S.M., Adam, S.X., Setiaputra, D., Barazas, M., Pettitt, S.J., Ling, A.K., Olivieri, M., n, A.A.X.L.-Q.X., Moatti, N., Zimmermann, M., et al. (2018). The shieldin complex mediates 53BP1-dependent DNA repair. Nature 1–25.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment 1 1Edited by J. Thornton. Journal of Molecular Biology *302*, 205–217.

Nurse, P. (1975). Genetic control of cell size at cell division in yeast. Nature *256*, 547–551.
.

O'Driscoll, M., and Jeggo, P.A. (2008). The role of the DNA damage response pathways in brain development and microcephaly: insight from human disorders. DNA Repair *7*, 1039–1050.

O'Driscoll, M., Ruiz-Perez, V.L., Woods, C.G., Jeggo, P.A., and Goodship, J.A. (2003). A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. Nature Genetics *33*, 497–501.

O'Driscoll, M., Cerosaletti, K.M., Girard, P.-M., Dai, Y., Stumm, M., Kysela, B., Hirsch, B., Gennery, A., Palmer, S.E., Seidel, J., et al. (2001). DNA Ligase IV Mutations Identified in Patients Exhibiting Developmental Delay and Immunodeficiency. Mol Cell *8*, 1175–1185. https://doi.org/10.1016/s1097-2765(01)00408-7.

O'Driscoll, M., Ruiz-Perez, V.L., Woods, C.G., Jeggo, P.A., and Goodship, J.A. (2003). A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. Nature Genetics *33*, 497–501. https://doi.org/10.1038/ng1129.

Obenauer, J.C., Cantley, L.C., and Yaffe, M.B. (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Research *31*, 3635–3641.

Ochi, T., Blackford, A.N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., Travers, J., Wu, Q., Draviam, V.M., Robinson, C.V., et al. (2015). PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. Science *347*, 185–188.

Ochs, F., Karemore, G., Miron, E., Brown, J., Sedlackova, H., Rask, M.-B., Lampe, M., Buckle, V., Schermelleh, L., Lukas, J., et al. (2019). Stabilization of chromatin topology safeguards genome integrity. Nature *574*, 571–574.

Ochs, F., Somyajit, K., Altmeyer, M., Rask, M.-B., Lukas, J., and Lukas, C. (2016). 53BP1 fosters fidelity of homology-directed DNA repair. Nature Structural \& Molecular Biology *23*, 714–721.

Oettinger, M.A., Schatz, D.G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, Adjacent Genes That Synergistically Activate V(D)J Recombination. Science *248*, 1517–1523. https://doi.org/10.1126/science.2360047.

Ogi, T., Walker, S., Stiff, T., Hobson, E., Limsirichaikul, S., Carpenter, G., Prescott, K., Suri, M., Byrd, P.J., Matsuse, M., et al. (2012). Identification of the first ATRIP-deficient patient and novel mutations in ATR define a clinical spectrum for ATR-ATRIP Seckel Syndrome. PLoS Genetics *8*, e1002945.

Ogi, T., Walker, S., Stiff, T., Hobson, E., Limsirichaikul, S., Carpenter, G., Prescott, K., Suri, M., Byrd, P.J., Matsuse, M., et al. (2012). Identification of the First ATRIP–Deficient

Patient and Novel Mutations in ATR Define a Clinical Spectrum for ATR–ATRIP Seckel Syndrome. Plos Genet *8*, e1002945. https://doi.org/10.1371/journal.pgen.1002945.

Okazaki, R., Okazaki, T., Sakabe, K., and Sugimoto, K. (1967). Mechanism of DNA replication possible discontinuity of DNA chain growth. Jpn J Med Sci Biol *20*, 255–260. .

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., and Sugino, A. (1968). Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. Proc National Acad Sci *59*, 598–605. https://doi.org/10.1073/pnas.59.2.598.

Oliver, T.G., Mercer, K.L., Sayles, L.C., Burke, J.R., Mendus, D., Lovejoy, K.S., Cheng, M.-H., Subramanian, A., Mu, D., Powers, S., et al. (2010). Chronic cisplatin treatment promotes enhanced damage repair and tumor progression in a mouse model of lung cancer. Genes \& Development *24*, 837–852.

Olivieri, M., Cho, T., Álvarez-Quilón, A., Li, K., Schellenberg, M.J., Zimmermann, M., Hustedt, N., Rossi, S.E., Adam, S., Melo, H., et al. (2019). A genetic map of the response to DNA damage in human cells. *1703*, 283–252.

Olovnikov, A.M. (1973). A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. Journal of Theoretical Biology *41*, 181–190. .

Olson, E., Nievera, C.J., Liu, E., Lee, A.Y.-L., Chen, L., and Wu, X. (2007). The Mre11 complex mediates the S-phase checkpoint through an interaction with replication protein A. Molecular and Cellular Biology *27*, 6053–6067.

Orthwein, A., Fradet-Turcotte, A., Noordermeer, S.M., Canny, M.D., Brun, C.M., Strecker, J., Escribano-Díaz, C., and Durocher, D. (2014). Mitosis inhibits DNA double-strand break repair to guard against telomere fusions. Science *344*, 189–193.

Orthwein, A., Noordermeer, S.M., Wilson, M.D., Landry, S., Enchev, R.I., Sherker, A., Munro, M., Pinder, J., Salsman, J., Dellaire, G., et al. (2015). A mechanism for the suppression of homologous recombination in G1 cells. Nature *528*, 422–426. https://doi.org/10.1038/nature16142.

Osheroff, N. (1989). Effect of antineoplastic agents on the DNA cleavage/religation reaction of eukaryotic topoisomerase II: inhibition of DNA religation by etoposide. Biochemistry *28*, 6157–6160. .

Pacek, M., and Walter, J.C. (2004). A requirement for MCM7 and Cdc45 in chromosome unwinding during eukaryotic DNA replication. The EMBO Journal *23*, 3667–3676.

Pagano, M., Pepperkok, R., Verde, F., Ansorge, W., and Draetta, G. (1992). Cyclin A is required at two points in the human cell cycle. The EMBO Journal *11*, 961–971. .

Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T., and Jiricny, J. (1996). hMutSβ, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. Curr Biol *6*, 1181–1184. https://doi.org/10.1016/s0960-9822(02)70685-4.

Panier, S., Maric, M., Hewitt, G., Mason-Osann, E., Gali, H., Dai, A., Labadorf, A., Guervilly, J.-H., Ruis, P., Segura-Bayona, S., et al. (2019). SLX4IP Antagonizes Promiscuous BLM Activity during ALT Maintenance. Molecular Cell *76*, 27–43.e11.

Papadopoulos, N., Nicolaides, N.C., Wei, Y.-F., Ruben, S.M., Carter, K.C., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., Adams, M.D., et al. (1994). Mutation of a mutL Homolog in Hereditary Colon Cancer. Science *263*, 1625–1629. https://doi.org/10.1126/science.8128251.

Patel, D.S., Misenko, S.M., Her, J., and Bunting, S.F. (2017). BLM helicase regulates DNA repair by counteracting RAD51 loading at DNA double-strand break sites. The Journal of Cell Biology *216*, 3521–3534.

Paull, T.T., and Gellert, M. (1998). The 3′ to 5′ Exonuclease Activity of Mre11 Facilitates Repair of DNA Double-Strand Breaks. Mol Cell *1*, 969–979. https://doi.org/10.1016/s1097-2765(00)80097-0.

Pearl, L.H., Schierz, A.C., Ward, S.E., Al-Lazikani, B., and Pearl, F.M.G. (2015). Therapeutic opportunities within the DNA damage response. Nat Rev Cancer *15*, 166–180. https://doi.org/10.1038/nrc3891.

Pearl, L.H., Schierz, A.C., Ward, S.E., Al-Lazikani, B., and Pearl, F.M.G. (2015). Therapeutic opportunities within the DNA damage response. Nature Publishing Group *15*, 166–180.

Pellegrini, L., Yu, D.S., Lo, T., Anand, S., Lee, M., Blundell, T.L., and Venkitaraman, A.R. (2002). Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. Nature *420*, 287–293.

Pellegrino, S., Michelena, J., Teloni, F., Imhof, R., and Altmeyer, M. (2017). Replication-Coupled Dilution of H4K20me2 Guides 53BP1 to Pre-replicative Chromatin. CellReports *19*, 1819–1831.

Peng, C.-Y., Graves, P.R., Thoma, R.S., Wu, Z., Shaw, A.S., and Piwnica-Worms, H. (1997). Mitotic and G2 Checkpoint Control: Regulation of 14-3-3 Protein Binding by Phosphorylation of Cdc25C on Serine-216. Science *277*, 1501–1505. https://doi.org/10.1126/science.277.5331.1501.

Pepe, A., and West, S.C. (2014). MUS81-EME2 promotes replication fork restart. CellReports *7*, 1048–1055.

Peschiaroli, A., Dorrello, N.V., Guardavaccaro, D., Venere, M., Halazonetis, T., Sherman, N.E., and Pagano, M. (2006). SCFβTrCP-Mediated Degradation of Claspin Regulates Recovery from the DNA Replication Checkpoint Response. Mol Cell *23*, 319–329. https://doi.org/10.1016/j.molcel.2006.06.013.

Petermann, E., and Helleday, T. (2010). Pathways of mammalian replication fork restart. Nature Reviews Molecular Cell Biology *11*, 683–687.

Petermann, E., Orta, M.L., Issaeva, N., Schultz, N., and Helleday, T. (2010). Hydroxyurea-Stalled Replication Forks Become Progressively Inactivated and Require Two Different RAD51-Mediated Pathways for Restart and Repair. Molecular Cell *37*, 492–502.

Peterson, S.E., Li, Y., Chait, B.T., Gottesman, M.E., Baer, R., and Gautier, J. (2011). Cdk1 uncouples CtIP-dependent resection and Rad51 filament formation during M-phase double-strand break repair. J Cell Biol *194*, 705–720. https://doi.org/10.1083/jcb.201103103.

Phillips, D.H., Hewer, A., Martin, C.N., Garner, R.C., and King, M.M. (1988). Correlation of DNA adduct levels in human lung with cigarette smoking. Nature *336*, 790–792.

Pichierri, P., Franchitto, A., and Rosselli, F. (2004). BLM and the FANC proteins collaborate in a common pathway in response to stalled replication forks. The EMBO Journal *23*, 3154–3163.

Pines, J., and Hunter, T. (1989). Isolation of a human cyclin cDNA: evidence for cyclin mRNA and protein regulation in the cell cycle and for interaction with p34cdc2. Cell *58*, 833–846. .

Plank, J.L., Wu, J., and Hsieh, T. (2006). Topoisomerase IIIa and Bloom's helicase can resolve a mobile double Holliday junction substrate through convergent branch migration. Proc Natl Acad Sci USA *103*, 11118–11123.

Podust, L.M., Podust, V.N., Sogo, J.M., and Hübscher, U. (1995). Mammalian DNA polymerase auxiliary proteins: analysis of replication factor C-catalyzed proliferating cell nuclear antigen loading onto circular double-stranded DNA. Mol Cell Biol *15*, 3072–3081. https://doi.org/10.1128/mcb.15.6.3072.

Podust, V.N., and Hübscher, U. (1993). Lagging strand DNA synthesis by calf thymus DNA polymerases α, β, δ and ε in the presence of auxiliary proteins. Nucleic Acids Res *21*, 841–846. https://doi.org/10.1093/nar/21.4.841.

Polo, S.E., Blackford, A.N., Chapman, J.R., Baskcomb, L., Gravel, S., Rusch, A., Thomas, A., Blundred, R., Smith, P., Kzhyshkowska, J., et al. (2012). Regulation of DNA-End Resection by hnRNPU-like Proteins Promotes DNA Double-Strand Break Signaling and Repair. Molecular Cell *45*, 505–516.

Postow, L., and Funabiki, H. (2013). An SCF complex containing Fbxl12 mediates DNA damage-induced Ku80 ubiquitylation. Cell Cycle *12*, 587–595. .

Postow, L., Ghenoiu, C., Woo, E.M., Krutchinsky, A.N., Chait, B.T., and Funabiki, H. (2008). Ku80 removal from DNA through double strand break–induced ubiquitylation. J Cell Biology *182*, 467–479. https://doi.org/10.1083/jcb.200802146.

Prakash, R., Zhang, Y., Feng, W., and Jasin, M. (2015). Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. Cold Spring Harbor Perspectives in Biology *7*, a016600.

Prescott, D.M., and Bender, M.A. (1962). Synthesis of RNA and protein during mitosis in mammalian tissue culture cells. Exp Cell Res *26*, 260–268. https://doi.org/10.1016/0014-4827(62)90176-3.

Pursell, Z.F., Isoz, I., Lundström, E.-B., Johansson, E., and Kunkel, T.A. (2007). Yeast DNA Polymerase ε Participates in Leading-Strand DNA Replication. Science *317*, 127–130. https://doi.org/10.1126/science.1144067.

Qi, Z., Redding, S., Lee, J.Y., Gibb, B., Kwon, Y., Niu, H., Gaines, W.A., Sung, P., and Greene, E.C. (2015). DNA sequence alignment by microhomology sampling during homologous recombination. Cell *160*, 856–869.

Qiu, R., DeRocco, V.C., Harris, C., Sharma, A., Hingorani, M.M., Erie, D.A., and Weninger, K.R. (2012). Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling. Embo J *31*, 2528–2540. https://doi.org/10.1038/emboj.2012.95.

Raderschall, E., Golub, E.I., and Haaf, T. (1999). Nuclear foci of mammalian recombination proteins are located at single-stranded DNA regions formed after DNA damage. Proceedings of the National Academy of Sciences *96*, 1921–1926. .

Ralf, C., Hickson, I.D., and Wu, L. (2006). The Bloom's syndrome helicase can promote the regression of a model replication fork. J. Biol. Chem. *281*, 22839–22846.

Ramberger, E., Suarez-Artiles, L., Perez-Hernandez, D., Haji, M., Popp, O., Reimer, U., Leutz, A., Dittmar, G., and Mertins, P. (2021). A Universal Peptide Matrix Interactomics Approach to Disclose Motif-Dependent Protein Binding. Molecular \& Cellular Proteomics *20*, 100135.

Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell *154*, 1380–1389.

Rastogi, R.P., Richa, Kumar, A., Tyagi, M.B., and Sinha, R.P. (2010). Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. Journal of Nucleic Acids *2010*, 592980.

Raynard, S., Bussen, W., and Sung, P. (2006). A double Holliday junction dissolvasome comprising BLM, topoisomerase IIIalpha, and BLAP75. J. Biol. Chem. *281*, 13861–13864.

Raynard, S., Zhao, W., Bussen, W., Lu, L., Ding, Y.-Y., Busygina, V., Meetei, A.R., and Sung, P. (2008). Functional Role of BLAP75 in BLM-Topoisomerase III□-dependent Holliday Junction Processing. J. Biol. Chem. *283*, 15701–15708.

Reardon, J.T., Nichols, A.F., Keeney, S., Smith, C.A., Taylor, J.S., Linn, S., and Sancar, A. (1993). Comparative analysis of binding of human damaged DNA-binding protein (XPE) and Escherichia coli damage recognition protein (UvrA) to the major ultraviolet photoproducts: T[c,s]T, T[t,s]T, T[6-4]T, and T[Dewar]T. J Biological Chem *268*, 21301–21308. .

Reczek, C.R., Szabolcs, M., Stark, J.M., Ludwig, T., and Baer, R. (2013). The interaction between CtIP and BRCA1 is not essential for resection-mediated DNA repair or tumor suppression. The Journal of Cell Biology *201*, 693–707.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F.X. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell *139*, 719–730.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F.X. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell *139*, 719–730. https://doi.org/10.1016/j.cell.2009.10.015.

replication forks to maintain genome integrity. Nature Cell Biology *18*, 1185–1195.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A., et al. (2000). Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2). Cell *102*, 565–575. https://doi.org/10.1016/s0092-8674(00)00079-9.

Rhee, H.-W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science *339*, 1328–1331.

Riballo, E., Critchlow, S.E., Teo, S.-H., Doherty, A.J., Priestley, A., Broughton, B., Kysela, B., Beamish, H., Plowman, N., Arlett, C.F., et al. (1999). Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient. Curr Biol *9*, 699-S2. https://doi.org/10.1016/s0960-9822(99)80311-x.

Rieder, C.L., and Cole, R.W. (1998). Entry into mitosis in vertebrate somatic cells is guarded by a chromosome damage checkpoint that reverses the cell cycle when triggered during early but not late prophase. J Cell Biol *142*, 1013–1022. https://doi.org/10.1083/jcb.142.4.1013.

Rieder, C.L., and Cole, R.W. (1998). Entry into mitosis in vertebrate somatic cells is guarded by a chromosome damage checkpoint that reverses the cell cycle when triggered during early but not late prophase. The Journal of Cell Biology *142*, 1013–1022.

Riley, T., Sontag, E., Chen, P., and Levine, A. (2008). Transcriptional control of human p53-regulated genes. Nature Reviews Molecular Cell Biology *9*, 402–412.

Robey, R.W., Pluchino, K.M., Hall, M.D., Fojo, A.T., Bates, S.E., and Gottesman, M.M. (2018). Revisiting the role of ABC transporters in multidrug-resistant cancer. Nature Reviews Cancer *18*, 452–464. .

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. Journal of Biological Chemistry *273*, 5858–5868.

Romanienko, P.J., and Camerini-Otero, R.D. (1999). Cloning, Characterization, and Localization of Mouse and Human SPO11. Genomics *61*, 156–169. https://doi.org/10.1006/geno.1999.5955.

Roninson, I.B., Chin, J.E., Choi, K.G., Gros, P., Housman, D.E., Fojo, A., Shen, D.W., Gottesman, M.M., and Pastan, I. (1986). Isolation of human mdr DNA sequences amplified in multidrug-resistant KB carcinoma cells. Proceedings of the National Academy of Sciences *83*, 4538–4542. .

Rosenberg, B., VanCamp, L., Trosko, J.E., and Mansour, V.H. (1969). Platinum compounds: a new class of potent antitumour agents. Nature *222*, 385–386. .

Rossi, D.J., Bryder, D., Seita, J., Nussenzweig, A., Hoeijmakers, J., and Weissman, I.L. (2007). Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. Nature *447*, 725–729.

Rossi, M.L., Pike, J.E., Wang, W., Burgers, P.M.J., Campbell, J.L., and Bambara, R.A. (2008). Pif1 Helicase Directs Eukaryotic Okazaki Fragments toward the Two-nuclease Cleavage Pathway for Primer Removal*. J Biol Chem *283*, 27483–27493. https://doi.org/10.1074/jbc.m804550200.

Royou, A., Gagou, M.E., Karess, R., and Sullivan, W. (2010). BubR1- and Polo-coated DNA tethers facilitate poleward segregation of acentric chromatids. Cell *140*, 235–245.

Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173–1178.

Ruis, P., and Boulton, S.J. (2021). The end protection problem-an unexpected twist in the tail. Genes \& Development *35*, 1–21.

Russell, P., and Nurse, P. (1986). cdc25+ functions as an inducer in the mitotic control of fission yeast. Cell *45*, 145–153. https://doi.org/10.1016/0092-8674(86)90546-5.

Saini, N., Ramakrishnan, S., Elango, R., Ayyar, S., Zhang, Y., Deem, A., Ira, G., Haber, J.E., Lobachev, K.S., and Malkova, A. (2013). Migrating bubble during break-induced replication drives conservative DNA synthesis. Nature *502*, 389–392.

Salguero, I., Belotserkovskaya, R., Coates, J., Sczaniecka-Clift, M., Demir, M., Jhujh, S., Wilson, M.D., and Jackson, S.P. (2019). MDC1 PST-repeat region promotes histone H2AX- independent chromatin association and DNA damage tolerance. Nature Communications 1–11.

Sanchez, Y., Wong, C., Thoma, R.S., Richman, R., Wu, Z., Piwnica-Worms, H., and Elledge, S.J. (1997). Conservation of the Chk1 Checkpoint Pathway in Mammals: Linkage of DNA Damage to Cdk Regulation Through Cdc25. Science *277*, 1497–1501. https://doi.org/10.1126/science.277.5331.1497.

Saredi, G., Huang, H., Hammond, C.M., Alabert, C., Bekker-Jensen, S., Forne, I., Reverón-Gómez, N., Foster, B.M., Mlejnkova, L., Bartke, T., et al. (2016). H4K20me0 marks post-replicative chromatin and recruits the TONSL–MMS22L DNA repair complex. Nature *534*, 714–718.

Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S.P. (2007). Human CtIP promotes DNA end resection. Nature *450*, 509–514.

Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D.A., Smith, S., Uziel, T., Sfez, S., et al. (1995). A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science *268*, 1749–1753. .

Schatz, D.G., Oettinger, M.A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. Cell *59*, 1035–1048. https://doi.org/10.1016/0092-8674(89)90760-5.

Schayek, H., Laitman, Y., Katz, L.H., Pras, E., Ries-Levavi, L., Barak, F., and Friedman, E. (2017). Colorectal and Endometrial Cancer Risk and Age at Diagnosis in BLMAsh Mutation Carriers. Israel Medical Assoc J Imaj *19*, 365–367. .

Schendel, R.V., Roerink, S.F., Portegijs, V., Heuvel, S.V.D., and Tijsterman, M. (2015). Polymerase Θ is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis. Nature Communications *6*, 7394–7398.

Schimke, R.N., Horton, W.A., King, C.R., and Martin, N.L. (1974). Chondroitin-6-sulfate mucopoly-saccharidosis in conjunction with lymphopenia, defective cellular immunity and the nephrotic syndrome. Birth Defects Orig Article Ser *10*, 258–266. .

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods *9*, 676–682.

Schlacher, K., Christ, N., Siaud, N., Egashira, A., Wu, H., and Jasin, M. (2011). Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. Cell *145*, 529–542.

Schlacher, K., Wu, H., and Jasin, M. (2012). A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. Cancer Cell *22*, 106–116.

Schmidt, C.K., Galanty, Y., Sczaniecka-Clift, M., Coates, J., Jhujh, S., Demir, M., Cornwell, M., Beli, P., and Jackson, S.P. (2015). Systematic E2 screening reveals a UBE2D-RNF138-CtIP axis promoting DNA repair. Nature Cell Biology *17*, 1458–1470.

Schreiber, V., Dantzer, F. coise, Ame, J.-C., and Murcia, G. de (2006). Poly(ADP-ribose): novel functions for an old molecule. Nature Reviews Molecular Cell Biology *7*, 517–528. .

Schreiber, V., Dantzer, F., Ame, J.-C., and Murcia, G. de (2006). Poly(ADP-ribose): novel functions for an old molecule. Nature Reviews Molecular Cell Biology *7*, 517–528.

Schubert, E.L., Lee, M.K., Mefford, H.C., Argonza, R.H., Morrow, J.E., Hull, J., Dann, J.L., and King, M.C. (1997). BRCA2 in American families with four or more cases of breast or ovarian cancer: recurrent and novel mutations, variable expression, penetrance, and the possibility of families whose cancer is not attributable to BRCA1 or BRCA2. Am J Hum Genet *60*, 1031–1040. .

Schwertman, P., Lagarou, A., Dekkers, D.H.W., Raams, A., Hoek, A.C. van der, Laffeber, C., Hoeijmakers, J.H.J., Demmers, J.A.A., Fousteri, M., Vermeulen, W., et al. (2012). UV-sensitive syndrome protein UVSSA recruits USP7 to regulate transcription-coupled repair. Nat Genet *44*, 598–602. https://doi.org/10.1038/ng.2230.

Scully, R., and Xie, A. (2013). Double strand break repair functions of histone H2AX. Mutation Research *750*, 5–14.

Sears, R.M., May, D.G., and Roux, K.J. (2019). BioID as a Tool for Protein-Proximity Labeling in Living Cells. Methods in Molecular Biology (Clifton, N.J.) *2012*, 299–313.

Sebesta, M., Burkovics, P., Juhasz, S., Zhang, S., Szabo, J.E., Lee, M.Y.W.T., Haracska, L., and Krejci, L. (2013). Role of PCNA and TLS polymerases in D-loop extension during homologous recombination in humans. DNA Repair *12*, 691–698.

Seckel HP. Bird-headed Dwarfs: Studies in developmental anthropology including human proportions. S Karger, Springfield, C.C. Thomas: Basel-New York; 1960

Sedelnikova, O.A., Rogakou, E.P., Panyutin, I.G., and Bonner, W.M. (2002). Quantitative detection of (125)IdU-induced DNA double-strand breaks with gamma-H2AX antibody. Radiat Res *158*, 486–492.

Seeman, P., Gebertová, K., Paděrová, K., Sperling, K., and Seemanová, E. (2004). Nijmegen breakage syndrome in 13% of age-matched Czech children with primary microcephaly. Pediatr Neurol *30*, 195–200. https://doi.org/10.1016/j.pediatrneurol.2003.07.003.

Seidman, J.G., Leder, A., Edgell, M.H., Polsky, F., Tilghman, S.M., Tiemeier, D.C., and Leder, P. (1978). Multiple related immunoglobulin variable-region genes identified by cloning and sequence analysis. Proc National Acad Sci *75*, 3881–3885. https://doi.org/10.1073/pnas.75.8.3881.

Setiaputra, D., and Durocher, D. (2019). Shieldin - the protector of DNA ends. EMBO Reports *20*, e47560.

Shiotani, B., and Zou, L. (2009). Single-stranded DNA orchestrates an ATM-to-ATR switch at DNA breaks. Molecular Cell *33*, 547–558. .

Shorrocks, A.-M.K., Jones, S.E., Tsukada, K., Morrow, C.A., Belblidia, Z., Shen, J., Vendrell, I., Fischer, R., Kessler, B.M., and Blackford, A.N. (2021). The Bloom syndrome complex senses RPA-coated single-stranded DNA to restart stalled replication forks. Nature Communications *12*, 585–515.

Shoulkamy, M.I., Nakano, T., Ohshima, M., Hirayama, R., Uzawa, A., Furusawa, Y., and Ide, H. (2012). Detection of DNA-protein crosslinks (DPCs) by novel direct fluorescence labeling methods: distinct stabilities of aldehyde and radiation-induced DPCs. Nucleic Acids Research *40*, e143.

Siddiqui, K., On, K.F., and Diffley, J.F.X. (2013). Regulating DNA replication in eukarya. Cold Spring Harbor Perspectives in Biology *5*, a012930.

Siddiqui, K., On, K.F., and Diffley, J.F.X. (2013). Regulating DNA replication in eukarya. Cold Spring Harbor Perspectives in Biology *5*, a012930. .

Siitonen, H.A., Kopra, O., Kääriäinen, H., Haravuori, H., Winter, R.M., Säämänen, A.-M., Peltonen, L., and Kestilä, M. (2003). Molecular defect of RAPADILINO syndrome expands the phenotype spectrum of RECQL diseases. Human Molecular Genetics *12*, 2837–2844.

Simonetta, M., Krijger, I. de, Serrat, J., Moatti, N., Fortunato, D., Hoekman, L., Bleijerveld, O.B., Altelaar, A.F.M., and Jacobs, J.J.L. (2018). H4K20me2 distinguishes pre-replicative from post-replicative chromatin to appropriately direct DNA repair pathway choice by 53BP1-RIF1-MAD2L2. Cell Cycle *17*, 124–136.

Singh, T.R., Ali, A.M., Busygina, V., Raynard, S., Fan, Q., Du, C.-H., Andreassen, P.R., Sung, P., and Meetei, A.R. (2008). BLAP18/RMI2, a novel OB-fold-containing protein, is an essential component of the Bloom helicase-double Holliday junction dissolvasome. Genes \& Development *22*, 2856–2868.

Skoog, L., and NordenskjÖLD, B. (1971). Effects of Hydroxyurea and 1-β-D-Arabinofuranosyl-cytosine on Deoxyribonucleotide Pools in Mouse Embryo Cells. Eur J Biochem *19*, 81–89. https://doi.org/10.1111/j.1432-1033.1971.tb01290.x.

Slater, M.L. (1973). Effect of Reversible Inhibition of Deoxyribonucleic Acid Synthesis on the Yeast Cell Cycle. J Bacteriol *113*, 263–270. https://doi.org/10.1128/jb.113.1.263-270.1973.

Snapper, C.M., and Paul, W.E. (1987). Interferon-γ and B Cell Stimulatory Factor-1 Reciprocally Regulate Ig Isotype Production. Science *236*, 944–947. https://doi.org/10.1126/science.3107127.

Sneeden, J.L., Grossi, S.M., Tappin, I., Hurwitz, J., and Heyer, W.-D. (2013). Reconstitution of recombination-associated DNA synthesis with human proteins. Nucleic Acids Research *41*, 4913–4925.

Solinger, J.A., Kiianitsa, K., and Heyer, W.-D. (2002). Rad54, a Swi2/Snf2-like Recombinational Repair Protein, Disassembles Rad51:dsDNA Filaments. Mol Cell *10*, 1175–1188. https://doi.org/10.1016/s1097-2765(02)00743-8.

Somyajit, K., Spies, J., Coscia, F., Kirik, U., Rask, M.-B., Lee, J.-H., Neelsen, K.J., Mund, A., Jensen, L.J., Paull, T.T., et al. (2021). Homology-directed repair protects the replicating genome from metabolic assaults. Developmental Cell *56*, 461–477.e467.

Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human deubiquitinating enzyme interaction landscape. Cell *138*, 389–403.

Spies, J., Lukas, C., Somyajit, K., Rask, M.-B., Lukas, J., and Neelsen, K.J. (2019). 53BP1 nuclear bodies enforce replication timing at under-replicated DNA to limit heritable DNA damage. Nature Cell Biology *21*, 487–497.

Spivak, G. (2015). Nucleotide excision repair in humans. Dna Repair *36*, 13–18. https://doi.org/10.1016/j.dnarep.2015.09.003.

Spranger, J., Hinkel, G.K., Stöss, H., Thoenes, W., Wargowski, D., and Zepp, F. (1991). Schimke immuno-osseous dysplasia: A newly recognized multisystem disease. J Pediatrics *119*, 64–72. https://doi.org/10.1016/s0022-3476(05)81040-6.

Spycher, C., Miller, E.S., Townsend, K., Pavic, L., Morrice, N.A., Janscak, P., Stewart, G.S., and Stucki, M. (2008). Constitutive phosphorylation of MDC1 physically links the MRE11-RAD50-NBS1 complex to damaged chromatin. The Journal of Cell Biology *181*, 227–240.

Stasiak, A., and Egelman, E.H. (1994). Structure and function of RecA-DNA complexes. Experientia *50*, 192–203.

Stavnezer, J., Guikema, J.E.J., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. Annu Rev Immunol *26*, 261–292. https://doi.org/10.1146/annurev.immunol.26.021607.090248.

Steensel, B. van, Smogorzewska, A., and Lange, T. de (1998). TRF2 Protects Human Telomeres from End-to-End Fusions. Cell *92*, 401–413. https://doi.org/10.1016/s0092-8674(00)80932-0.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell *144*, 27–40.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell *144*, 27–40. https://doi.org/10.1016/j.cell.2010.11.055.

Stewart, G.S., Maser, R.S., Stankovic, T., Bressan, D.A., Kaplan, M.I., Jaspers, N.G., Raams, A., Byrd, P.J., Petrini, J.H., and Taylor, A.M. (1999). The DNA double-strand

break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. Cell *99*, 577–587. .

Stewart, G.S., Panier, S., Townsend, K., Al-Hakim, A.K., Kolas, N.K., Miller, E.S., Nakada, S., Ylanko, J., Olivarius, S., Mendez, M., et al. (2009). The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. Cell *136*, 420–434.

Stewart, M.D., Ritterhoff, T., Klevit, R.E., and Brzovic, P.S. (2016). E2 enzymes: more than just middle men. Cell Research *26*, 423–440.

Stiff, T., O'Driscoll, M., Rief, N., Iwabuchi, K., Löbrich, M., and Jeggo, P.A. (2004). ATM and DNA-PK function redundantly to phosphorylate H2AX after exposure to ionizing radiation. Cancer Research *64*, 2390–2396.

Stigger, E., Drissi, R., and Lee, S.H. (1998). Functional analysis of human replication protein A in nucleotide excision repair. Journal of Biological Chemistry *273*, 9337–9343.

Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 Directly Binds Phosphorylated Histone H2AX to Regulate Cellular Responses to DNA Double-Strand Breaks. Cell *123*, 1213–1226.

Stults, D.M., Killen, M.W., Marco-Casanova, P., and Pierce, A.J. (2020). The Sister Chromatid Exchange (SCE) Assay. Methods in Molecular Biology (Clifton, N.J.) *2102*, 441–457.

Sudakin, V., Ganoth, D., Dahan, A., Heller, H., Hershko, J., Luca, F.C., Ruderman, J.V., and Hershko, A. (1995). The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. Molecular Biology of the Cell *6*, 185–197. .

Sugasawa, K., Ng, J.M.Y., Masutani, C., Iwai, S., Spek, P.J. van der, Eker, A.P.M., Hanaoka, F., Bootsma, D., and Hoeijmakers, J.H.J. (1998). Xeroderma Pigmentosum Group C Protein Complex Is the Initiator of Global Genome Nucleotide Excision Repair. Mol Cell *2*, 223–232. https://doi.org/10.1016/s1097-2765(00)80132-x.

Sugimoto, K., Okazaki, T., and Okazaki, R. (1968). Mechanism of DNA chain growth, II. Accumulation of newly synthesized short chains in E. coli infected with ligase-defective T4 phages. Proc National Acad Sci *60*, 1356–1362. https://doi.org/10.1073/pnas.60.4.1356.

Sun, H., Karow, J.K., Hickson, I.D., and Maizels, N. (1998). The Bloom's Syndrome Helicase Unwinds G4 DNA. Journal of Biological Chemistry *273*, 27587–27592.

Sun, H., Treco, D., and Szostak, J.W. (1991). Extensive 3′-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the ARG4 recombination initiation site. Cell *64*, 1155–1161. https://doi.org/10.1016/0092-8674(91)90270-9.

Sun, L., Huang, Y., Edwards, R.A., Yang, S., Blackford, A.N., Niedzwiedz, W., and Glover, J.N.M. (2017). Structural Insight into BLM Recognition by TopBP1. Structure/Folding and Design *25*, 1582–1588.e1583.

Szilard, L. (1959). ON THE NATURE OF THE AGING PROCESS. Proc National Acad Sci *45*, 30–45. https://doi.org/10.1073/pnas.45.1.30.

Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., et al. (2021). Correction to 'The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets'. Nucleic Acids Research *49*, 10800.

Takata, M., Sasaki, M.S., Tachiiri, S., Fukushima, T., Sonoda, E., Schild, D., Thompson, L.H., and Takeda, S. (2001). Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. MCB *21*, 2858–2866.

Takedachi, A., Despras, E., Scaglione, S., Guérois, R., Guervilly, J.H., Blin, M., Audebert, S., Camoin, L., Hasanova, Z., Schertzer, M., et al. (2020). SLX4 interacts with RTEL1 to prevent transcription-mediated DNA replication perturbations. Nature Structural \& Molecular Biology *27*, 438–449.

Tanaka, S., Nakato, R., Katou, Y., Shirahige, K., and Araki, H. (2011). Origin Association of Sld3, Sld7, and Cdc45 Proteins Is a Key Step for Determination of Origin-Firing Timing. Curr Biol *21*, 2055–2063. https://doi.org/10.1016/j.cub.2011.11.038.

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., and Araki, H. (2007). CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. Nature *445*, 328–332. https://doi.org/10.1038/nature05465.

Tatsumi, Y., Sugimoto, N., Yugawa, T., Narisawa-Saito, M., Kiyono, T., and Fujita, M. (2006). Deregulation of Cdt1 induces chromosomal damage without rereplication and leads to chromosomal instability. J Cell Sci *119*, 3128–3140. https://doi.org/10.1242/jcs.03031.

Taylor, E.R., and McGowan, C.H. (2008). Cleavage mechanism of human Mus81-Eme1 acting on Holliday-junction structures. Proceedings of the National Academy of Sciences of the United States of America *105*, 3757–3762.

Teng, F.-Y., Jiang, Z.-Z., Guo, M., Tan, X.-Z., Chen, F., Xi, X.-G., and Xu, Y. (2021). G-quadruplex DNA: a novel target for drug design. Cellular and Molecular Life Sciences : CMLS *78*, 6557–6583.

Tercero, J.A., Longhese, M.P., and Diffley, J.F.X. (2003). A Central Role for DNA Replication Forks in Checkpoint Activation and Response. Mol Cell *11*, 1323–1336. https://doi.org/10.1016/s1097-2765(03)00169-2.

Teyra, J., Kelil, A., Jain, S., Helmy, M., Jajodia, R., Hooda, Y., Gu, J., D'Cruz, A.A., Nicholson, S.E., Min, J., et al. (2020). Large-scale survey and database of high affinity ligands for peptide recognition modules. Mol Syst Biol *16*, e9310.

Thangavel, S., Berti, M., Levikova, M., Pinto, C., Gomathinayagam, S., Vujanovic, M., Zellweger, R., Moore, H., Lee, E.H., Hendrickson, E.A., et al. (2015). DNA2 drives processing and restart of reversed replication forks in human cells. The Journal of Cell Biology *208*, 545–562.

Thorslund, T., McIlwraith, M.J., Compton, S.A., Lekomtsev, S., Petronczki, M., Griffith, J.D., and West, S.C. (2010). The breast cancer tumor suppressor BRCA2 promotes the specific targeting of RAD51 to single-stranded DNA. Nature Structural \& Molecular Biology *17*, 1263–1265.

Thorslund, T., Ripplinger, A., Hoffmann, S., Wild, T., Uckelmann, M., Villumsen, B., Narita, T., Sixma, T.K., Choudhary, C., Bekker-Jensen, S., et al. (2015). Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage. Nature *527*, 389–393.

Toledo, L.I., Altmeyer, M., Rask, M.-B., Lukas, C., Larsen, D.H., Povlsen, L.K., Bekker-Jensen, S., Mailand, N., Bartek, J., and Lukas, J. (2013). ATR prohibits replication catastrophe by preventing global exhaustion of RPA. Cell *155*, 1088–1103. https://doi.org/10.1016/j.cell.2013.10.043.

Toledo, L.I., Murga, M., Zur, R., Soria, R., Rodriguez, A., Martinez, S., Oyarzabal, J., Pastor, J., Bischoff, J.R., and Fernandez-Capetillo, O. (2011). A cell-based screen identifies ATR inhibitors with synthetic lethal properties for cancer-associated mutations. Nat Struct Mol Biol *18*, 721–727. https://doi.org/10.1038/nsmb.2076.

Tomida, J., Takata, K.I., Bhetawal, S., Person, M.D., Chao, H.P., Tang, D.G., and Wood, R.D. (2018). FAM35A associates with REV7 and modulates DNA damage responses of normal and BRCA1-defective cells. The EMBO Journal *37*, e99543–14.

Tompa, P., Davey, N.E., Gibson, T.J., and Babu, M.M. (2014). A Million Peptide Motifs for the Molecular Biologist. Molecular Cell *55*, 161–169.

Toshiyuki, M., and Reed, J.C. (1995). Tumor suppressor p53 is a direct transcriptional activator of the human bax gene. Cell *80*, 293–299. https://doi.org/10.1016/0092-8674(95)90412-3.

Truong, L.N., Li, Y., Shi, L.Z., Hwang, P.Y.-H., He, J., Wang, H., Razavian, N., Berns, M.W., and Wu, X. (2013). Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *110*, 7720–7725.

Tsurimoto, T., and Stillman, B. (1989). Purification of a cellular replication factor, RF-C, that is required for coordinated synthesis of leading and lagging strands during simian virus 40 DNA replication in vitro. Mol Cell Biol *9*, 609–619. https://doi.org/10.1128/mcb.9.2.609-619.1989.

Tsurimoto, T., and Stillman, B. (1990). Functions of replication factor C and proliferating-cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from

human cells and bacteriophage T4. Proc National Acad Sci *87*, 1023–1027. https://doi.org/10.1073/pnas.87.3.1023.

Tsurimoto, T., and Stillman, B. (1991). Replication factors required for SV40 DNA replication in vitro. II. Switching of DNA polymerase alpha and delta during initiation of leading and lagging strand synthesis. J Biological Chem *266*, 1961–1968. .

Tugendreich, S., Tomkiel, J., Earnshaw, W., and Hieter, P. (1995). CDC27Hs colocalizes with CDC16Hs to the centrosome and mitotic spindle and is essential for the metaphase to anaphase transition. Cell *81*, 261–268. .

Turchi, J.J., Huang, L., Murante, R.S., Kim, Y., and Bambara, R.A. (1994). Enzymatic completion of mammalian lagging-strand DNA replication. Proc National Acad Sci *91*, 9803–9807. https://doi.org/10.1073/pnas.91.21.9803.

Ueda, K., Cornwell, M.M., Gottesman, M.M., Pastan, I., Roninson, I.B., Ling, V., and Riordan, J.R. (1986). The mdr1 gene, responsible for multidrug-resistance, codes for P-glycoprotein. Biochemical and Biophysical Research Communications *141*, 956–962. .

Umar, A., Buermeyer, A.B., Simon, J.A., Thomas, D.C., Clark, A.B., Liskay, R.M., and Kunkel, T.A. (1996). Requirement for PCNA in DNA Mismatch Repair at a Step Preceding DNA Resynthesis. Cell *87*, 65–73. https://doi.org/10.1016/s0092-8674(00)81323-9.

Uyar, B., Weatheritt, R.J., Dinkel, H., Davey, N.E., and Gibson, T.J. (2014). Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer? Mol Biosyst *10*, 2626–2642.

Uziel, T., Lerenthal, Y., Moyal, L., Andegeko, Y., Mittelman, L., and Shiloh, Y. (2003). Requirement of the MRN complex for ATM activation by DNA damage. The EMBO Journal *22*, 5612–5621. .

Van Maldergem, L., Siitonen, H.A., Jalkh, N., Chouery, E., De Roy, M., Delague, V., Muenke, M., Jabs, E.W., Cai, J., Wang, L.L., et al. (2006). Revisiting the craniosynostosis-radial ray hypoplasia association: Baller-Gerold syndrome caused by mutations in the RECQL4 gene. J Med Genet *43*, 148–152.

Van Roey, K., and Davey, N.E. (2015). Motif co-regulation and co-operativity are common mechanisms in transcriptional, post-transcriptional and post-translational regulation. Cell Commun Signal *13*, 45–16.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research *50*, D439–D444.

Vaziri, C., Saxena, S., Jeon, Y., Lee, C., Murata, K., Machida, Y., Wagle, N., Hwang, D.S., and Dutta, A. (2003). A p53-Dependent Checkpoint Pathway Prevents Rereplication. Mol Cell *11*, 997–1008. https://doi.org/10.1016/s1097-2765(03)00099-6.

Virgilio, M.D., Callen, E., Gomez, G., Wu, J., Lawhon, S.D., Popova, I.E., Parikh, S.J., Adams, L.G., Tsolis, R.M., Stewart, V.J., et al. (2013). Rif1 Prevents Resection of DNA Breaks and Promotes Immunoglobulin Class Switching. Science *339*, 708–711.

Vugt, M.A.T.M. van, Gardino, A.K., Linding, R., Ostheimer, G.J., Reinhardt, H.C., Ong, S.-E., Tan, C.S., Miao, H., Keezer, S.M., Li, J., et al. (2010). A Mitotic Phosphorylation Feedback Network Connects Cdk1, Plk1, 53BP1, and Chk2 to Inactivate the G2/M DNA Damage Checkpoint. Plos Biol *8*, e1000287. https://doi.org/10.1371/journal.pbio.1000287.

Waga, S., and Stillman, B. (1994). Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication in vitro. Nature *369*, 207–212. https://doi.org/10.1038/369207a0.

Waldman, A.D., Fritz, J.M., and Lenardo, M.J. (2020). A guide to cancer immunotherapy: from T cell basic science to clinical practice. Nat Rev Immunol *20*, 651–668.

Walther, A.P., Gomes, X.V., Lao, Y., Lee, C.G., and Wold, M.S. (1999). Replication protein A interactions with DNA. 1. Functions of the DNA-binding and zinc-finger domains of the 70-kDa subunit. Biochemistry *38*, 3963–3973.

Wang, A.C., Wilson, S.K., Hopper, J.E., Fudenberg, H.H., and Nisonoff, A. (1970). Evidence for Control of Synthesis of the Variable Regions of the Heavy Chains of Immunoglobulins G and M by the Same Gene*. Proc National Acad Sci *66*, 337–343. https://doi.org/10.1073/pnas.66.2.337.

Wang, B., Matsuoka, S., Ballif, B.A., Zhang, D., Smogorzewska, A., Gygi, S.P., and Elledge, S.J. (2007). Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. Science *316*, 1194–1198.

Wang, F., Yang, Y., Singh, T.R., Busygina, V., Guo, R., Wan, K., Wang, W., Sung, P., Meetei, A.R., and Lei, M. (2010). Crystal structures of RMI1 and RMI2, two OB-fold regulatory subunits of the BLM complex. Structure (London, England : 1993) *18*, 1159–1170.

Wang, H., Li, Y., Truong, L.N., Shi, L.Z., Hwang, P.Y.-H., He, J., Do, J., Cho, M.J., Li, H., Negrete, A., et al. (2014). CtIP Maintains Stability at Common Fragile Sites and Inverted Repeats by End Resection-Independent Endonuclease Activity. Mol Cell *54*, 1012–1021. https://doi.org/10.1016/j.molcel.2014.04.012.

Wang, J., Okkeri, J., Pavic, K., Wang, Z., Kauko, O., Halonen, T., Sarek, G., Ojala, P.M., Rao, Z., Xu, W., et al. (2017). Oncoprotein CIP2A is stabilized via interaction with tumor suppressor PP2A/B56. EMBO Reports *18*, 437–450.

Wang, T.C., and Smith, K.C. (1986). Postreplication repair in ultraviolet-irradiated human fibroblasts: formation and repair of DNA double-strand breaks. Carcinogenesis *7*, 389–392. .

Wang, W., Seki, M., Narita, Y., Sonoda, E., Takeda, S., Yamada, K., Masuko, T., Katada, T., and Enomoto, T. (2000). Possible association of BLM in decreasing DNA double strand breaks during DNA replication. The EMBO Journal *19*, 3428–3435.

Wang, Y., Gao, W., Shi, X., Ding, J., Liu, W., He, H., Wang, K., and Shao, F. (2017). Chemotherapy drugs induce pyroptosis through caspase-3 cleavage of a gasdermin. Nature *547*, 99–103. .

Wardlaw, C.P., Carr, A.M., and Oliver, A.W. (2014). TopBP1: A BRCT-scaffold protein functioning in multiple cellular pathways. DNA Repair *22*, 165–174.

Watanabe, N., Arai, H., Nishihara, Y., Taniguchi, M., Watanabe, N., Hunter, T., and Osada, H. (2004). M-phase kinases induce phospho-dependent ubiquitination of somatic Wee1 by SCFbeta-TrCP. P Natl Acad Sci Usa *101*, 4419–4424. https://doi.org/10.1073/pnas.0307700101.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Wechsler, T., Newman, S., and West, S.C. (2011). Aberrant chromosome morphology in human cells defective for Holliday junction resolution. Nature *471*, 642–646.

Weigert, M., Gatmaitan, L., Loh, E., Schilling, J., and Hood, L. (1978). Rearrangement of genetic information may produce immunoglobulin diversity. Nature *276*, 785–790. https://doi.org/10.1038/276785a0.

Weissman, L., Souza-Pinto, N.C. de, Stevnsner, T., and Bohr, V.A. (2007). DNA repair, mitochondria, and neurodegeneration. Neuroscience *145*, 1318–1329.

White, C.I., and Haber, J.E. (1990). Intermediates of recombination during mating type switching in Saccharomyces cerevisiae. Embo J *9*, 663–673. https://doi.org/10.1002/j.1460-2075.1990.tb08158.x.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331–338.

Wiese, C., Dray, E., Groesser, T., Filippo, J.S., Shi, I., Collins, D.W., Tsai, M.-S., Williams, G.J., Rydberg, B., Sung, P., et al. (2007). Promotion of homologous recombination and genomic stability by RAD51AP1 via RAD51 recombinase enhancement. Molecular Cell *28*, 482–490.

Wietmarschen, N., Merzouk, S., Halsema, N., Spierings, D.C.J., Guryev, V., and Lansdorp, P.M. (2018). BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes. Nature Communications 1–12.

Wilson, M.A., Kwon, Y., Xu, Y., Chung, W.-H., Chi, P., Niu, H., Mayle, R., Chen, X., Malkova, A., Sung, P., et al. (2013). Pif1 helicase and Polδ promote recombination-coupled DNA synthesis via bubble migration. Nature *502*, 393–396.

Wong, A.K., Pero, R., Ormonde, P.A., Tavtigian, S.V., and Bartel, P.L. (1997). RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene brca2. Journal of Biological Chemistry *272*, 31941–31944.

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., Micklem, G., et al. (1995). Identification of the breast cancer susceptibility gene BRCA2. Nature *378*, 789–792. https://doi.org/10.1038/378789a0.

Wright, W.D., Shah, S.S., and Heyer, W.-D. (2018). Homologous recombination and the repair of DNA double-strand breaks. J. Biol. Chem. *293*, 10524–10535.

Wu, C.-C., Li, T.-K., Farh, L., Lin, L.-Y., Lin, T.-S., Yu, Y.-J., Yen, T.-J., Chiang, C.-W., and Chan, N.-L. (2011). Structural basis of type II topoisomerase inhibition by the anticancer drug etoposide. Science *333*, 459–462.

Wu, L., and Hickson, I.D. (2003). The Bloom's syndrome helicase suppresses crossing over during homologous recombination. Nature *426*, 870–874.

Wu, L., Bachrati, C.Z., Ou, J., Xu, C., Yin, J., Chang, M., Wang, W., Li, L., Brown, G.W., and Hickson, I.D. (2006). BLAP75/RMI1 promotes the BLM-dependent dissolution of homologous recombination intermediates. Proceedings of the National Academy of Sciences *103*, 4068–4073.

Wuite, G.J.L., Smith, S.B., Young, M., Keller, D., and Bustamante, C. (2000). Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. Nature *404*, 103–106. https://doi.org/10.1038/35003614.

Wyatt, H.D.M., Sarbajna, S., Matos, J., and West, S.C. (2013). Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells. Molecular Cell *52*, 234–247.

Wyman, C., and Botchan, M. (1995). DNA Replication: A familiar ring to DNA polymerase processivity. Curr Biol *5*, 334–337. https://doi.org/10.1016/s0960-9822(95)00065-0.

Xu, D., Guo, R., Sobeck, A., Bachrati, C.Z., Yang, J., Enomoto, T., Brown, G.W., Hoatlin, M.E., Hickson, I.D., and Wang, W. (2008a). RMI, a new OB-fold complex essential for Bloom syndrome protein to maintain genome stability. Genes \& Development *22*, 2843–2855.

Xu, G., Chapman, J.R., Brandsma, I., Yuan, J., Mistrik, M., Bouwman, P., Bartkova, J., Gogola, E., Warmerdam, D., Barazas, M., et al. (2015). REV7 counteracts DNA double-strand break resection and affects PARP inhibition. Nature *521*, 541–544.

Xu, X., Vaithiyalingam, S., Glick, G.G., Mordes, D.A., Chazin, W.J., and Cortez, D. (2008b). The basic cleft of RPA70N binds multiple checkpoint proteins, including RAD9, to regulate ATR signaling. Molecular and Cellular Biology *28*, 7345–7353.

Xue, X., Raynard, S., Busygina, V., Singh, A.K., and Sung, P. (2013). Role of replication protein A in double holliday junction dissolution mediated by the BLM-Topo IIIα-RMI1-RMI2 protein complex. J. Biol. Chem. *288*, 14221–14227.

Yabuuchi, H., Yamada, Y., Uchida, T., Sunathvanichkul, T., Nakagawa, T., and Masukata, H. (2006). Ordered assembly of Sld3, GINS and Cdc45 is distinctly regulated by DDK and CDK for activation of replication origins. Embo J *25*, 4663–4674. https://doi.org/10.1038/sj.emboj.7601347.

Yaffe, M.B., Leparc, G.G., Lai, J., Obata, T., Volinia, S., and Cantley, L.C. (2001). A motif-based profile scanning approach for genome-wide prediction of signaling pathways. Nature Biotechnology *19*, 348–353.

Yamamoto, O., Fuji, I., Yoshida, T., Cox, A.B., and Lett, J.T. (1988). Age Dependency of Base Modification in Rabbit Liver DNA. J Gerontology *43*, B132–B136. https://doi.org/10.1093/geronj/43.5.b132.

Yan, S., and Michael, W.M. (2009). TopBP1 and DNA polymerase-α directly recruit the 9-1-1 complex to stalled DNA replication forks. J Cell Biology *184*, 793–804. https://doi.org/10.1083/jcb.200810185.

Yan, Z., Delannoy, M., Ling, C., Daee, D., Osman, F., Muniandy, P.A., Shen, X., Oostra, A.B., Du, H., Steltenpool, J., et al. (2010). A Histone-Fold Complex and FANCM Form a Conserved DNA-Remodeling Complex to Maintain Genome Stability. Mol Cell *37*, 865–878. https://doi.org/10.1016/j.molcel.2010.01.039.

Yang, H., Jeffrey, P.D., Miller, J., Kinnucan, E., Sun, Y., Thoma, N.H., Zheng, N., Chen, P.-L., Lee, W.-H., and Pavletich, N.P. (2002). BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. Science *297*, 1837–1848.

Yang, J., O'Donnell, L., Durocher, D., and Brown, G.W. (2012). RMI1 promotes DNA replication fork progression and recovery from replication fork stress. MCB *32*, 3054–3064.

Ye, J.Z.-S., Donigian, J.R., Overbeek, M. van, Loayza, D., Luo, Y., Krutchinsky, A.N., Chait, B.T., and Lange, T. de (2004). TIN2 Binds TRF1 and TRF2 Simultaneously and Stabilizes the TRF2 Complex on Telomeres*. J Biol Chem *279*, 47264–47271. https://doi.org/10.1074/jbc.m409047200.

Yeeles, J.T.P., Deegan, T.D., Janska, A., Early, A., and Diffley, J.F.X. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. Nature *519*, 431–435. https://doi.org/10.1038/nature14285.

Yeom, G., Kim, J., and Park, C.-J. (2019). Investigation of the core binding regions of human Werner syndrome and Fanconi anemia group J helicases on replication protein A. Scientific Reports *9*, 14016–10.

Yu, C.-E., Oshima, J., Fu, Y.-H., Wijsman, E.M., Hisama, F., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S., et al. (1996). Positional Cloning of the Werner's Syndrome Gene. Science *272*, 258–262. https://doi.org/10.1126/science.272.5259.258.

Yu, J., Zhang, L., Hwang, P.M., Kinzler, K.W., and Vogelstein, B. (2001). PUMA Induces the Rapid Apoptosis of Colorectal Cancer Cells. Mol Cell *7*, 673–682. https://doi.org/10.1016/s1097-2765(01)00213-1.

Yu, X., and Chen, J. (2004). DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains. Molecular and Cellular Biology *24*, 9478–9486.

Yusufzai, T., and Kadonaga, J.T. (2008). HARP is an ATP-driven annealing helicase. Science *322*, 748–750.

Yusufzai, T., and Kadonaga, J.T. (2010). Annealing helicase 2 (AH2), a DNA-rewinding motor with an HNH motif. Proceedings of the National Academy of Sciences of the United States of America *107*, 20970–20973.

Zakharyevich, K., Ma, Y., Tang, S., Hwang, P.Y.-H., Boiteux, S., and Hunter, N. (2010). Temporally and Biochemically Distinct Activities of Exo1 during Meiosis: Double-Strand Break Resection and Resolution of Double Holliday Junctions. Mol Cell *40*, 1001–1015. https://doi.org/10.1016/j.molcel.2010.11.032.

Zegerman, P., and Diffley, J.F.X. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature *445*, 281–285. https://doi.org/10.1038/nature05432.

Zellweger, R., Dalcher, D., Mutreja, K., Berti, M., Schmid, J.A., Herrador, R., Vindigni, A., and Lopes, M. (2015). Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. The Journal of Cell Biology *208*, 563–579.

Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. Nature Cell Biology *16*, 2–9.

Zhang, H., Kobayashi, R., Galaktionov, K., and Beach, D. (1995). p19Skp1 and p45Skp2 are essential elements of the cyclin A-CDK2 S phase kinase. Cell *82*, 915–925. .

Zhang, J. (2000). Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. Journal of Molecular Evolution *50*, 56–68.

Zhang, J., Dewar, J.M., Budzowska, M., Motnenko, A., Cohn, M.A., and Walter, J.C. (2015). DNA interstrand cross-link repair requires replication-fork convergence. Nat Struct Mol Biol *22*, 242–247. https://doi.org/10.1038/nsmb.2956.

Zhao, B., Rothenberg, E., Ramsden, D.A., and Lieber, M.R. (2020). The molecular basis and disease relevance of non-homologous DNA end joining. Nature Reviews Molecular Cell Biology *21*, 765–781.

Zhou, C., Pourmal, S., and Pavletich, N.P. (2015). Dna2 nuclease-helicase structure, mechanism and regulation by Rpa. eLife *4*, 213.

Zhu, J., Zhu, S., Guzzo, C.M., Ellis, N.A., Sung, K.S., Choi, C.Y., and Matunis, M.J. (2008). Small ubiquitin-related modifier (SUMO) binding determines substrate recognition and paralog-selective SUMO modification. J. Biol. Chem. *283*, 29405–29415.

Zhu, Z., Chung, W.-H., Shim, E.Y., Lee, S.E., and Ira, G. (2008). Sgs1 Helicase and Two Nucleases Dna2 and Exo1 Resect DNA Double-Strand Break Ends. Cell *134*, 981–994. https://doi.org/10.1016/j.cell.2008.08.037.

Zimmermann, M., Lottersberger, F., Buonomo, S.B., Sfeir, A., and Lange, T. de (2013). 53BP1 regulates DSB repair using Rif1 to control 5' end resection. Science *339*, 700–704.

Zirkle, R.E., and Bloom, W. (1953). Irradiation of Parts of Individual Cells. Science *117*, 487–493. https://doi.org/10.1126/science.117.3045.487.

Zong, D., Adam, S., Wang, Y., Sasanuma, H., CallEn, E., Murga, M., Day, A., Kruhlak, M.J., Wong, N., Munro, M., et al. (2019). BRCA1 Haploinsufficiency Is Masked by RNF168-Mediated Chromatin Ubiquitylation. Molecular Cell *73*, 1267–1281.e1267.

Zou, L., and Elledge, S.J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science *300*, 1542–1548.

Zou, L., Cortez, D., and Elledge, S.J. (2002). Regulation of ATR substrate selection by Rad17-dependent loading of Rad9 complexes onto chromatin. Gene Dev *16*, 198–208. https://doi.org/10.1101/gad.950302.

Zou, L., Liu, D., and Elledge, S.J. (2003). Replication protein A-mediated recruitment and activation of Rad17 complexes. Proc National Acad Sci *100*, 13827–13832. https://doi.org/10.1073/pnas.2336100100.

Chapter 8: Appendices

8.1 Manual Sequence Alignments Used to Predict SLiMs

Sequence alignments of orthologues from representative vertebrates. Underlined region in red corresponds to region synthesized as biotinylated peptide. Sequence alignments can be found at https://data.mendeley.com/datasets/7cfzh9p2ct/1.

8.2 Silver Stain Gels of Peptide Pulldowns for Mass Spectrometry

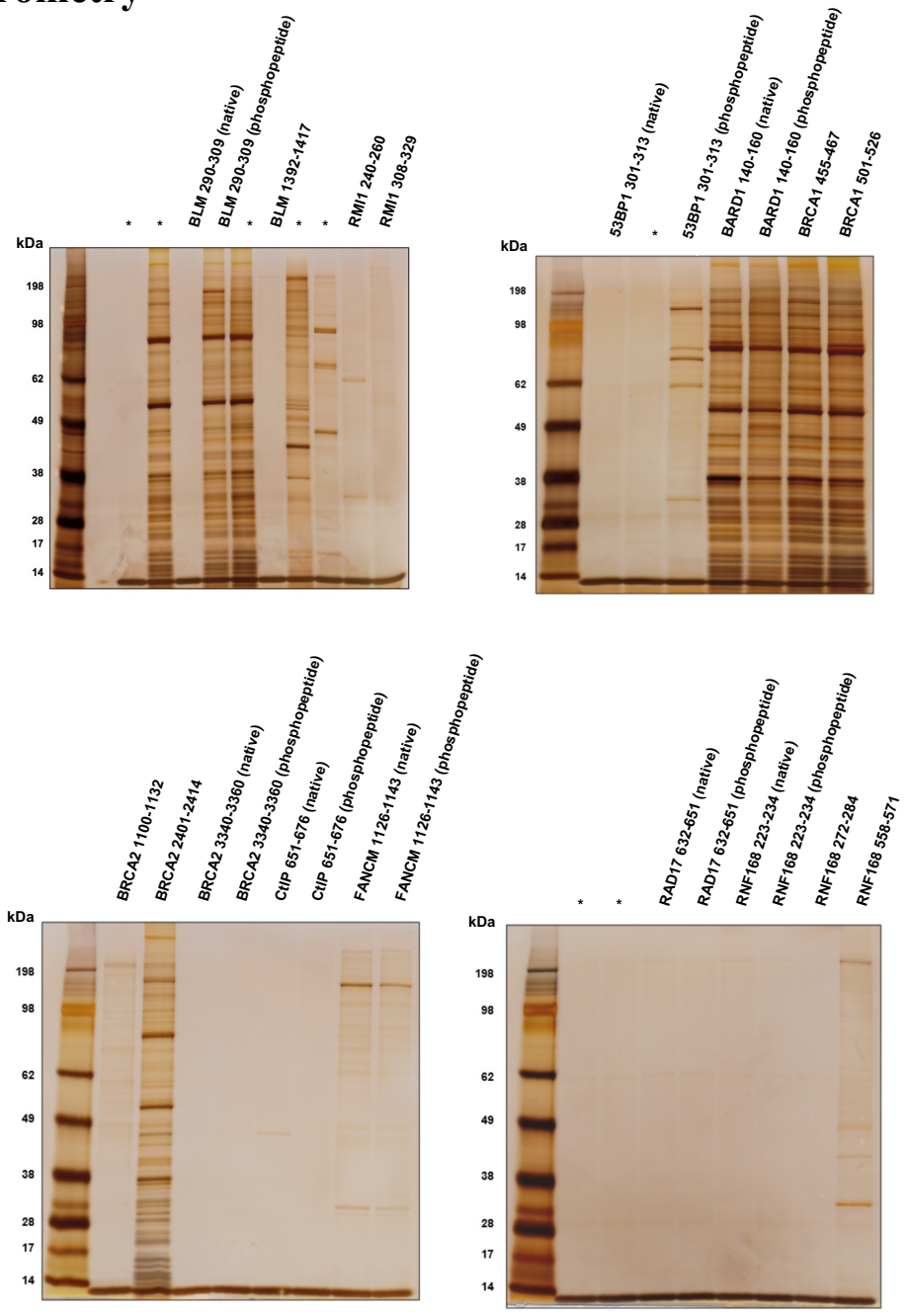


Fig. 8.2.1. Silver stain gels of peptide pulldowns conducted in Chapter 3.
One repeat shown. *, peptide pulldowns outside of the work in this thesis.

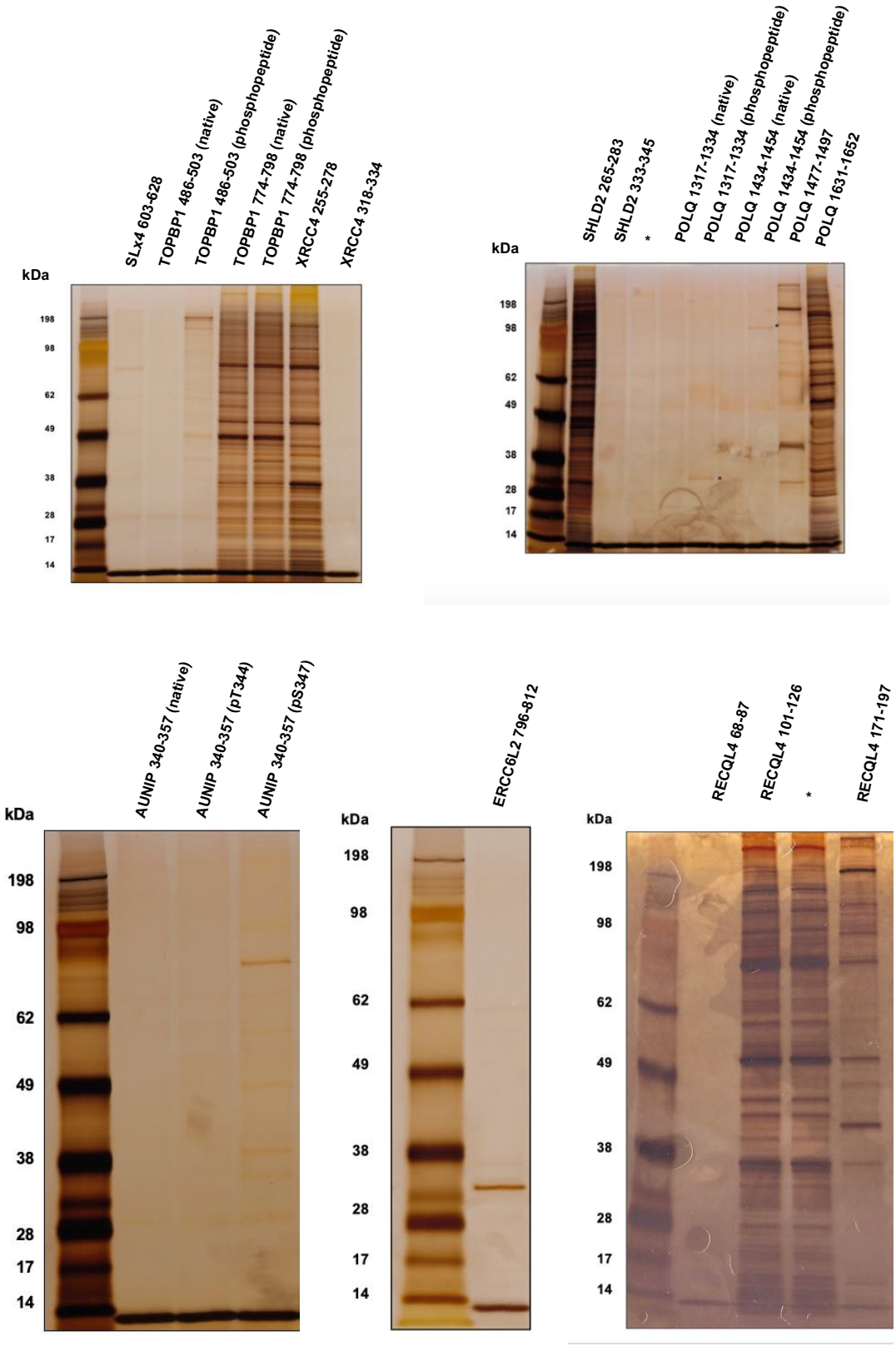


Fig. 8.2.1. Silver stain gels of peptide pulldowns conducted in Chapter 3.
One repeat shown. *, peptide pulldowns outside of work in this thesis.

8.3 Proteins Assigned to the Background Proteome

Table 8.3.1. Proteins assigned to the background proteome

ABCF1	ACACA	ACACB	ACIN1	ACLY	ACTA1	ACTA2	ACTB	ACTBL2	ACTC1
ACTG1	ACTG2	ACTN1	ACTN2	ACTN4	ACTR1A	ADAR	ADRM1	AFF4	AHCY
AHNAK	AHSA1	AIFM1	ALB	ALBU	ALDH18A1	ALDOA	ALYREF	ANXA2	ANXA2P2
AP3D1	ARCN1	ARF1	ARF3	ARF4	ARF5	ARGLU1	ARL6IP4	ATAD3A	ATP1A1
ATP1A2	ATP1A3	ATP2A2	ATP4A	ATP5A1	ATP5B	ATP5C1	ATP5O	ATXN2L	B4DLN1
BAG2	BCLAF1	BIRA	BOLA2	BOLA2B	BUB3	C14ORF166	C19ORF43	C1QBP	CACYBP
CAD	CALM1	CALM2	CALM3	CALR	CAND1	CANX	CAPRIN1	CAPZA1	CAPZA2
CAPZB	CAS1	CASP14	CBX1	CBX3	CCAR2	CCDC124	CCT2	CCT3	CCT4
CCT5	CCT6A	CCT6B	CCT7	CCT8	CD3EAP	CDC2	CDC37	CDC5L	CDK1
CDK12	CDK13	CDK2	CDK3	CDK9	CFL1	CFL2	CHD3	CHD4	CHD5
CHERP	CIR1	CIRBP	CKAP5	CKB	CLASP1	CLASP2	CLINT1	CLNS1A	CLTC
CLTCL1	CNN3	COPA	COPB1	COPG1	COPG2	CORO1C	CP	CPSF6	CPSF7
CRKL	CS	CSDE1	CSE1L	CSRP2	CSTF2	CTNND1	CTPS1	CTTN	CUL4B
CYFIP1	CYFIP2	DARS	DBN1	DCD	DCTN1	DCTN2	DDB1	DDX1	DDX17
DDX18	DDX21	DDX39A	DDX39B	DDX3X	DDX3Y	DDX4	DDX42	DDX46	DDX5
DDX50	DDX6	DEK	DES	DHE3	DHX15	DHX30	DHX9	DKFZP686E0752	DKFZP686J1372
DKFZP686L22104	DLD	DNAJA1	DNAJA2	DPYSL2	DSP	DUT	DYNC1H1	DYNLL1	E1B
EDF1	EEF1A1	EEF1A1P5	EEF1A2	EEF1B2	EEF1D	EEF1E1	EEF1G	EEF2	EFTUD2
EIF2S1	EIF2S2	EIF2S3	EIF2S3B	EIF2S3L	EIF3A	EIF3B	EIF3C	EIF3CL	EIF3D
EIF3E	EIF3F	EIF3G	EIF3H	EIF3I	EIF3L	EIF3S3	EIF4A1	EIF4A2	EIF4A3
EIF4B	EIF4G1	EIF4H	EIF5A	EIF5A2	EIF5AL1	EIF5B	EIF6	ELAVL1	EMD
ENO1	ENO2	ENO3	EPB41L2	EPPK1	EPRS	ERH	EWSR1	EXOSC10	EZR

F5H423	F8W810	FABP5	FASN	FAU	FBL	FEN1	FIP1L1	FKBP4	FLG2
FLNA	FLNB	FLNC	FMC1	FUBP1	FUBP3	FUS	G3BP1	G3BP2	GANAB
GAPDH	GART	GCN1	GFAP	GFP	GIGYF2	GLUD1	GLUD2	GMPS	GNAS
GNL3	GTF2I	H0YHG0	H1FX	H2AFJ	H2AFV	H2AFX	H2AFZ	H2BFS	H3BN98
H3BTW0	H3BVE0	H3F3A	H3F3B	H3F3C	HCFC1	HDAC1	HDAC2	HDLBP	HIST1H1A
HIST1H1B	HIST1H1C	HIST1H1D	HIST1H1E	HIST1H1T	HIST1H2AA	HIST1H2AB	HIST1H2AC	HIST1H2AD	HIST1H2AE
HIST1H2AG	HIST1H2AH	HIST1H2AI	HIST1H2AJ	HIST1H2AK	HIST1H2AL	HIST1H2AM	HIST1H2BA	HIST1H2BB	HIST1H2BC
HIST1H2BD	HIST1H2BE	HIST1H2BF	HIST1H2BG	HIST1H2BH	HIST1H2BI	HIST1H2BJ	HIST1H2BK	HIST1H2BL	HIST1H2BM
HIST1H2BN	HIST1H2BO	HIST1H3A	HIST1H3B	HIST1H3C	HIST1H3D	HIST1H3E	HIST1H3F	HIST1H3G	HIST1H3H
HIST1H3I	HIST1H3J	HIST1H4A	HIST1H4B	HIST1H4C	HIST1H4D	HIST1H4E	HIST1H4F	HIST1H4H	HIST1H4I
HIST1H4J	HIST1H4K	HIST1H4L	HIST2H2AA3	HIST2H2AA4	HIST2H2AB	HIST2H2AC	HIST2H2BC	HIST2H2BD	HIST2H2BE
HIST2H2BF	HIST2H3A	HIST2H3C	HIST2H3D	HIST2H3PS2	HIST2H4A	HIST2H4B	HIST3H2A	HIST3H2BB	HIST3H3
HIST4H4	HLTF	HMGB1	HMGB1P1	HMGN2	HNRNPA0	HNRNPA1	HNRNPA1L2	HNRNPA2B1	HNRNPA3
HNRNPAB	HNRNPC	HNRNPCL1	HNRNPCL2	HNRNPCL3	HNRNPCL4	HNRNPD	HNRNPDL	HNRNPF	HNRNPH1
HNRNPH2	HNRNPH3	HNRNPK	HNRNPL	HNRNPM	HNRNPR	HNRNPU	HNRNPUL1	HNRNPUL2	HP1BP3
HPRT1	HRNR	HSD17B10	HSP90AA1	HSP90AA2P	HSP90AA4P	HSP90AA5P	HSP90AB1	HSP90AB2P	HSP90AB3P
HSP90AB4P	HSP90B1	HSP90B2P	HSPA1A	HSPA1B	HSPA1L	HSPA2	HSPA4	HSPA5	HSPA6
HSPA7	HSPA8	HSPA9	HSPB1	HSPD1	HSPH1	HUWE1	IARS	IGF2BP1	IGF2BP2
IGF2BP3	IGKV2-28	IGKV2-29	IGKV2-30	IGKV2-40	IGKV2D-26	IGKV2D-28	IGKV2D-29	IGKV2D-30	IGKV2D-40
ILF2	ILF3	IMPDH2	INA	IPO5	IPO7	IQGAP1	IRS4	J3QQQ9	JPT2
JUP	K1C10	K1C15	K1C9	K1H1	K1H2	K1H5	K1H6	K1H7	K1H8
K1HB	K1M1	K22E	K2C1	KARS	KHDRBS1	KHSRP	KIF11	KIF5A	KIF5B
KIF5C	KPNA2	KPNB1	KRHB4	KRT1	KRT10	KRT13	KRT14	KRT15	KRT16
KRT17	KRT18	KRT19	KRT2	KRT20	KRT24	KRT25	KRT26	KRT27	KRT28
KRT3	KRT31	KRT32	KRT33B	KRT35	KRT36	KRT37	KRT38	KRT4	KRT5
KRT6A	KRT6B	KRT6C	KRT7	KRT72	KRT73	KRT74	KRT75	KRT76	KRT77
KRT78	KRT79	KRT8	KRT80	KRT84	KRT9	KRT9	LARP1	LARS	LBR

LDHA	LDHB	LIMA1	LMNA	LMNB1	LRPPRC	LSM12	LSM14A	LUC7L	LUC7L2
LUC7L3	M0QYT0	M0R076	M0R2N4	MAP1B	MAP4	MAP7	MARCKS	MARS	MAT2A
MATR3	MCCC1	MCCC2	MCM3	MCM4	MCM5	MCM6	MCM7	MDH2	MFAP1
MIF	MKI67	MMTAG2	MRE11	MSH2	MSH6	MSN	MTA1	MTA2	MTCL1
MTHFD1	MYBBP1A	MYH10	MYH11	MYH14	MYH9	MYL12A	MYL12B	MYL6	MYL6B
MYL9	NACA	NAP1L1	NAP1L4	NASP	NAT10	NBN	NCAPH	NCL	NCOR1
NEFH	NEFL	NEFM	NME1	NME2	NME2P1	NOLC1	NONO	NOP2	NOP56
NOP58	NPM1	NPM3	NSUN2	NUDC	NUDT21	NUFIP2	NUMA1	NUP153	NUP214
NUP93	NXF1	OAT	P03243	P4HB	PA2G4	PABPC1	PABPC1L	PABPC1L2A	PABPC1L2B
PABPC3	PABPC4	PABPN1	PAICS	PARK7	PARP1	PC	PCBP1	PCBP2	PCBP3
PCCA	PCCB	PCMT1	PCNA	PDAP1	PDCD5	PDCD6IP	PDIA3	PDIA6	PFKP
PFN1	PGAM1	PGAM2	PGAM4	PGAM5	PGK1	PHB	PHB2	PHF3	PHGDH
PKLR	PKM	PLEC	PLS3	POLDIP3	POLR1A	POLR1C	POLR2E	POLR2H	POTEE
POTEF	POTEI	POTEJ	POTEKP	PPA1	PPIA	PPIB	PPIL4	PPP1CA	PPP1CB
PPP1CC	PPP2CA	PPP2CB	PPP2R1A	PPP2R2A	PPP2R2B	PPP2R2D	PRDX1	PRDX2	PRDX3
PRDX4	PRDX6	PRKDC	PRMT1	PRMT5	PRPF19	PRPF3	PRPF31	PRPF4	PRPF40A
PRPF6	PRPF8	PRPH	PRPS1	PRPS1L1	PRPS2	PRRC2A	PRRC2C	PRSS1	PSIP1
PSMA2	PSMA4	PSMA5	PSMA6	PSMA7	PSMB5	PSMB6	PSMC1	PSMC2	PSMC3
PSMC4	PSMC5	PSMD1	PSMD11	PSMD12	PSMD2	PSMD3	PSMD4	PSMD6	PSME3
PSPC1	PTBP1	PTMA	PUF60	QARS	RAB1A	RAB1B	RACK1	RAD50	RAN
RANBP1	RANBP2	RANGAP1	RARS	RAVER1	RBBP4	RBBP7	RBM10	RBM14	RBM17
RBM25	RBM26	RBM39	RBMX	RBMXL1	RBMXL2	RBMXL3	RCC2	RCN1	RCN2
RDX	RFC1	RFC2	RFC3	RFC4	RGPD3	RGPD4	RIF1	RNPS1	RPL10
RPL10	RPL10A	RPL10L	RPL11	RPL12	RPL13	RPL13A	RPL13AP3	RPL14	RPL15
RPL17	RPL18	RPL18A	RPL19	RPL21	RPL22	RPL23	RPL23A	RPL24	RPL26
RPL26L1	RPL27	RPL27A	RPL28	RPL29	RPL3	RPL30	RPL31	RPL34	RPL35
RPL35A	RPL36	RPL36A	RPL36AL	RPL37A	RPL38	RPL4	RPL5	RPL6	RPL7

RPL7A	RPL8	RPL9	RPL9P7	RPL9P8	RPL9P9	RPLP0	RPLP0P6	RPLP1	RPLP2
RPN1	RPS10	RPS10P5	RPS11	RPS12	RPS13	RPS14	RPS15	RPS15A	RPS16
RPS17	RPS18	RPS19	RPS2	RPS20	RPS21	RPS23	RPS24	RPS25	RPS26
RPS26P11	RPS27	RPS27A	RPS27L	RPS28	RPS29	RPS3	RPS3A	RPS4X	RPS4Y1
RPS4Y2	RPS5	RPS6	RPS7	RPS8	RPS9	RPSA	RRP15	RRP1B	RS27A
RSL1D1	RTCB	RUVBL1	RUVBL2	SAFB	SARNP	SART1	SART3	SCAF11	SEC13
SEPT7	SERBP1	SERPINH1	SET	SETP2	SETSIP	SF1	SF3A1	SF3A3	SF3B1
SF3B2	SF3B3	SF3B4	SF3B6	SFN	SFPQ	SHMT2	SKIV2L2	SKP1	SLC1A5
SLC25A11	SLC25A3	SLC25A31	SLC25A4	SLC25A5	SLC25A6	SLC3A2	SMARCA5	SMARCC1	SMC1A
SMC2	SMC3	SMC4	SND1	SNRNP200	SNRNP70	SNRPA	SNRPA1	SNRPB	SNRPB2
SNRPC	SNRPD1	SNRPD2	SNRPD3	SNRPE	SNRPF	SNRPN	SNU13	SNW1	SON
SPTAN1	SPTBN1	SRP14	SRP68	SRP9	SRRM1	SRRM2	SRSF1	SRSF11	SRSF2
SRSF3	SRSF4	SRSF5	SRSF6	SRSF7	SRSF8	SSB	SSBP1	SSR4	SSRP1
STAU1	STIP1	STK38	STMN1	STMN2	STRAP	SUB1	SUGT1	SUMO2	SUMO3
SUMO4	SUPT16H	SYNCRIP	TAF15	TAGLN2	TARDBP	TBL1XR1	TCEAL4	TCOF1	TCP1
TECR	THIO	THOC2	THRAP3	TIMM50	TJP2	TLN1	TMPO	TOP1	TOP2A
TOP2B	TPI1	TPM1	TPM2	TPM3	TPM4	TPR	TPX2	TRAP1	TRIM21
TRIM28	TRYP	TUBA1A	TUBA1B	TUBA1C	TUBA3C	TUBA3D	TUBA3E	TUBA4A	TUBA4B
TUBA8	TUBAL3	TUBB	TUBB1	TUBB2A	TUBB2B	TUBB3	TUBB4A	TUBB4B	TUBB6
TUBB8	TUFM	TXN	TXNDC12	U2AF1	U2AF1L4	U2AF1L5	U2AF2	U2SURP	UBA1
UBA52	UBAP2L	UBB	UBC	UBTF	UQCRC2	USP7	VARS	VCP	VDAC1
VDAC2	VDAC3	VIM	WARS	WDR77	XPO1	XRCC5	XRCC6	XRN2	YBX1
YBX2	YBX3	YTHDF2	YWHAB	YWHAE	YWHAG	YWHAH	YWHAQ	YWHAZ	YY1
ZC3HAV1	ZFR	ZGPAT	ZNF326	ZNF787	ZNF90				

8.4. Full STRING Networks

53BP1 301-313 (phosphopeptide)

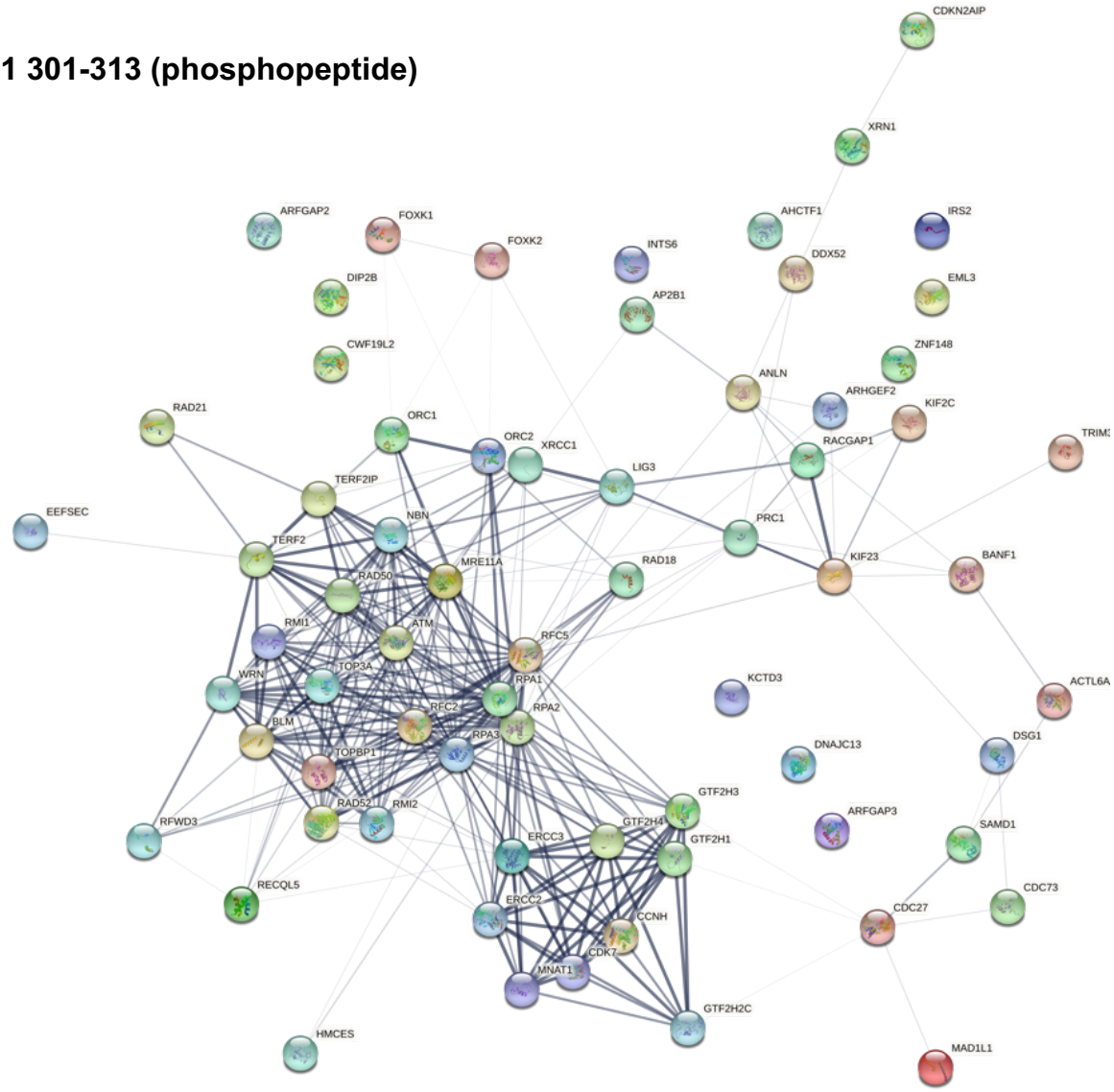
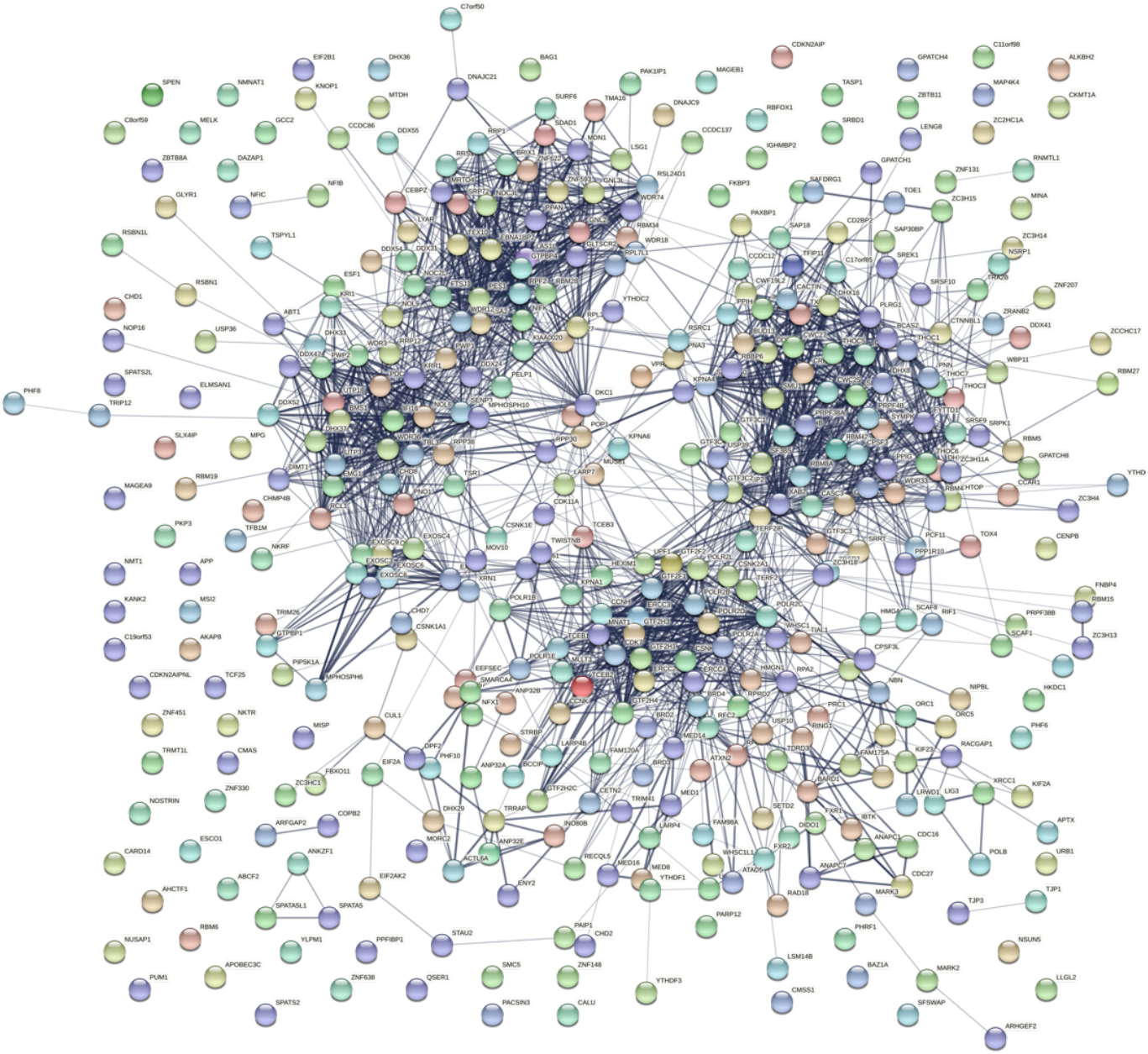
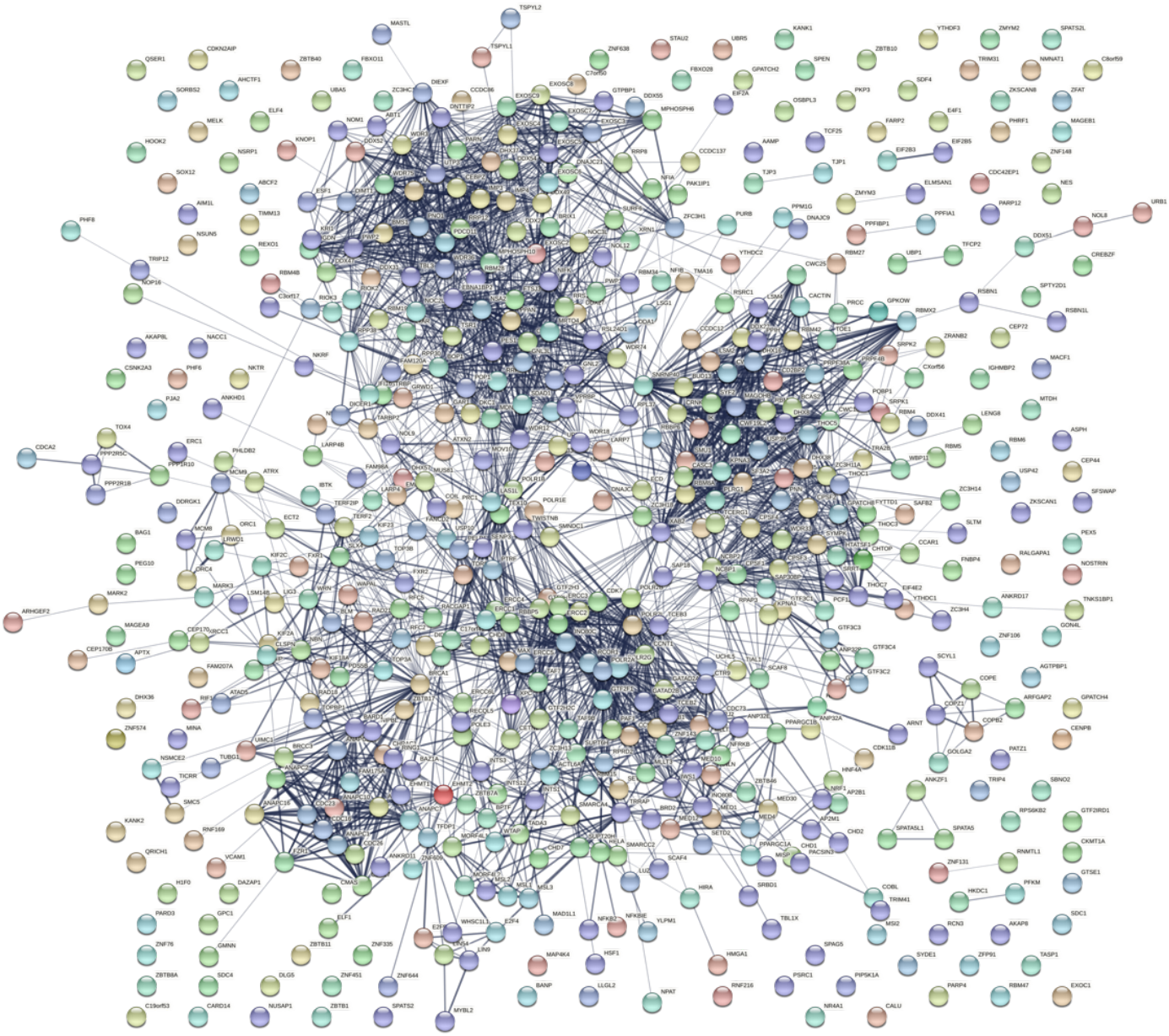


Fig. 8.4.1. Full STRING networks of interacting proteins peptide pulldowns conducted in Chapter 3.

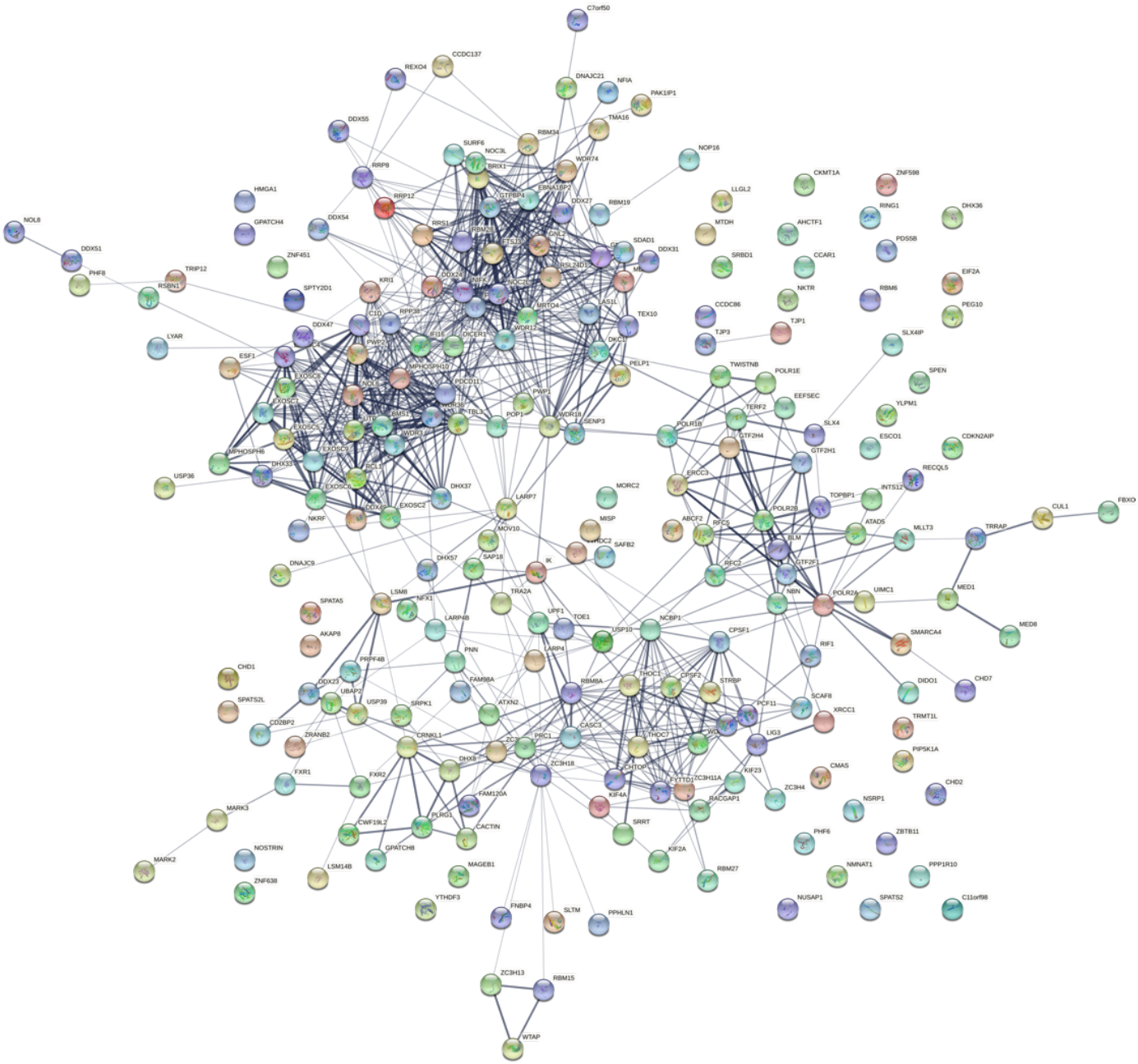
BARD1 140-160 (native)



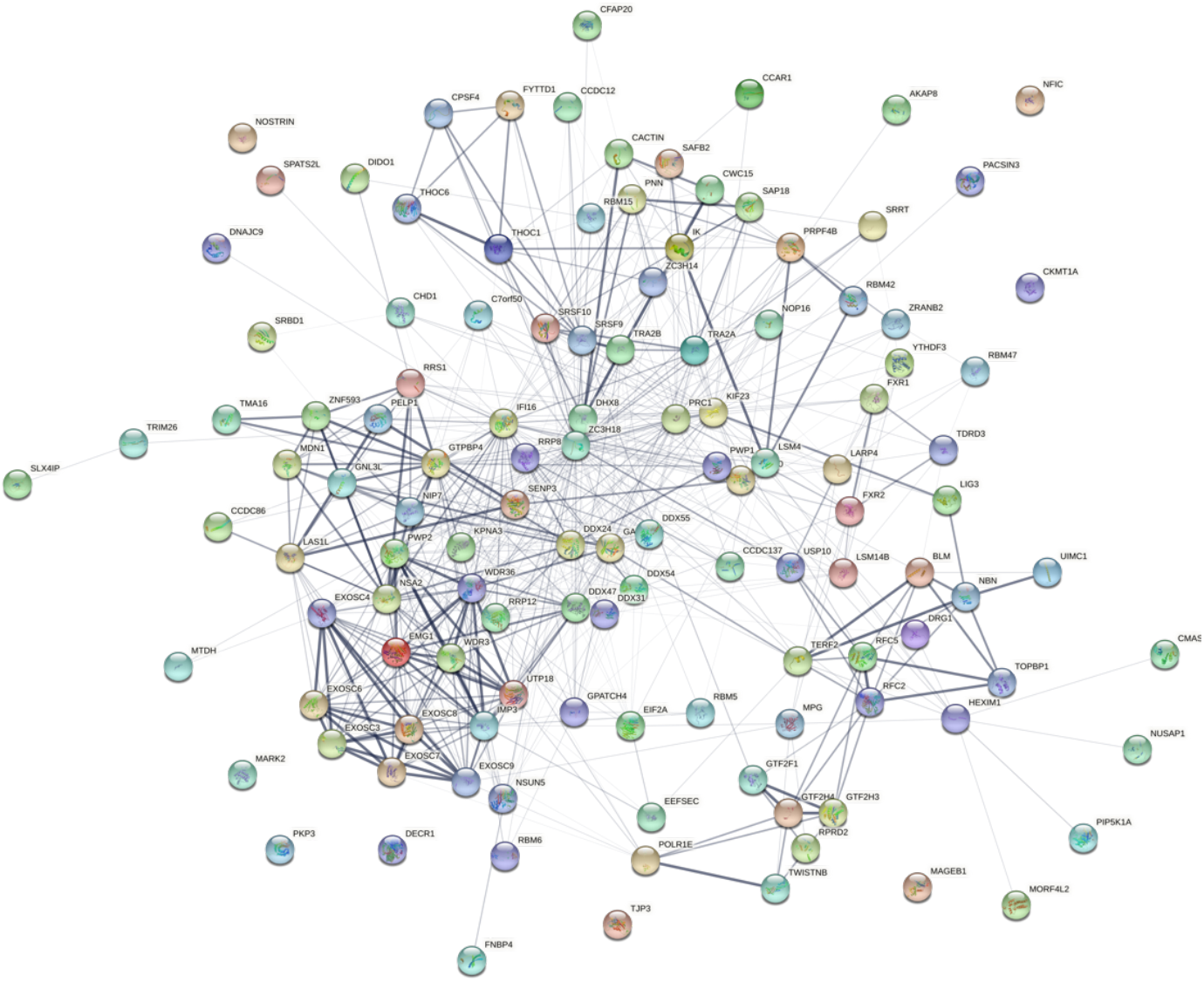
BARD1 140-160 (phosphopeptide)



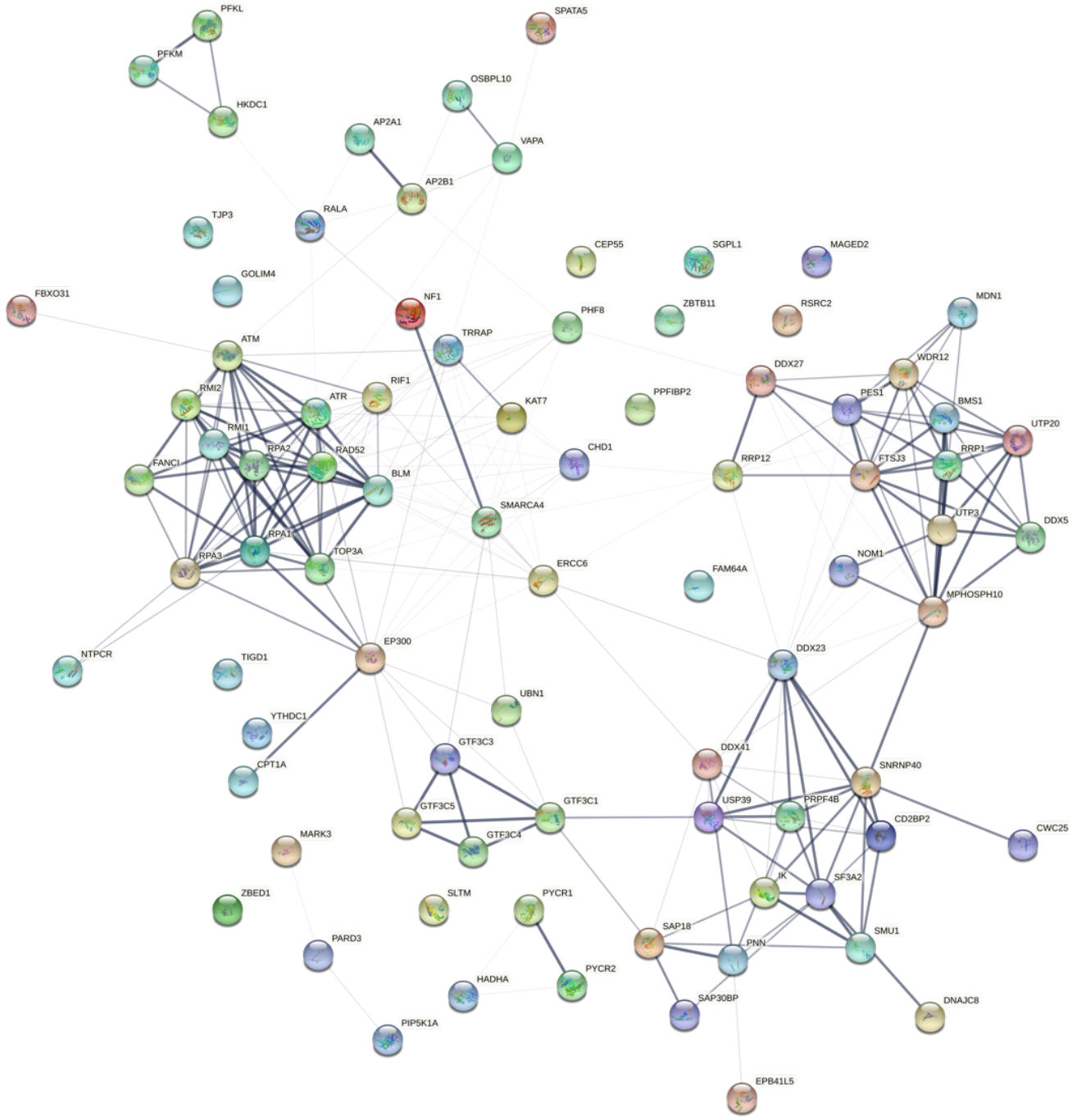
BRCA1 455-467



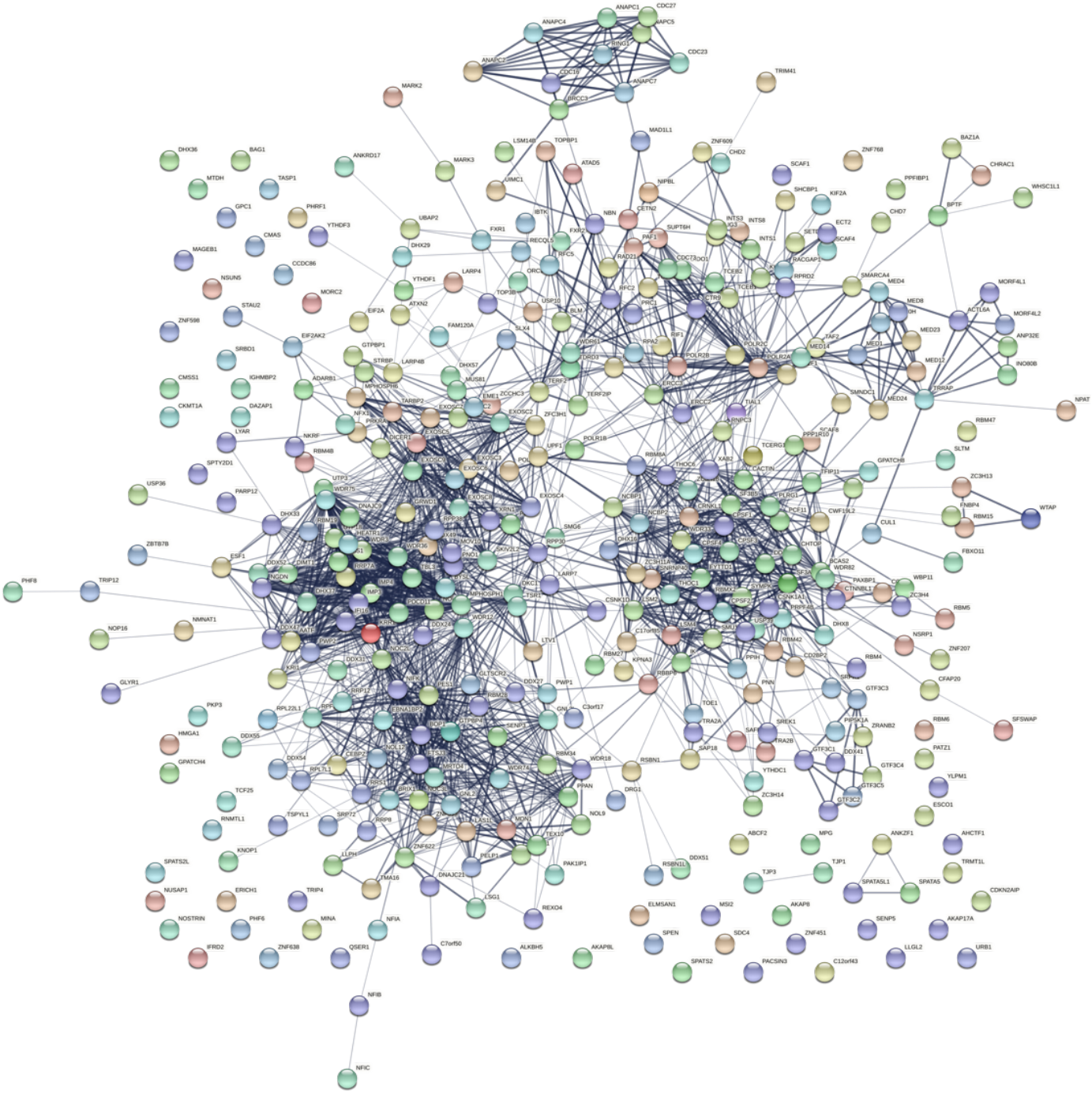
BRCA1 501-527



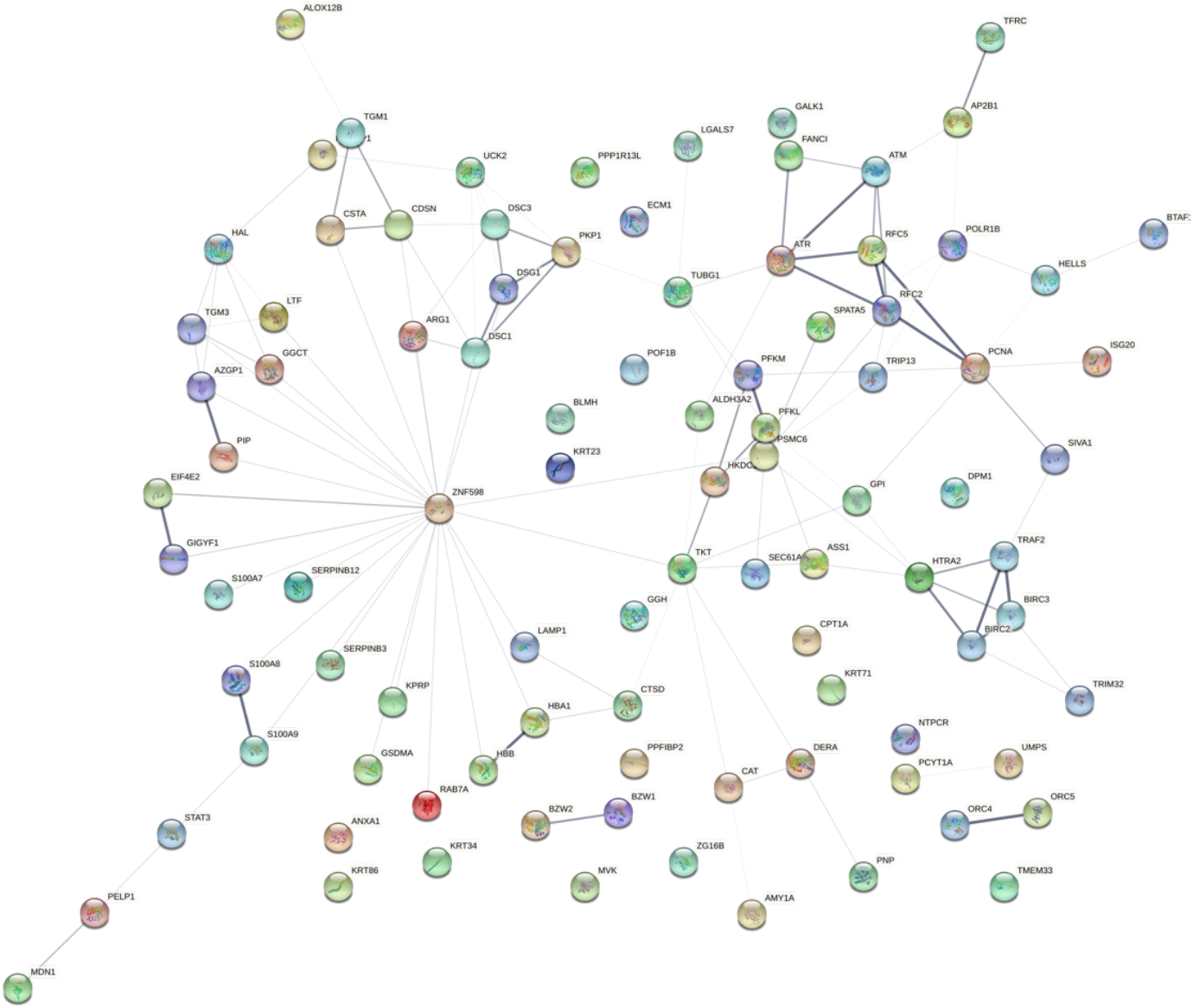
BRCA2 1100-1132



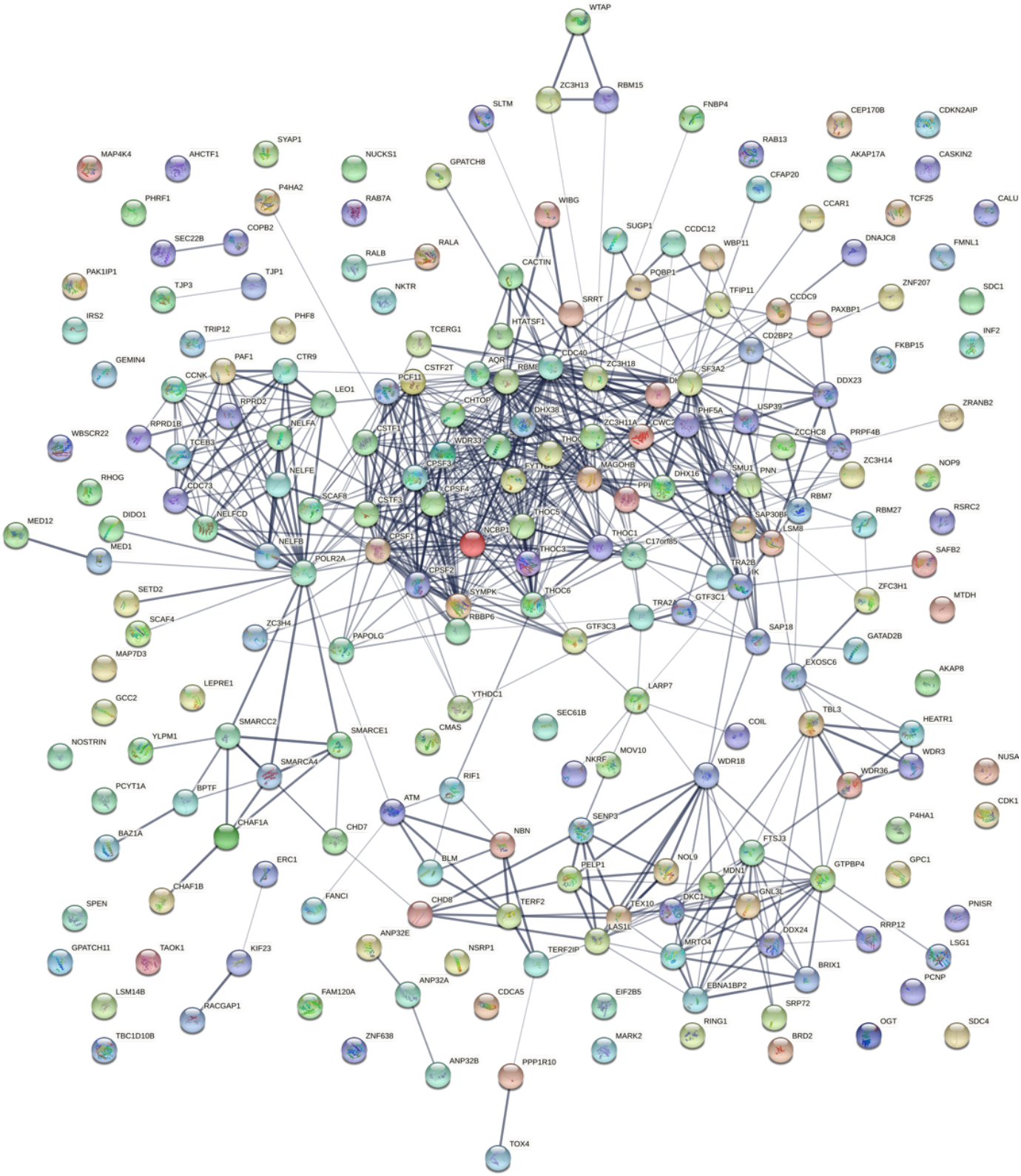
BRCA2 2401-2414



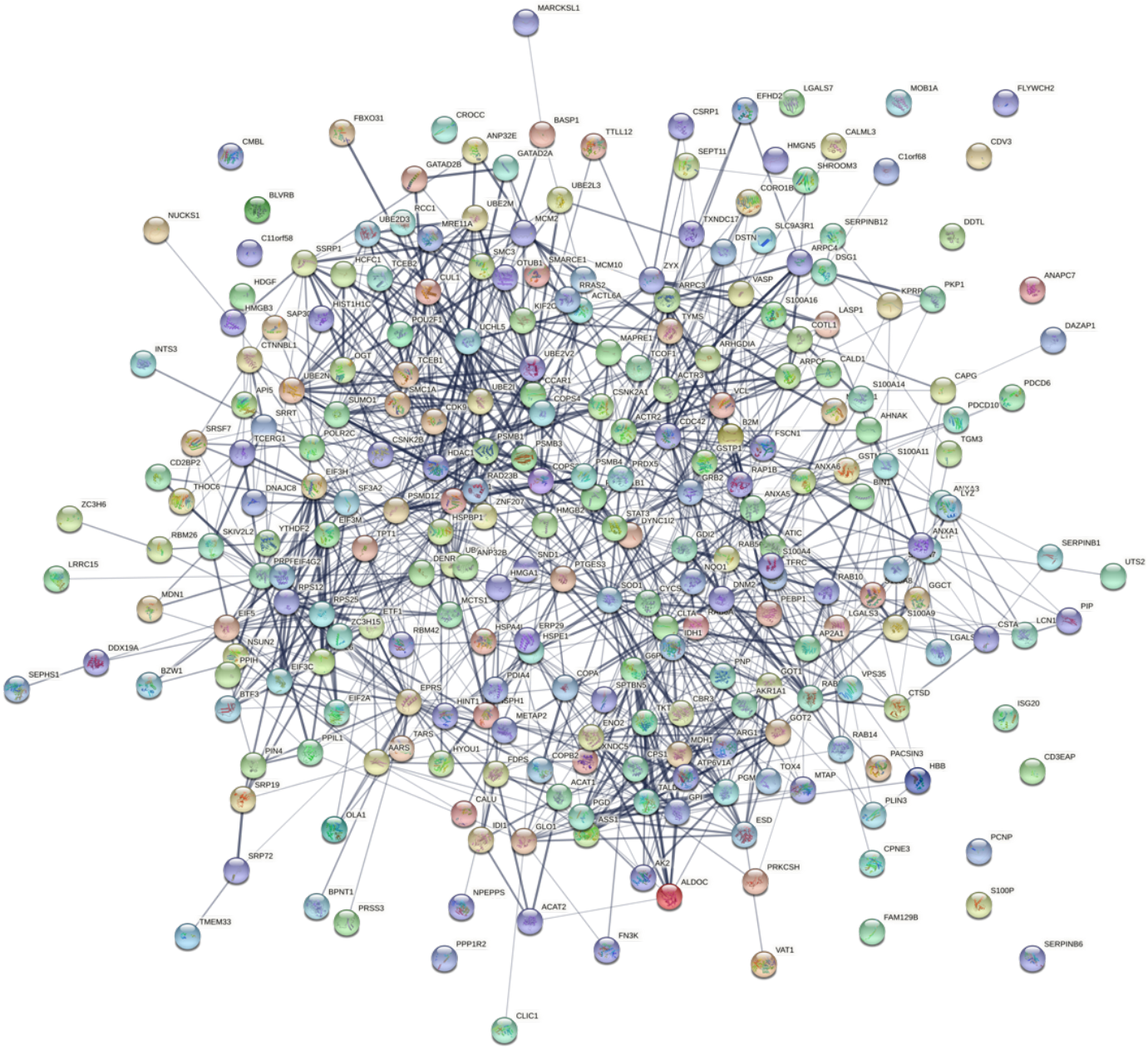
ERCC6L2 796-812



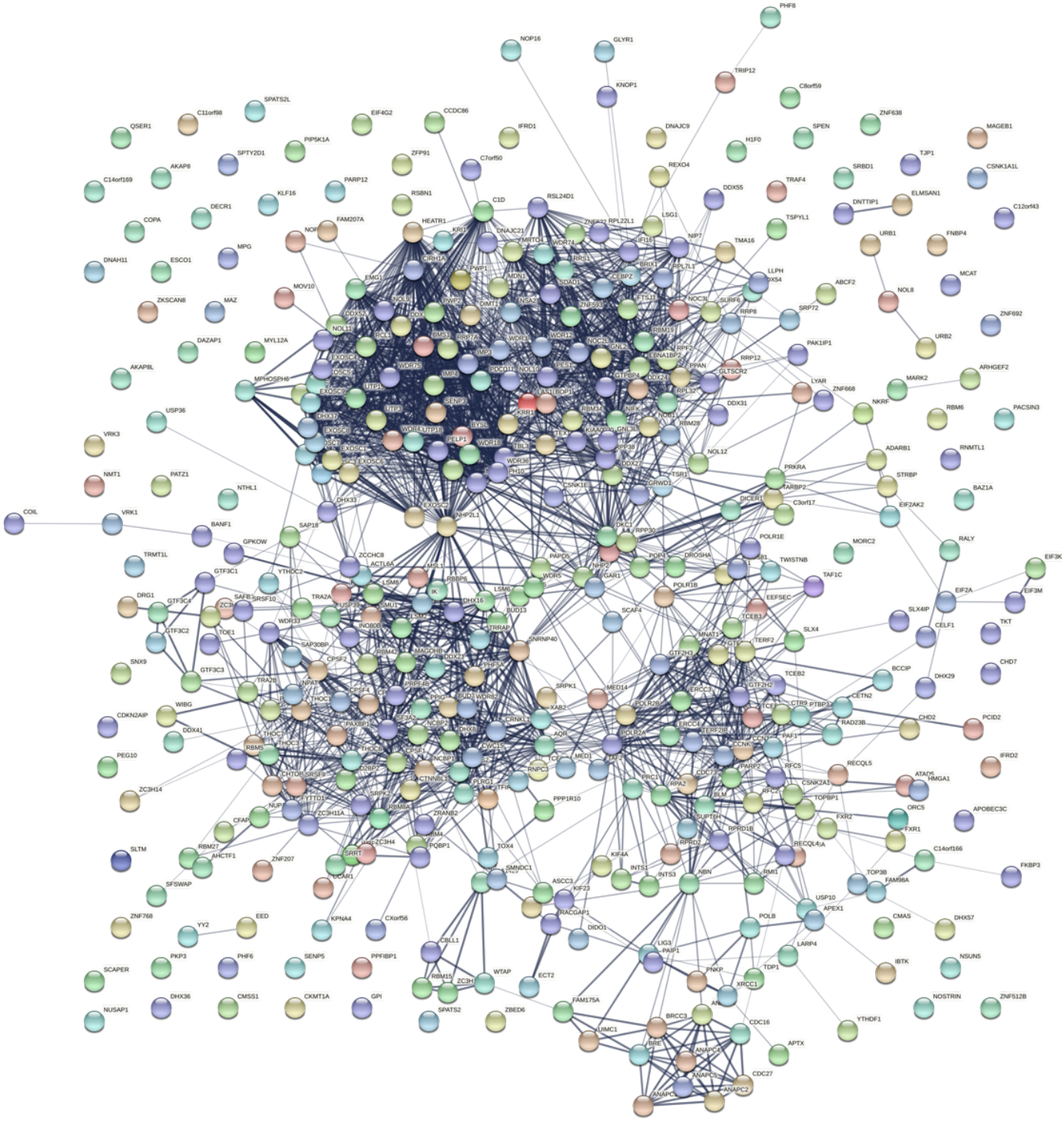
POLQ 1631-1652



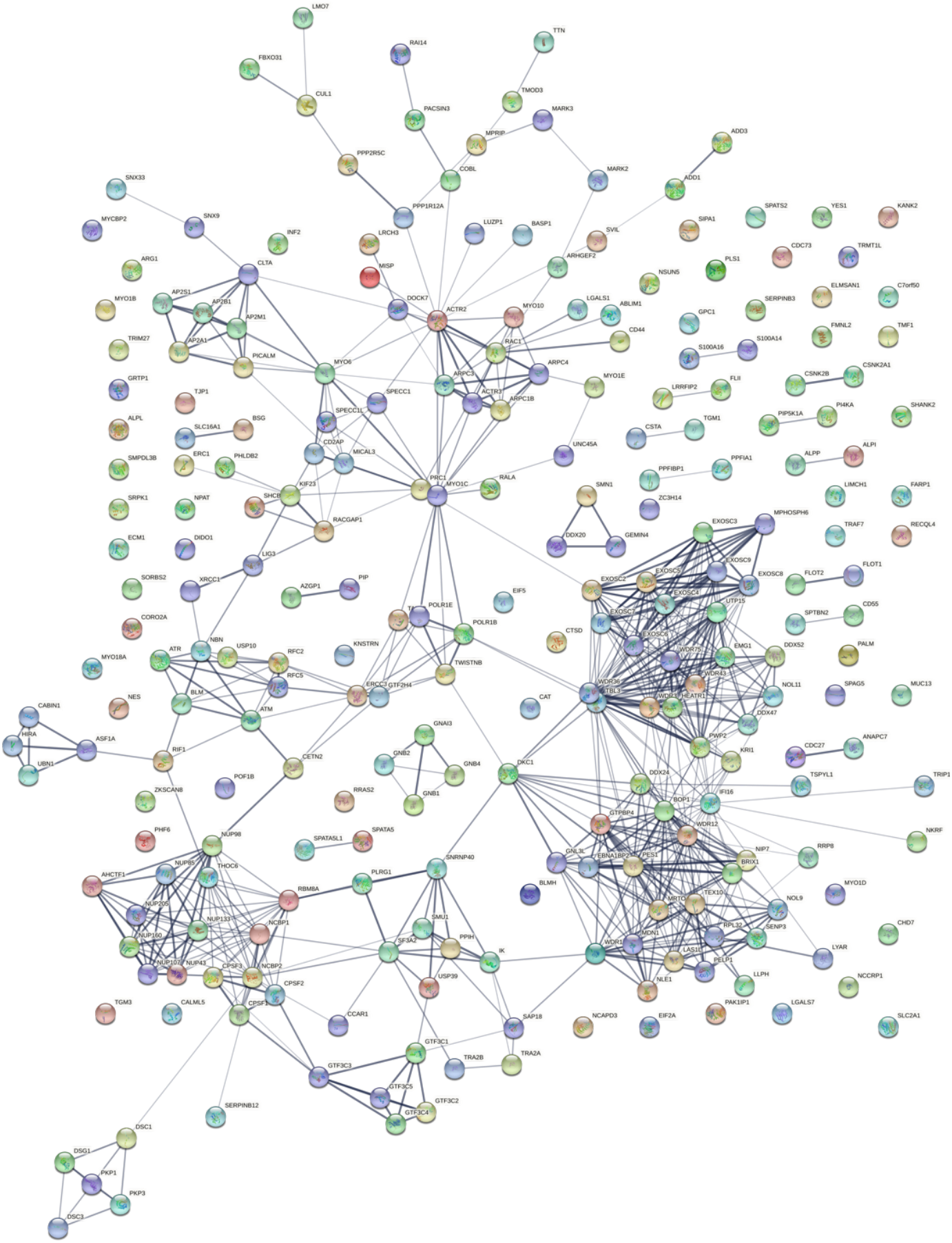
RECQL4 68-87


MARCKSL1
EFHD2
LGALS7
MOB1A
FLYWCH2
CSRP1
FBXO31
CROCC
BASP1
TTLL12
CMBL
ANP32E
CALML3
HMGN5
CDV3
SEPT11
GATAD2B
GATAD2A
SHROOM3
C1orf68
UBE2L3
CORO1B
BLVRB
NUCKS1
UBE2M
UBE2D3
RCC1
MRE11A
MCM2
TXNDC17
SLC9A3R1
SERPINB12
DDTL
C11orf58
SSRP1
HCFC1
TCEB2
SMC3
OTUB1
SMARCE1
MCM10
ZYX
DSTN
ARPC4
DSG1
HDGF
CUL1
RRAS2
ACTL6A
ARPC3
VASP
KPRP
PKP1
ANAPC7
HMGB3
HIST1H1C
SAP30
KIF2C
S100A16
LASP1
TYMS
COTL1
POU2F1
UCHL5
MAPRE1
DAZAP1
INTS3
CTNNBL1
OGT
TCEB1
UBE2V2
TCOF1
ARHGDIA
ARPC1
CALD1
UBE2N
ACTR3
CAPG
API5
SMC1A
UBE2I
CCAR1
CSNK2A1
VCL
S100A14
PDCD6
SUMO1
COPS4
CDK9
ACTR2
CDC42
B2M
PDCD10
SRRT
AHNAK
SRSF7
TCERG1
POLR2C
PSMB1
HDAC1
PSMB3
FSCN1
GSTP1
TGM3
CD2BP2
EIF3H
PSMB4
PRDX5
COPS5
RAD23B
GRB2
RAP1B
ANXA6
GSTT1
S100A11
THOC6
DNAJC8
SF3A2
PSMD12
ZNF207
HMGB2
ANXA5
BIN1
ANXA3
LY2
ZC3H6
EIF3M
HSPBP1
STAT3
DYNC1I2
GDI2
ATIC
ANXA1
RBM26
SKIV2L2
YTHDF2
TPT1
RAB5C
S100A4
SERPINB1
LRRC15
PRPF19
EIF4G2
DENR
UBA52
ANP32B
SND1
PTGES3
NQO1
TFRC
UTS2
RPS12
RPS25
HMGA1
SOD1
CYCS
DNM2
RAB10
PEBP1
GGCT
MDN1
EIF5
ETF1
MCTS1
ZC3H15
CLTA
LGALS3
S100A9
PIP
RBM42
HSPA4L
ERP29
HSPE1
IDH1
AP2A1
CSTA
LCN1
DDX19A
NSUN2
PPIH
G6PD
PNP
LGALS1
SEPHS1
BZW1
EIF3C
PDIA4
COPA
GOT1
VPS35
BTF3
EIF2A
EPRS
HINT1
SPH1
SPTBN1
AKR1A1
RAB1B
CTSD
METAP2
CBR3
TKT
ISG20
ENO2
GOT2
TARS
MDH1
ARG1
RAB14
PPIL1
PIN4
AARS
HYOU1
XNDC1
GPS1
PACSIN3
CD3EAP
COPB2
FDPS
ATP6V1A
TOX4
PGM
MTAP
HBB
SRP19
ACAT1
TALDO1
GPI
PLIN3
CALU
PGD
OLA1
IDI1
GLO1
ASS1
ESD
PCNP
SRP72
CPNE3
AK2
PRKCSH
ALDOC
BPNT1
NPEPPS
S100P
PRSS3
ACAT2
TMEM33
FAM129B
FN3K
PPP1R2
SERPINB6
VAT1
CLIC1

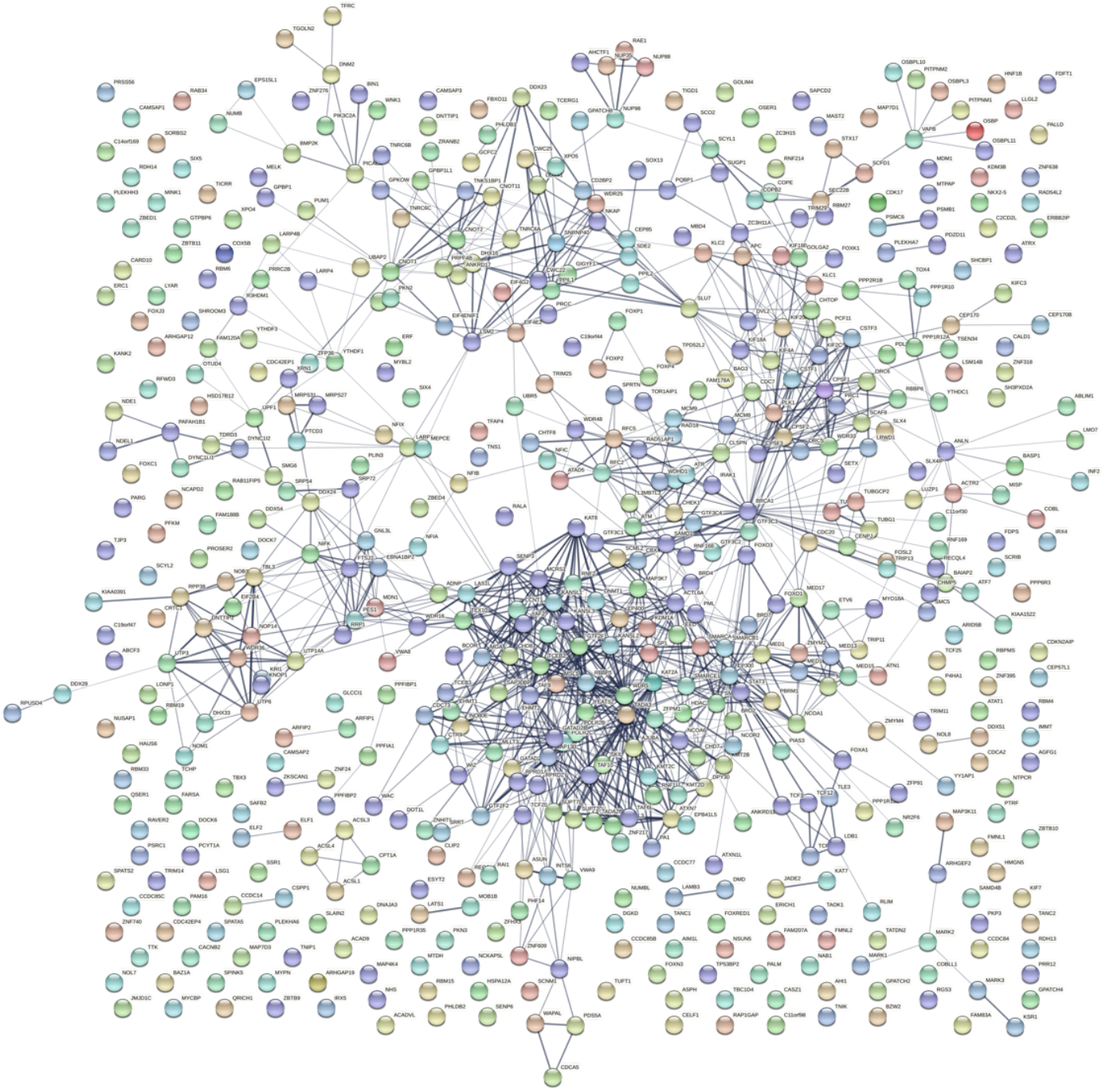
RECQL4 101-124



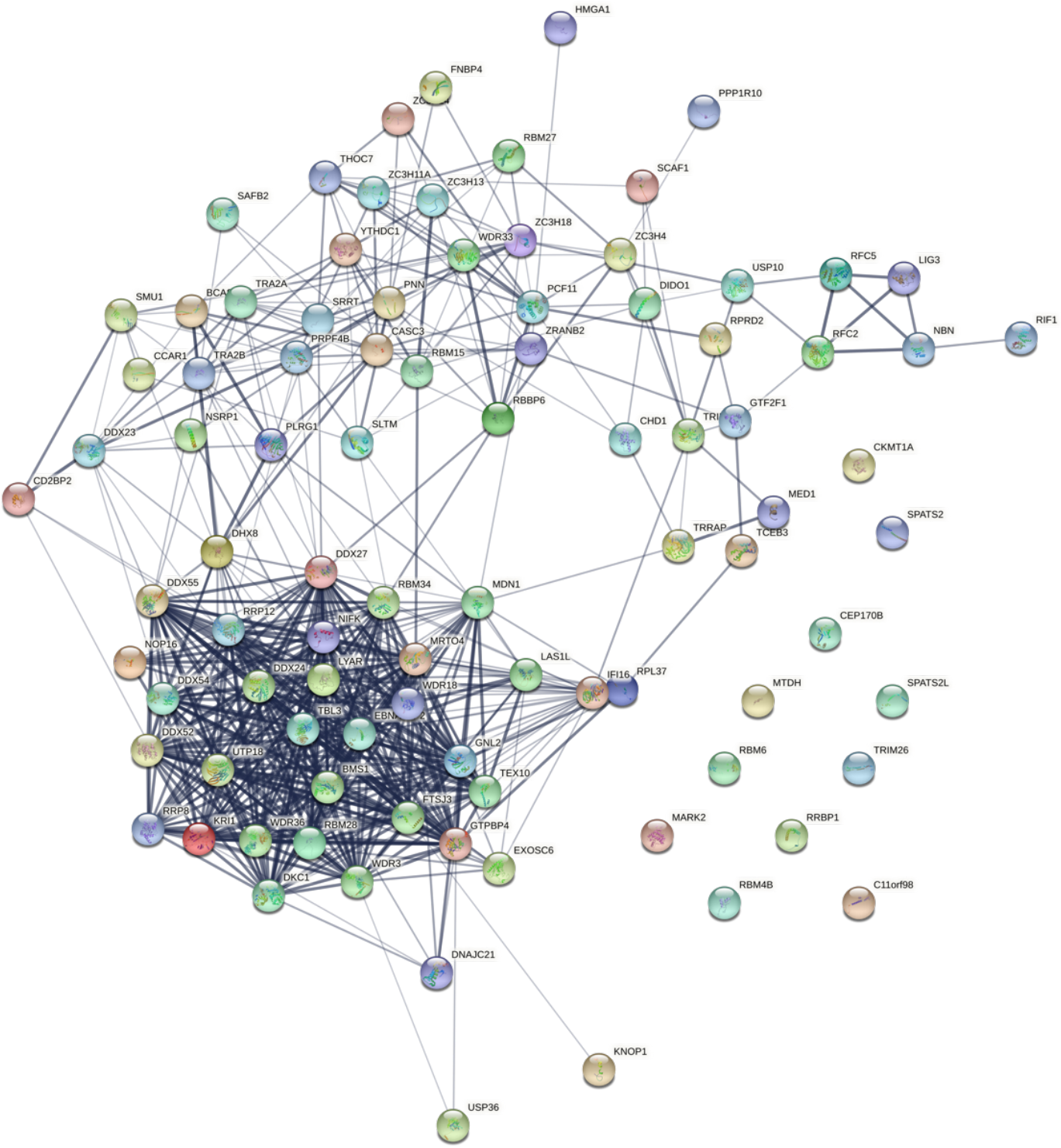
RECQL4 171-197



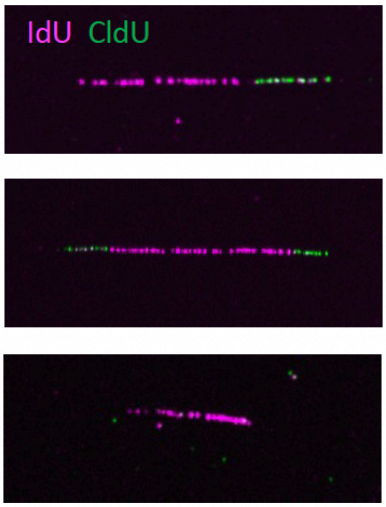
SHLD2 265-283



XRCC4 255-278



8.5. Representative images of DNA fibres

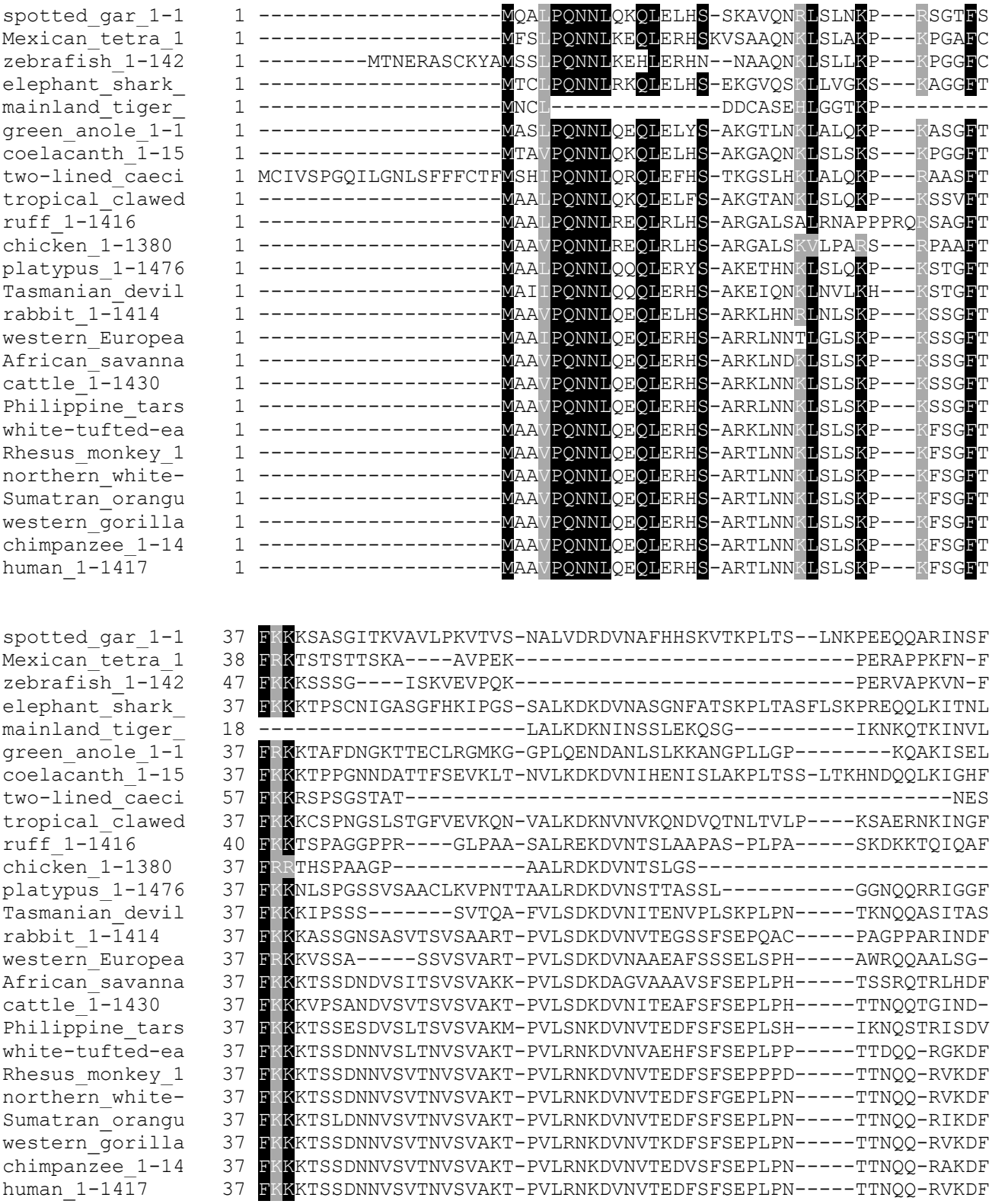


1) Fork restart (unidirectional)

2) Fork restart (bidirectional)

3) Stalled fork

8.6. Automated Sequence Alignment of BLM Orthologs from the SLiMCon Tool

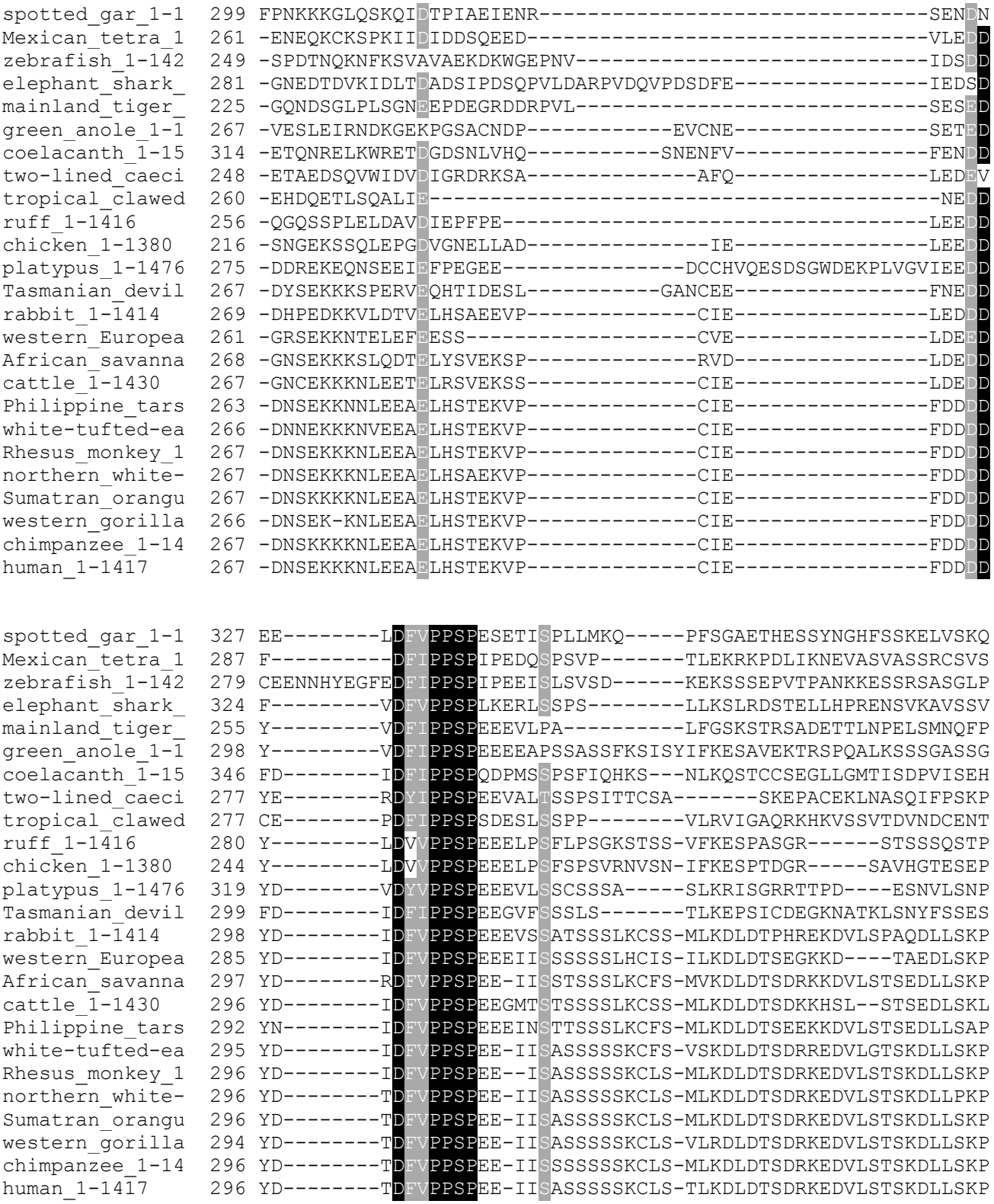


```
spotted_gar_1-1    94 FELTHKAKSQNFTPVVTRPAVSNIKLSLSTTKASPQSALVNTDVG---------KKDDSE
Mexican_tetra_1    65 FSAPSKPKTQSINPSANPFAVTKPTLIQSVIKTSAKSETIPPDTH---------EDKSAE
zebrafish_1-142    74 FTAPSKPKTNTVNPLANPFALNKTPLVQSTIKA---LPKSAALPK---------DSTNDQ
elephant_shark_    96 FEQNTKNSA----------LSVGATSTPRETRI-------TASVS---------NTQLTK
mainland_tiger_    46 FKTTLKVQE------GKQEYSNDQHAKNLDCFI---SSVKEGDAEGKV------GGSEIP
green_anole_1-1    88 FGTTQKGQE------GQQSHPNDPSESSLGHFT---FSVKERDAG---------SQLENP
coelacanth_1-15    95 FEKNCKSQPQVTQMFSGQPFLNCSESVRQTSKILPVFTRPEETIK-EPKQPSAKPEESSK
two-lined_caeci    72 FKGADRGEA----------AGPSRTALESQRVC---SPPPGQQEL---------ADSPST
tropical_clawed    92 FTPVYTKST----------QLPQTVAVKDHVHI-NDSANKHPRVE---------DSASK
ruff_1-1416        89 FPVVSGGCSLQPSPVLARPQEASRGPVASGQRSVSEGLRGTPAPR----------------
chicken_1-1380     62 -GGVARPKDLRVGPPGARPALPGP----------AGSSGPTAG---------APPGSV
platypus_1-1476    86 FEKSLKNEE--TQPLGTDRAGSSSRKTQTAVSRGSQPPAKGMTITPDP------PDIVSK
Tasmanian_devil    84 SEKTSKGQQ--IHKFCSNQILSDSSKNTHETSCYLQPSTNENSVK-KV------LCKDSK
rabbit_1-1414      91 FKNAPAGLQ--SRGAGSTPLLPDVSQTPQEASC---TPRLAPAAK-TP------RDTVFK
western_Europea    85 FSHSPAGPQ--TPKAGSEPSSADLLHTPQGAFC---TTQNTPAVK-AS------RGSTMK
African_savanna    91 LQNAPADQE--ITRGGSKPLSPDLLQTPGEVSC---TARNTPAVR-ES------HDVTCK
cattle_1-1430      90 FKSAPSGPQ--TKRVGSKPLLPSLSQVPQEVSC---TAQNTPVIK-QS------LNATFK
Philippine_tars    91 FKNAPEGQQ--TQRLASKPSL----QIPKEVLC---TTQNTPVIK-KS------RDAAFK
white-tufted-ea    90 SKNVPAGQQ--TQRVGSKPLLPDFLQTPKEVLC---TTQNTPTLE-KS------RDTAFK
Rhesus_monkey_1    90 FKNAPAGQQ--TQRVGSKPLLPDFLQTPKEVLC---TTQNTPTVK-KS------QNTALK
northern_white-    90 FKNAPEGQQ--TQRVGSKPLLPDFLQTPKEVLC---ATQNTPTVK-KS------RDTVLK
Sumatran_orangu    90 FKNAPAGQQ--TQRVESKPLLPDFLQTPKEVLC---TTQNTPTVK-KS------WDTALK
western_gorilla    90 FKNATAGQQ--TQRVESKSLLPDFLQTPKEVLC---TTQNTPTIK-KS------RDTALK
chimpanzee_1-14    90 FKNAPAGQQ--TQRVGSKSLLPDFLQTPKEVLC---TTQNTPTVK-KS------RDTALK
human_1-1417       90 FKNAPAGQE--TQRGGSKSLLPDFLQTPKEVVC---TTQNTPTVK-KS------RDTALK

spotted_gar_1-1   145 KTALNNSLS--VSIDD-WDDFDDFD---------------FGTPVKDKSVTRLSQSNS-G
Mexican_tetra_1   116 SLVGNDSFA--IPVDD-WDDLDDFETPSK--VRA-------SQSPQIPVKKTSGLDKN-V
zebrafish_1-142   122 TVELKHDISQL-SFNE-WDDLDDFETPVK--SR--------VASPVAGTSTKKPSVS---D
elephant_shark_   130 KPDADHSFESVVSLDD-WDDFDDFDTSWK--GKSNVKLT-SASSNKALQVQGCKTSQN-S
mainland_tiger_    91 TPDNNCLNNSFIDLTDIWEDMEDFETTGP--QK-------ISPKLSVSFRDKLSVNNGKT
green_anole_1-1   130 KADHNSSYSSIINIDDVWDDIDDFEIARQ--EKKCSKSP-ASLSSNTLLAKKSKAPKN-Q
coelacanth_1-15   154 KSELDCSFSTVISIED-WDDLDDFDTSWK--DKKSSQT--ISPPDKVRQAKHYKSTPSSH
two-lined_caeci   110 ETCLNSSFSSVISLDG-WDDIDDFDTSWK--EKKTCRTSVPVKSCQLPEKGAARTWIL-K
tropical_clawed   131 KTGINTSFGSVTSLEE-WDDLDDFDT--------------SVSPPKSQAGKIKITPQNCK
ruff_1-1416       134 -------PVAVPAVGDEWDDIDDFDLSGI--EKK------YCRPPVLSPKGQRATCKASQ
chicken_1-1380    100 SVSLPSSGPAAYTGDE-WDDMDDFDLSGF--EKK------FSRPA-VLSPKGPRTPQG-R
platypus_1-1476   138 RLDFSSALDTILSLDD-WDDLDDFDTSGKNCPKA------FVTPPSSQAVKAGRTPVS-S
Tasmanian_devil   135 KLELNSSFDSTSLVSD-WDDIDDFDTSGN--SKT------FTTPSRSNFAKLNSTQRS-K
rabbit_1-1414     139 KLDFSSSLDSLGALNG-WDDMDDFDTSGT--PKA------FAASPRKPVARVSTAQKS-K
western_Europea   133 KLEVSSSSDSFLLPNE-WDDMDDFDTSGN--SKA------FVTPPKNHFVRVSTAQKS-K
African_savanna   139 KLKFSSSTASLITIND-WDDMDDFDISGN--SEA------FVTPLRTHNVRISTAQKS-K
cattle_1-1430     138 KLEFSSSSDSLIAFND-WDDMDDFDTSGN--SKA------FVTPCKNHFVRVSTAQKS-K
Philippine_tars   135 TLEFSSSSDSFNAIDD-WDDMDDFDTTGT--SKA------FVTPSRNHFVRVSTAQKS-K
white-tufted-ea   138 KLEFSSSPDSLSTIND-WDDMDDFDTSGT--SKA------FVTPPQSHFVRVSTAQKS-K
Rhesus_monkey_1   138 KLEFSSSPDSFRTIND-WDDMDDFDTSRT--SKA------FVTPPQSHFVRVSTAQKS-K
northern_white-   138 KLEFSSSPDSLSTIND-WDDMDDFDTSGT--SKS------FVTPPQSHFVRVSTAQKS-K
Sumatran_orangu   138 RLEFSSSPDSLSTIND-WDDMDDFDTSET--SKS------FVTPPQSHFVRVSTAQKS-K
western_gorilla   138 KLEFSSSPDSLSTIND-WDDMDDFDTSET--SKS------FVTPPQSHFVRVSTAQKS-K
chimpanzee_1-14   138 KLEFSSSPDSLSTIND-WDDMDDFDTSET--SKS------FVTPPQSHFVRVSTAQKS-K
human_1-1417      138 KLEFSSSPDSLSTIND-WDDMDDFDTSET--SKS------FVTPPQSHFVRVSTAQKS-K
```

```
spotted_gar_1-1  186 ENFKKESASNLNKPCLKTSRAAEKILSESSDHNISNGLQNQEEPLEAQNGPAALPVQCR-
Mexican_tetra_1  163 SPKAASAKSA------DTTENGSLPVGPSFGACGEALEIPTDTSGADRHDAISAGEP---
zebrafish_1-142  168 QNTSSSCSSKCEETKVNEESQATETITAPKDVLSAANGVSTE---TAEREP---------
elephant_shark_  185 KHSTPTATKSVTTPDLDTSKVN---GELKVDFLPKDSTHSLS---QREPKGSDDPVICLV
mainland_tiger_  142 AKNPSVSKSKISSTGRDNEKCDDVYICTNRKTQTL----------HEEQEGGSLPVPCL-
green_anole_1-1  186 KPLTSEQLVPLVSR--QGSQLCNEGFVNNKSNTIQIPDDDDD---QEENAPGSQSVICL-
coelacanth_1-15  209 QSKINLSSGTFNDS--LSTRNSKNVLSPQKSIQDTMDKDNQVGY-ANSFEISERSLTCV-
two-lined_caeci  166 GSGQKLVQQRLAKPCAWKRGST----FPRNGLVGSLDGSDP----CREAQPRDTEASSL-
tropical_clawed  176 NTSPVSSKIQCLSP--KGPTIEKQMCAKKPYENNEIASEPQKLVHAQTAEMPDKSLVCL-
ruff_1-1416      179 RPEPRRDELPGSSP--GAVGQD---TGPGSRGTPGRGETPPE---AEQQPSSQQSVICL-
chicken_1-1380   149 GLRVSKARAD---------------VGPQEHGSEV----------SCSQGSSQGSLICL-
platypus_1-1476  190 R-KDDASNARVSKA--KTGKVGVELVEPRPGGGDRVARTSDP---LREPEPSGDSVIY--
Tasmanian_devil  185 GDKNNLSKTRARKA--NGKKVDPELLNFPSGKNKQAELIKKE---LPESEELSSPMICL-
rabbit_1-1414    189 RAKRNFCKAQLKKA--DAAEAD---LTPASAASKPVELAGAG---KGDCDGVGGGVICI-
western_Europea  183 KSKRNFFKAPRPKA--NKVEIN----LTLSSESSHIHSAKEQ---KGALEWSHSGVICI-
African_savanna  189 KGKRNFLKSQLHKA--DVVKVD---LTSSSAESEQVDLTKEQ---QNDSEWLSGGVVCI-
cattle_1-1430    188 KSKRNLVKAQLNKT--HTGETD---LAPFSSESKQEHLTEKQ---KDDSEWLSSDVICI-
Philippine_tars  185 KAKRNFLKAQLYKT--NTVNID----LTSSSESKQVDLTKEQ---KDDSEWLSSDVICI-
white-tufted-ea  188 KAKRNFFKAQLCAT--NTVKAD----LPPSSESKQIDLTVEQ---KDDSEWFSSDVICI-
Rhesus_monkey_1  188 KAKRNFFKAQLYTT--NTVKAD---LPPSFSESEQIDLTKEQ---KDDSEWLSSDVICI-
northern_white-  188 KGKRNFFKAQLYTT--NTVKAD---LPPPSSESEQIDLTEEQ---KDDSEWLSSDVICI-
Sumatran_orangu  188 KGKRNFFKAQLYTT--NTVKAD---LPPPSSESEQIDLTEEQ---KDDSEWLSSDVICI-
western_gorilla  188 KGKRNFFKAQLYTT--NTVKTD----LPPPSESEQIDLTEEQ---KDDSEWLSSDVICI-
chimpanzee_1-14  188 KGKRNFFKAQLYTT--NTVKTD---LPPPSSESEQIDLTEEQ---KDDSEWLSSDVICI-
human_1-1417     188 KGKRNFFKAQLYTT--NTVKTD---LPPPSSESEQIDLTEEQ---KDDSEWLSSDVICI-


spotted_gar_1-1  245 ------PLPDAFAAVEKPQGCVEDKEEEDDVPIKFARKRHLSQIAQILSDSEDDCVEEGD
Mexican_tetra_1  214 ----------DDSPIKLTKRR-LSAVKQKPLLSDSEDESITDNIKPHNKADENFKYSE--
zebrafish_1-142  216 ----------EDSPIKKSKRP-KKSVQHTALLSDTEDEEIIHCV----------------
elephant_shark_  239 KPHLEEIKNSDDEDCVESFAR-RKISRSSGARVLLSDDETDVS----------------
mainland_tiger_  191 ----------DSSNSSQRDFSNAEREEEKSLETFQTEKKEKSWI----------------
green_anole_1-1  240 ------------GPITSHDGSYEDVPANRPLGSPPGSKK--------------------
coelacanth_1-15  265 ---DLVSLNSTGKDCNDAAEVPAAKKKSLSTGSVVIDEEETNCIKENERRDS--------
two-lined_caeci  217 ------------KAATETEGT-QRKGDTKNYKVPGEPDRRIDPE----------------
tropical_clawed  233 ------------GSVEHTNLGSAEKDTCRNIDYSGTDDI----------------------
ruff_1-1416      230 ---------EDSAPCGDEKAVGSGAGGGAGQRAGG-------------------------
chicken_1-1380   183 --------GDAGAPAAAVLSD-NGMEEARAATNGGSGESQKL------------------
platypus_1-1476  242 ----------AETADSRGIPA-EDVQEVRSFEAISHDEREKNHT----------------
Tasmanian_devil  239 ----------DDDPISEEFKN-EDIQEKQALKTQLDDER--------------------
rabbit_1-1414    240 --------DDDDGPVSQVLI--LDSQESHSLETHLGGER--------------------
western_Europea  233 ----------DDDSQHEEFLK-EETRGSRSPETHSEIER--------------------
African_savanna  240 ----------EDDPISKRLVD-EETQESDPLKTHLGDEK--------------------
cattle_1-1430    239 ----------DDDPDPEELIN-EDTQENHSLKTHLGAER--------------------
Philippine_tars  235 ----------DDGPVFKVPTS-EDTQKSHSLETHLGDER--------------------
white-tufted-ea  238 ----------DDGPIAEVHVN-EDAQESDSLKTHLQDER--------------------
Rhesus_monkey_1  239 ----------DDGPIDEVHIN-EDAQESDSLKTHLEDER--------------------
northern_white-  239 ----------DDGPIAEVHIN-EDAQESDSLKTHLEDER--------------------
Sumatran_orangu  239 ----------DDGPIAEVHIN-EDAQESDSLKTHLEDER--------------------
western_gorilla  238 ----------DDGPIAEVHIN-EDAQESDSLKTHLEDER--------------------
chimpanzee_1-14  239 ----------DDGPIAEVHIN-EDAQESDSLKTHLEDER--------------------
human_1-1417     239 ----------DDGPIAEVHIN-EDAQESDSLKTHLEDER--------------------
```

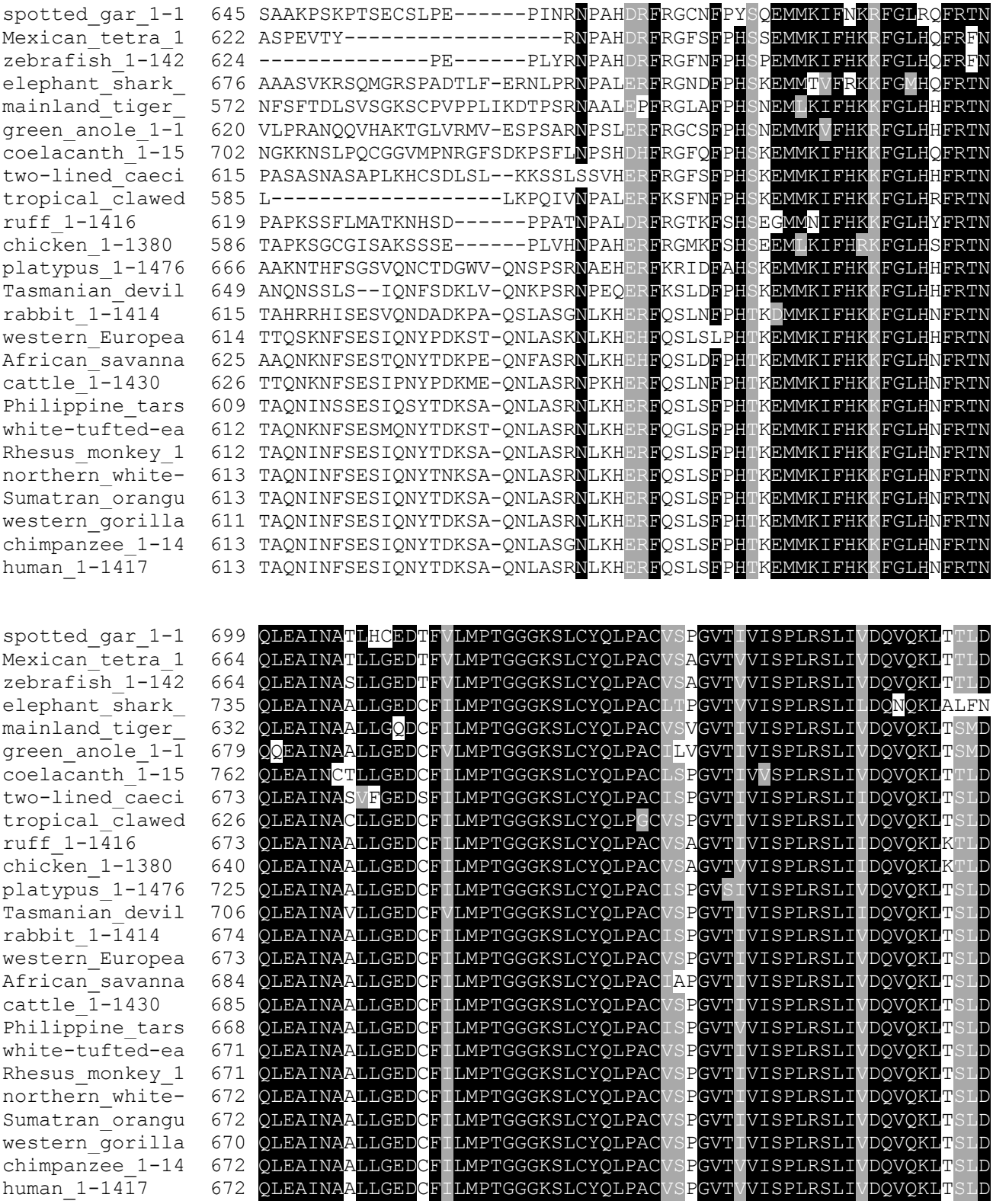

spotted_gar_1-1 299 FPNKKKGLQSKQIDTPIAEIENR----------SEND
Mexican_tetra_1 261 -ENEQKCKSPKIIDIDDSQEED----------VLED
zebrafish_1-142 249 -SPDTNQKNFKSVAVAEKDKWGEPNV----------IDSD
elephant_shark_ 281 -GNEDTDVKIDLTDADSIPDSQPVLDARPVDQVPDSDFE----------IEDSD
mainland_tiger_ 225 -GQNDSGLPLSGNEEPDEGRDDRPVL----------SESED
green_anole_1-1 267 -VESLEIRNDKGEKPGSACNDP----------EVCNE----------SETED
coelacanth_1-15 314 -ETQNRELKWRETDGDSNLVHQ----------SNENFV----------FEND
two-lined_caeci 248 -ETAEDSQVWIDVDIGRDRKSA----------AFQ----------LEDEV
tropical_clawed 260 -EHDQETLSQALIE----------NED
ruff_1-1416 256 -QGQSSPLELDAVDIEPFPE----------LEED
chicken_1-1380 216 -SNGEKSSQLEPGDVGNELLAD----------IE----------LEED
platypus_1-1476 275 -DDREKEQNSEEIEFPEGEE----------DCCHVQESDSGWDEKPLVGVIEED
Tasmanian_devil 267 -DYSEKKKSPERVEQHTIDESL----------GANCEE----------FNED
rabbit_1-1414 269 -DHPEDKKVLDTVELHSAEEVP----------CIE----------LEDD
western_Europea 261 -GRSEKKNTELEFEESS----------CVE----------LDEED
African_savanna 268 -GNSEKKKSLQDTELYSVEKSP----------RVD----------LDEDD
cattle_1-1430 267 -GNCEKKKNLEETELRSVEKSS----------CIE----------LDEDD
Philippine_tars 263 -DNSEKKNNLEEAELHSTEKVP----------CIE----------FDDDD
white-tufted-ea 266 -DNNEKKKNVEEAELHSTEKVP----------CIE----------FDDDD
Rhesus_monkey_1 267 -DNSEKKKNLEEAELHSTEKVP----------CIE----------FDDDD
northern_white- 267 -DNSEKKKNLEEAELHSAEKVP----------CIE----------FDDDD
Sumatran_orangu 267 -DNSKKKKNLEEAELHSTEKVP----------CIE----------FDDDD
western_gorilla 266 -DNSEK-KNLEEAELHSTEKVP----------CIE----------FDDDD
chimpanzee_1-14 267 -DNSKKKKNLEEAELHSTEKVP----------CIE----------FDDDD
human_1-1417 267 -DNSEKKKNLEEAELHSTEKVP----------CIE----------FDDDD
spotted_gar_1-1 327 EE--------LDFVPPSPESETISPLLMKQ-----PFSGAETHESSYNGHFSSKELVSKQ
Mexican_tetra_1 287 F----------DFIPPSPIPEDQSPSVP-------TLEKRKPDLIKNEVASVASSRCSVS
zebrafish_1-142 279 CEENNHYEGFEDFIPPSPIPEEISLSVSD------KEKSSSEPVTPANKKESSRSASGLP
elephant_shark_ 324 F---------VDFVPPSPLKERLSSPS--------LLKSLRDSTELLHPRENSVKAVSSV
mainland_tiger_ 255 Y---------VDFIPPSPEEEVLPA----------LFGSKSTRSADETTLNPELSMNQFP
green_anole_1-1 298 Y---------VDFIPPSPEEEEAPSSASSFKSISYIFKESAVEKTRSPQALKSSSGASSG
coelacanth_1-15 346 FD--------IDFIPPSPQDPMSSPSFIQHKS---NLKQSTCCSEGLLGMTISDPVISEH
two-lined_caeci 277 YE--------RDYIPPSPEEVALTSSPSITTCSA-------SKEPACEKLNASQIFPSKP
tropical_clawed 277 CE--------PDFIPPSPSDESLSSPP--------VLRVIGAQRKHKVSSVTDVNDCENT
ruff_1-1416 280 Y---------LDVVPPSPEEELPSFLPSGKSTSS-VFKESPASGR------STSSSQSTP
chicken_1-1380 244 Y---------LDVVPPSPEEELPSFSPSVRNVSN-IFKESPTDGR-----SAVHGTESEP
platypus_1-1476 319 YD--------VDYVPPSPEEEVLSSCSSSA-----SLKRISGRRTTPD----ESNVLSNP
Tasmanian_devil 299 FD--------IDFIPPSPEEGVFSSSLS-------TLKEPSICDEGKNATKLSNYFSSES
rabbit_1-1414 298 YD--------IDFVPPSPEEEVSSATSSSLKCSS-MLKDLDTPHREKDVLSPAQDLLSKP
western_Europea 285 YD--------IDFVPPSPEEEIISSSSSSLHCIS-ILKDLDTSEGKKD----TAEDLSKP
African_savanna 297 YD--------RDFVPPSPEE-IISSTSSSLKCFS-MVKDLDTSDRKKDVLSTSEDLLSKP
cattle_1-1430 296 YD--------IDFVPPSPEEGMTSTSSSSLKCSS-MLKDLDTSDKKHSL--STSEDLSKL
Philippine_tars 292 YN--------IDFVPPSPEEEINSTTSSSLKCFS-MLKDLDTSEEKKDVLSTSEDLLSAP
white-tufted-ea 295 YD--------IDFVPPSPEE-IISASSSSSKCFS-VSKDLDTSDRREDVLGTSKDLLSKP
Rhesus_monkey_1 296 YD--------IDFVPPSPEE--ISASSSSSKCLS-MLKDLDTSDRKEDVLSTSKDLLSKP
northern_white- 296 YD--------TDFVPPSPEE-IISASSSSSKCLS-MLKDLDTSDRKEDVLSTSKDLLPKP
Sumatran_orangu 296 YD--------TDFVPPSPEE-IISASSSSSKCLS-MLKDLDTSDRKEDVLSTSKDLLSKP
western_gorilla 294 YD--------TDFVPPSPEE-IISASSSSSKCLS-VLRDLDTSDRKEDVLSTSKDLLSKP
chimpanzee_1-14 296 YD--------TDFVPPSPEE-IISSSSSSSKCLS-MLKDLDTSDRKEDVLSTSKDLLSKP
human_1-1417 296 YD--------TDFVPPSPEE-IISASSSSSKCLS-TLKDLDTSDRKEDVLSTSKDLLSKP

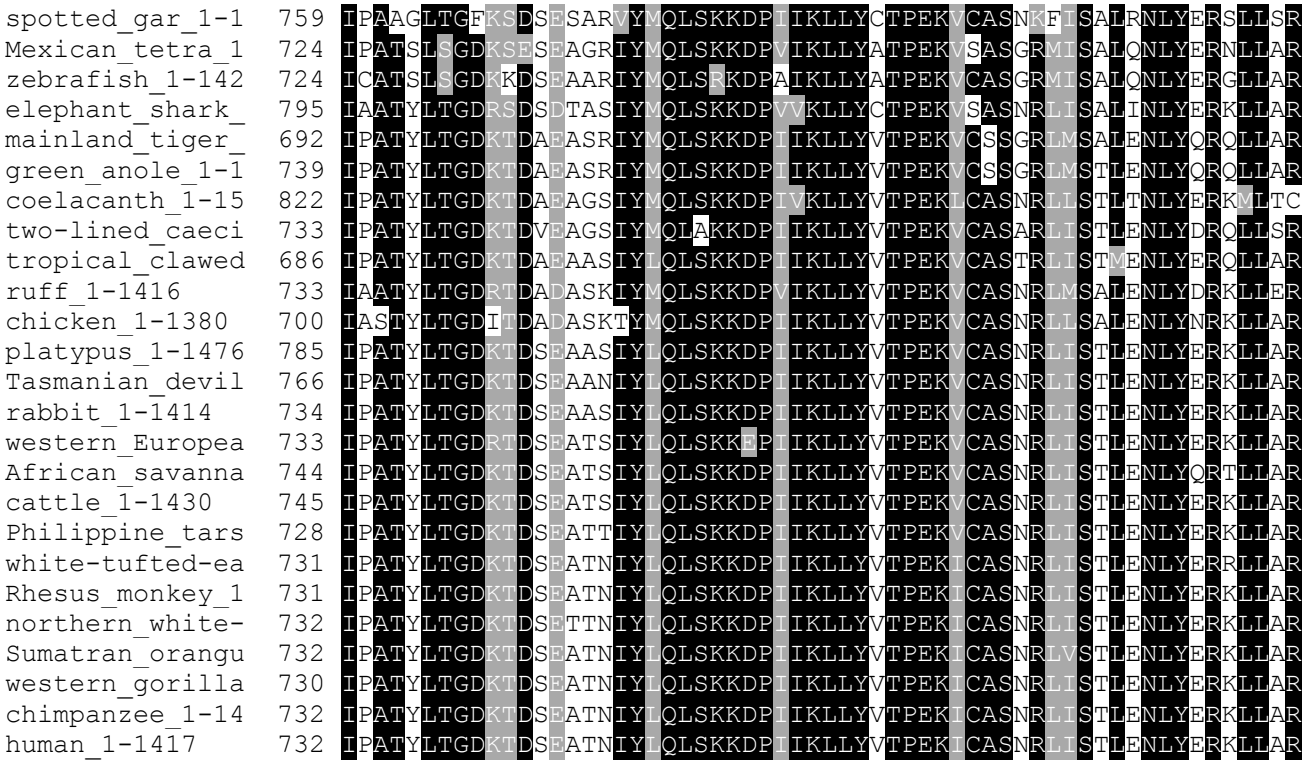

spotted_gar_1-1 374 QKHETA----DDGDSEEI--KISGTDSENQLFSIMEEICQLVDSIPEHELTALSCGRDLL
Mexican_tetra_1 330 SLSGQKVP----------AKDLKGPDDALFSIMESICCLVDKIPEHELIALTCGTELL
zebrafish_1-142 333 APLDQS----------AKGLKGADDALFSVMESICCLVDTIPEHELIALTCGTELL
elephant_shark_ 367 NPLSQERRLSNERACVSD--SSGERGQGTQLFSVMEKICNLVEQIPLTQLQSLACADQLF
mainland_tiger_ 296 EFGESTK----------RSEGRKEGQPLLCVMEDICRLVDSIPERELKSWVGGQELL
green_anole_1-1 349 SRKAQEDSDEGGFSEVQCYLSGKPFPSGQSLYSVMEDICKLVDSIPERDLRFLSCGSELI
coelacanth_1-15 395 NQDILRQP--PQWSNSNR--EAEAFGSQDKLYRVMEEICRLVDTIPLDELKALSCGSQLL
two-lined_caeci 322 AQSLAEEAPGQRRTDAEG--EKDCASQQRALQKVMERICQLLDSIPSAELKALSCGNELL
tropical_clawed 321 ADHLQSQ----------CVSSSLDSKVPSHLLSVMLEICDLVDKIPISELHVLSCGMDLK
ruff_1-1416 324 GPGATAQPVADSDPGAEG--ADKDLPLEQQLFGVMATICELVDAIPLRELQGLSCAKALL
chicken_1-1380 289 ELMAPKQPAAEQDSSAEH--ADKGLHLEQQLYSVMEDICKLVDAIPLHELTSISCAKELL
platypus_1-1476 362 KRTTSEER--KIETAMER--DDKQPPLPQQLLRVMEEMCEMVDAIPICELQALRCGNELL
Tasmanian_devil 344 EKTTSPQL--LQETSTKY--DGKLLTLQQQLFHVMEDICSLVDTIPTYQLEGLDCGKKLL
rabbit_1-1414 349 EKLTPQEP--DPETTAGC--EARPSPVQRRLLQVMEHICVLVDAIPAEALRALEGGSELL
western_Europea 332 EQTTSQQF--TQDTSTDN--NARQTSLQQQLICVMEHICNLVDAIPAGELRALGCGDELL
African_savanna 347 EKMTAQQP--NQETSTDC--DAERLSLQQQLIRVMEDICKLVDTIPADELETLDCGNELL
cattle_1-1430 345 KEMATQQP--NQETSTSC--EARQTSLQQQLIRVMDRICKLVDAVPDDELKALDCGNELL
Philippine_tars 343 ENVPTQKP--HPEASTDC--DAGQKSLQQQLIQVMDRICKLVDTIPADKLIALDCGNELL
white-tufted-ea 345 EKMTMQEP--NPETSTDC--DARRISLQQQLIHVMDRICKLVDIIPADKLKLLDCGNELL
Rhesus_monkey_1 345 EKMSTQEP--NPETSTDC--DARQINLQQQLIHVMDHICKLIDNIPADKLKLLDCGNELL
northern_white- 346 EKMSMQDP--NPETSTDC--DARQISLQQQLIHVMEHICKLIDTIPDDKLKLLDCGNELL
Sumatran_orangu 346 EKMSMQEP--NPETSTDY--DVGQISLQQQLIYVMEHICKLIDTIPDDKLKLLDCGNELL
western_gorilla 344 EKMSMQEL--NPETSTDC--DARQISLQQQLIHVMEHICKLIDTIPDDKLKLLDCGNELL
chimpanzee_1-14 346 EKTSTQEL--NPETSTDC--DARQISLQQQLIHVMEHICKLIDTIPDDKLKLLDCGNELL
human_1-1417 346 EKMSMQEL--NPETSTDC--DARQISLQQQLIHVMEHICKLIDTIPDDKLKLLDCGNELL
spotted_gar_1-1 428 FQRARRRRFFAESCRTPQAPRWTGS-HEASNAKTDNQWSHFKTPV-------SHSSSFKT
Mexican_tetra_1 378 LQRAHRKRILAEIASSGKSVSVSGA---ACDFRL-LHRGSFGVPT--------------P
zebrafish_1-142 379 LQRAHRKRILAKGGSSRTSHSDSVS-----------TPGFLNRPT--FGVTPSNLTSLTP
elephant_shark_ 425 HYRALRKQILANQSEMPIVQNRTCSAVTAMPNHNASQNNIPTYPV----------TPAVT
mainland_tiger_ 343 RLRDCRRKLLSNPRNGNRTGADTFL---DINCKHNFNYDGSHDGSYSSLLPQGGASTPGP
green_anole_1-1 409 RHRDYRRQLLNNL-TNGSGRDAGIPL--DISSKALFRYDGSLDSG-------VSFRLEEE
coelacanth_1-15 451 QQRQHRRAIISGPNEKLLTNQNEPSLSSVPTLAASVHKRSLDPTP------PSSFPTDQM
two-lined_caeci 380 QQRDLRRQLLADWTEGDGSSSDSGIHRVLGISSLSRGSQCAGYNS--------PLPPLHP
tropical_clawed 371 KKRDMRKRILLNNSVLRSSPADS-----STVSLTSCPSQNMEFDV----------SAPKD
ruff_1-1416 382 QQRDLRSRLLAGSVTLNKSAMTKGF---PGSWKGCVEQSSPAPPA----PGSGPTWSFSL
chicken_1-1380 347 QQRELRRKLLADSGALNTNSVNGPR-----NWKACVQQDPSSRPG--TPLCSGPGRGVSS
platypus_1-1476 418 RKREHRKRLLDNYFAGSDVSDISSLL--ASHWK--SSPIGCRRPG---PPSLSSPPSFQK
Tasmanian_devil 400 QHRDLRRKLLADADLNKNGASSSPL-----NWT--YQPNSHDQSV-NGKTVFSSLHSLQT
rabbit_1-1414 405 RQRDTRRKLLAEADFTRNGASLLGS-----LWE--VGPDSLGSPV--------------D
western_Europea 388 QQRDIRRKLLAEVDLGTCDASVPDA-----VWR--CRPRPHGSPVEDVVSASSSPFAFSG
African_savanna 403 QQRNIRRRLLADAGFNRSDAGFLGS-----LWR--CRPNLLDIPV--------EGVAFSQ
cattle_1-1430 401 QQRDIRRKLLAEADFDPSDASVLSA-----MWG--CRPDPLESPV-KDLASVSLPYAFRK
Philippine_tars 399 QQRNIRRKLLAEADCNTNDASLLGW-----IWR--RRPDSLCSPV--------------E
white-tufted-ea 401 QQRNIRRKLLTEVDLNKSDASLLGP-----VWR--CRPDSLDEPV--------------E
Rhesus_monkey_1 401 QQRNIRRKLLTEVDFNKSDASLLGS-----VWR--CRPDSLDGPV--------------E
northern_white- 402 QQRNIRRKLLTEVEFNKSDASLLGL-----MWR--CRPDSLDGPM--------------E
Sumatran_orangu 402 QQRNIRRKLLTEVDFNKSDASLLGS-----MWR--CRPDSLDGPM--------------E
western_gorilla 400 QQRNIRRKLLTEVDFNKSDASLLGS-----MWR--CRPDSLDGPM--------------E
chimpanzee_1-14 402 QQRNIRRKLLTEVDFNKSDASLLGS-----MWR--CRPDSLDGPM--------------K
human_1-1417 402 QQRNIRRKLLTEVDFNKSDASLLGS-----LWR--YRPDSLDGPM--------------E

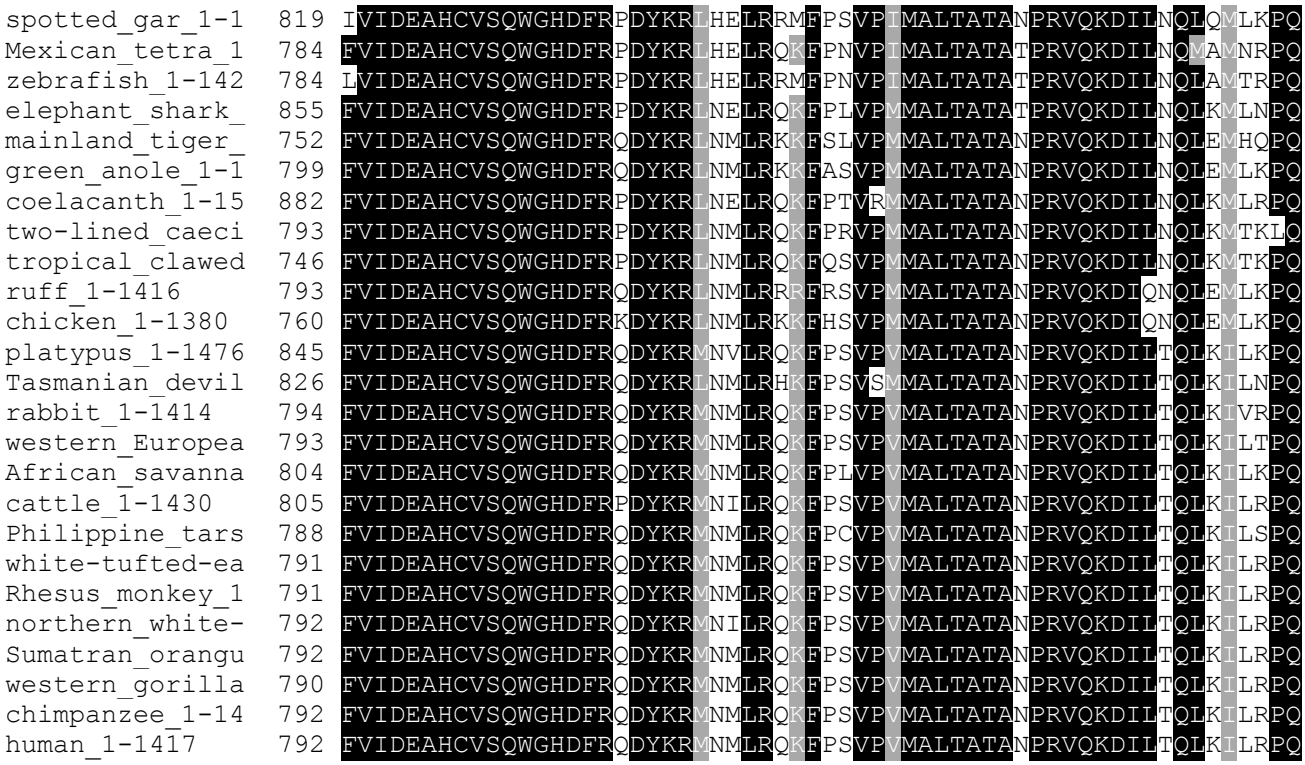
```
spotted_gar_1-1  480 IPSGKDQSSAKSFRFRKPSSVPFA-------DGDG--SV-----FEE-------------
Mexican_tetra_1  420 LISGKGDSALKGFQYRKSFASTMS-------VDYN--NS-----VFEDSDCII-----NG
zebrafish_1-142  426 VTSGKREGVKTGFSFRKSIASVMS-------VGDE--SV-----FEDSDCIING------
elephant_shark_  475 LNPSQELNSSKTFTFRKSFSTGFGGDGRLSDRGDC--SL-----LNQTDDRT--------
mainland_tiger_  400 FSGKRLGDVAPLHGLHPPPAQGWA-------SPDP--SVSVKGSTWDTS-----------
green_anole_1-1  459 NVDSPPGGMSTSTQPSVRKFGVFE-------RHHN--FH-----S-------------TP
coelacanth_1-15  505 LSSSAEGTSTKAFKFKKSSSTSFS-------PASC--GK-----SFFSDADF--------
two-lined_caeci  432 VCRDSHASPFTRSGVSDSPSHLFH-------VGNP--SA-----AKHTD-----------
tropical_clawed  416 SENFSGSSVNKIFKFNKLVADEFG-----------------------------------
ruff_1-1416      435 GRNSPQSPNLPSALSGTVPSSHFS-------TKKN--QT-----LDTPCASKPST---EE
chicken_1-1380   400 VGSTPKSTNLPSVLSRTVNSSSFS-------TIRN--QT-----LDKLDTSYSSKETDQE
platypus_1-1476  471 TDFRPPGNSAEDFKLPHLLCAASA-------AEHR--GSGLGQTPRELPNLDTSVNRTPE
Tasmanian_devil  452 SKSFFTGNLVKEFKSLHLPSNPVS-------TIDCQGNFFDKTSRTFPNLDVSTGMHDIQ
rabbit_1-1414    444 GESCSVGDSVKAPTFPHLPSNSLS-------NGKC--L----------------------
western_Europea  441 GKYCPTGRSAKEFNSPHLALTPLP-------ASES--LL----------------------
African_savanna  448 GDFCPTENSVQELNFPHLPSNSIS-------TGTC--LP----------------------
cattle_1-1430    453 GDSCLAGNAVKEFTLPHLPSNSMS-------TRER--LL----------------------
Philippine_tars  438 GDSYPREISMKELNSSHLSLDSVS-------PGEC--LL----------------------
white-tufted-ea  440 GDSCPTENSMKGLNFLHLSSDSVS-------PGEC--LL----------------------
Rhesus_monkey_1  440 GDSCSAGNSMKELNFSHLPSNSVS-------PGDY--LL----------------------
northern_white-  441 GDSCPTRNSMKELNFSHLPSNSVS-------PGDC--LL----------------------
Sumatran_orangu  441 GDSCPTGNSMKELNFSHLPSNAVS-------PGDC--LL----------------------
western_gorilla  439 GDSWPTGNSMKELNFSHLPSNSVS-------PGDC--LL----------------------
chimpanzee_1-14  441 GDSCPTGNSMKELNFSHLPSNSVS-------PGDC--LL----------------------
human_1-1417     441 GDSCPTGNSMKELNFSHLPSNSVS-------PGDC--LL----------------------

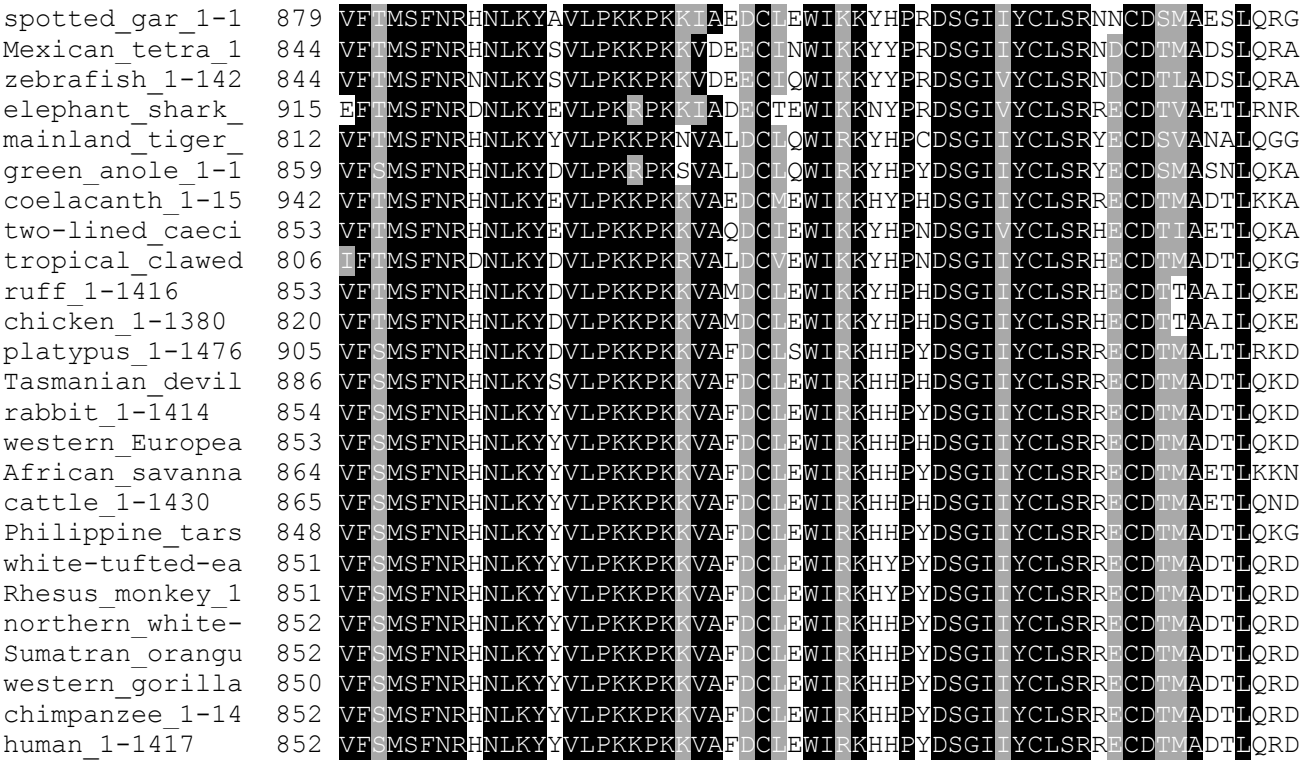

spotted_gar_1-1  513 SDLSDTGGIQSP--------SPSYRPRGVPNVSFNGESDSRKEKLGVGNNDT------PS
Mexican_tetra_1  461 VETPGGSWNPNPSLKSAAK-GDQENTCLKKNNVHIESSSLLLKQTDDQTEDL----LQTN
zebrafish_1-142  466 VETPGGTWNPNSSTKISAG-RDTFNGSIQTLSKPESKTDKCYSRLSFNESSNQTDLFYSP
elephant_shark_  520 SDIPVRSCASSFTCSQGFQSKPRLNT--GKLDPYTGDHGTLPTDTPRSSLSSFA--YVAP
mainland_tiger_  440 SPAVERSGASGPRREAGAI-AGLQNPNLEA----ETRW-WPPTLNGTSSESR----RKPA
green_anole_1-1  492 SVVPAVQGTSNVSL-----------------SRFPGNSALENGISPLQNASL----SFPK
coelacanth_1-15  543 SIASELQTPELSGKSHSLS-FTWLDPQKDQLSFTSGDYENTPINVFQSKASS----FKNN
two-lined_caeci  467 LTSIGTAPDLQLPHSKGDA-KGPFNFHFKA---LSCSSLHKNQRIGGEEEEEEDNCYFPE
tropical_clawed  440 TKGNDRASDFAPNFREKMENKTSFSLQAGGDSVMEDSFSFQSNVLSTSQFNT------PQ
ruff_1-1416      478 VTCPGTAAPSS-SRGHGGD-RASLSPHPET-SFHSSSCGEPGLRNG---ESC----LLLE
chicken_1-1380   446 VICLEPAALPSPKVNG----KGSTSLSRPSEASFNGSWCEKP--TGRDSGNW----RVPE
platypus_1-1476  522 TNAAGKPGWPS-SRQDASE-RTFFNSPSEEVSLVSSGWGDRS-RTERKDERR----HFPE
Tasmanian_devil  505 RTTPEKSKLSS-AKRNISE-SILLSSQSPE-PFISSNSAEIS-RTGEKCKRS----HLLG
rabbit_1-1414    473 TTPPGKTEFSA-PTKHLSE-GPSFSSHLHK-SFVSGNWADTP-RTGQRNGRC----YFPE
western_Europea  471 TITPAKTGFSATTTKTLPE-RPLLGSHLQK-SLVSDSWAETP-STDQRSERS----YFPG
African_savanna  478 TATPGKTGFLA-TPKNLSE-RPLFSSQLQR-SFVSSSCAVTPGREERNGRSS----YFPG
cattle_1-1430    483 TPPLGKTGVLA-TTKSPFE-RPLFSSHLQK-SFVSSNWAETP-RAEKRNESP----DFPG
Philippine_tars  468 TTTPGKTGFSA--RKDLFE-RPLFNSHLQE-SFVSSNWAETP-RLQKRNESS----YFPG
white-tufted-ea  470 TTTPGKTGFSA-TRENLFE-RPLFNTHLQK-SFVSSNWAETP-RLEKRNESP----YFPG
Rhesus_monkey_1  470 TTTLGKTGFSA-TRENLFE-RPSFNTPLQK-SFVSSNWAETP-RLGKKNESS----YFPG
northern_white-  471 TTTLGKTGFSA-TRENLFE-RPLFNTHLQK-SFVSSNWAETP-RLGKKNESS----YFPG
Sumatran_orangu  471 TTTLGKTGFSA-TRKNLFE-RPLFNTHLQK-SFVSSNWAETP-RLGKKNESS----YFPG
western_gorilla  469 TTTLGKTGFSA-TRKNLFE-RPLFNTHLQK-SFVNSNWAETP-RLGKKNESS----YFPG
chimpanzee_1-14  471 TTTLGKTGFSA-TRKNLFE-RPLFNTHLQK-SFVSSNWAETP-RLGKKNESS----YFPG
human_1-1417     471 TTTLGKTGFSA-TRKNLFE-RPLFNTHLQK-SFVSSNWAETP-RLGKKNESS----YFPG
```

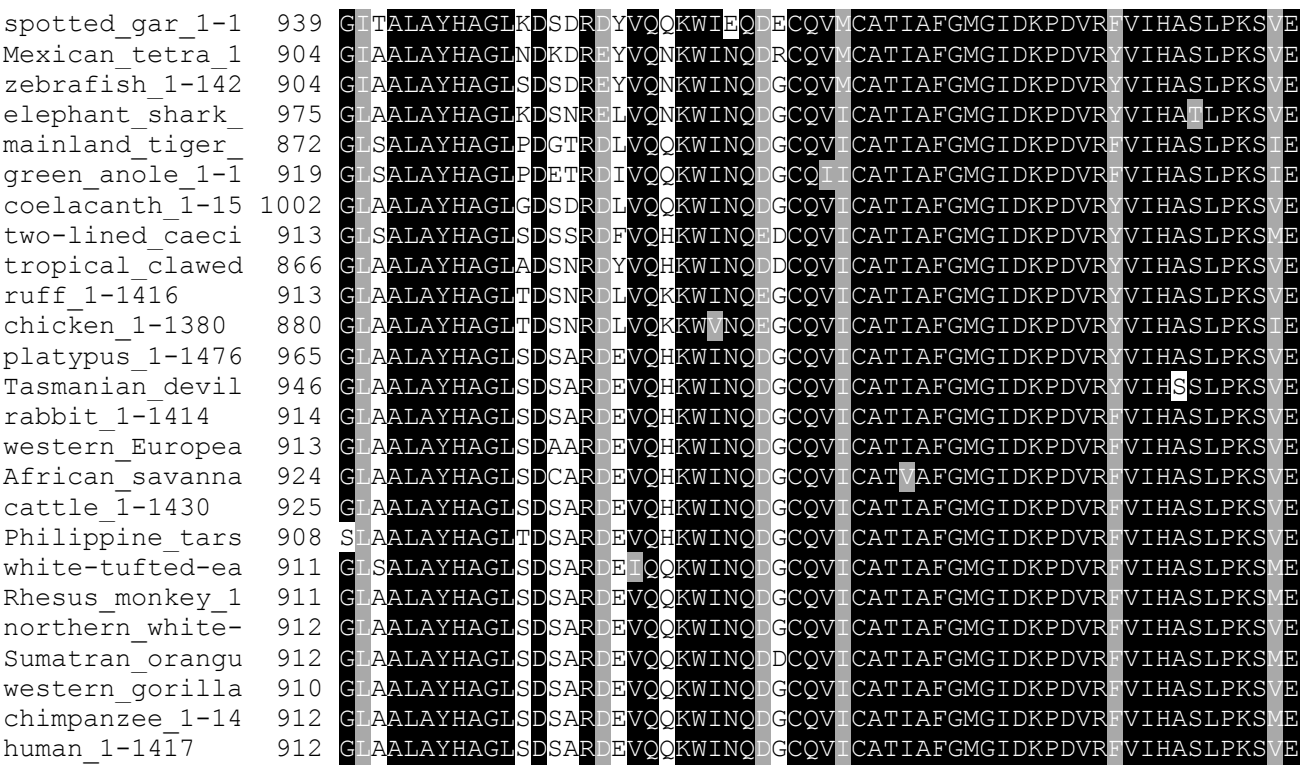

spotted_gar_1-1 559 LRIQN-KTNTPASD----------SYGVDPSLDDIDNE----------FTFDNFDIDD
Mexican_tetra_1 516 LFFSPKKRVGTIKK----------NKEDKVENNTAGSVYFDSEMDTVPDDFYDDDFDIDD
zebrafish_1-142 525 KRVDS----GS----------RNADSSVEINIAGSSSLRTGAEPVDDFLLDDFDIDD
elephant_shark_ 576 RSLEYSRLSSPSTG----------FAGLQARPFADGGGSHVNAGTPVDESYDIESFDIDD
mainland_tiger_ 490 APPLV----------SNGGSESPWATGGQK----------LHQDDFDMED
green_anole_1-1 531 RGSSG----ERMWS----------ENFMNPEMNVTAESDSLANR---AGNFDDDNFDIDD
coelacanth_1-15 598 LYKTPRTENNGSWCSLEIPVTSTIQNQSNNCKTDLDGRNPPLTEDADG--FDVDDFDIDD
two-lined_caeci 523 IPSTS----TNVKC----------QTQLNSSKEVVAGMEAMGAREME---FDLDNFLTDC
tropical_clawed 494 NEKPV----SSSTC----------ARPISQPIDDMDNPDLD---------FDIDNFDIED
ruff_1-1416 528 RPLTS----TALRD----------QGRAPTSSSEQDALGLNHTD------FDLDHFDIDD
chicken_1-1380 496 RPTAS----TALKA----------QHTAPAGNPASGCWDVNDTD------FDLDHFDIDD
platypus_1-1476 575 RALTS----TAVKI----------PKKHPAATPAPSDLDRGVSLEAQDPIFDLDAFDMDD
Tasmanian_devil 557 NFLTS----TTIKD----------QNKHINSVNDLKRNNFSET-------QDIDTFDIDD
rabbit_1-1414 525 NVLTS----TAVRD----------QSKHRASASDTARDARAA--------CDVDDFDIDD
western_Europea 524 KVLTS----TAVKD----------QNKHNASMNDSETELQAC--------CDIDNFDIDG
African_savanna 531 SVLTS----TAVKD----------PHKHTASTNDLERREVSEIQAS----YDIDHFDIDD
cattle_1-1430 535 NVLTS----TAVKD----------QNKYTTSINDLERELQAA--------CEIDNFDIDD
Philippine_tars 519 NVLTS----TAVKD----------QNKHTASISDLERETQAS--------YDIDNFDIDD
white-tufted-ea 522 NVLTS----TAVKD----------WNKHTASINDLERESQPS--------CDIGNFDIDD
Rhesus_monkey_1 522 NVLTS----TAVKD----------QNKHTASINDLERETQPS--------YDIDNFDIDD
northern_white- 523 NVLTS----TAVKD----------QNKHTASINDLERETQPS--------YDIDNFDIDD
Sumatran_orangu 523 NVLTS----TAVKD----------QNKHTASINDLERETQPS--------YDIDNFDIDD
western_gorilla 521 NVLTS----TAVKD----------QNKHTASINDLERETQPS--------YDIDNFDIDD
chimpanzee_1-14 523 NVLTS----TAVKD----------QNKHTASINDLERETQPS--------YDIDNFDIDD
human_1-1417 523 NVLTS----TAVKD----------QNKHTASINDLERETQPS--------YDIDNFDIDD
spotted_gar_1-1 596 FDEEDMGDYFEGNDS--ASKKQGTEGVQPVREGH----PSKSLWEK-----KSPSSSFSS
Mexican_tetra_1 566 FNESDIPKYNEEPVDSLATKKSSSATPKPIYEGG----PSKSCEKRTVTTPTAVVKPTKP
zebrafish_1-142 568 FDDNDIPDYYEEPPSVLESRNNSGVKTPSVQEGG----SSKSFERKTFTPPAPKSIKTPN
elephant_shark_ 626 LDD---EDVPECFLSPQSRSGTGGLSYRTFQ-------PIREETAKPRAFGDKNHTISPS
mainland_tiger_ 520 LDD-----LDDAVSASSARGAGSQGPARPAAGASPA--PKLFLEGNS-VASKEGADFLHS
green_anole_1-1 574 LNE-----LDEIIQL-SGAEQPLQASCLPA--------PASSMGNCSLASDAGRGLLSSA
coelacanth_1-15 656 FYD-DDIVCSDSPSILGAAEKLSTVLVQPIKEGG----PTKSFLEKE---------NSVL
two-lined_caeci 566 LDD--EVQFVACSQNPPATRDLSTLQFPAG--------QEEQIMEPF-VARQPPLSKESQ
tropical_clawed 531 FDD----QLMDSPAA-PPVSSKNAPQYPTICETQFDSGSKEKVTRHN-VGDKTNTELSDS
ruff_1-1416 568 FDE----DWENSVNV-SVPETPSTPLYQPVREGP----PARSLSSKILSRAKVSAAASNP
chicken_1-1380 536 FDE----GWEEAVAPEAAPEAPPAPQWQPLREGS----A--SLRCRLLAAAAGSAPGPHP
platypus_1-1476 621 FDE----DWEN------PAGPSPVGAYPPIREGQ----PAKSASEKT-HPTEGNRLPVFS
Tasmanian_devil 596 FDDFDDYDWENPSYN--SVEEKSKNMYLPIREGQ----PLKSSSERT-PLSKTSCLSVVP
rabbit_1-1414 563 FDD--DDDWEDIMHS-LAANKSSTAAYRPIREGG----PVRSASERI-SSAKTSCLPMVS
western_Europea 562 FDD--DDDWEDIMHN-LTANKSSTAAYHPIKEGR----PVKSVSERI-SSAKTNCLPVIA
African_savanna 573 FDD--DDDWENIMQN-LTASKSATAAYPPIKEGR----PIKSVAERI-SSAKTNCLPVAS
cattle_1-1430 573 FDD-DDDDWENLMQN-LAASKSSTAAYPPIKEGR----PVKSVSERP-SSAKTDCLPGAS
Philippine_tars 557 FDD--DDDWENIMHN-LAASKSSTAAYQPIKEGP----PIKSVTERI-ASAKTNCLPVAS
white-tufted-ea 560 FDD--DDDWEDIMHN-LAASKSSTAAYQPIKEGR----PIKSVSEQL-SSAKTNCLPVAS
Rhesus_monkey_1 560 FDD--DDDWENIMHN-LAASKSSTAAYQPIKEGR----PIKSVSERL-SSAKTNCLPVAS
northern_white- 561 FDD--DDDWENIMHN-LAASKSSTAAYQPIKEGR----PIKSVSERL-SSAKTNCLPVAS
Sumatran_orangu 561 FDD--DDDWENIMHN-LAASKSSTAAYQPIKEGR----PIKSVSERL-SSAKTNCLPVAS
western_gorilla 559 FDD--DDDWEDIMHN-LAASKSSTAAYQPIKEGR----PIKSVSERL-SSAKTNCLPVAS
chimpanzee_1-14 561 FDD--DDDWEDIMHN-LAASKSSTAAYQPIKEGR----PIKSVSERL-SSAKTNCLPVSS
human_1-1417 561 FDD--DDDWEDIMHN-LAASKSSTAAYQPIKEGR----PIKSVSERL-SSAKTDCLPVSS

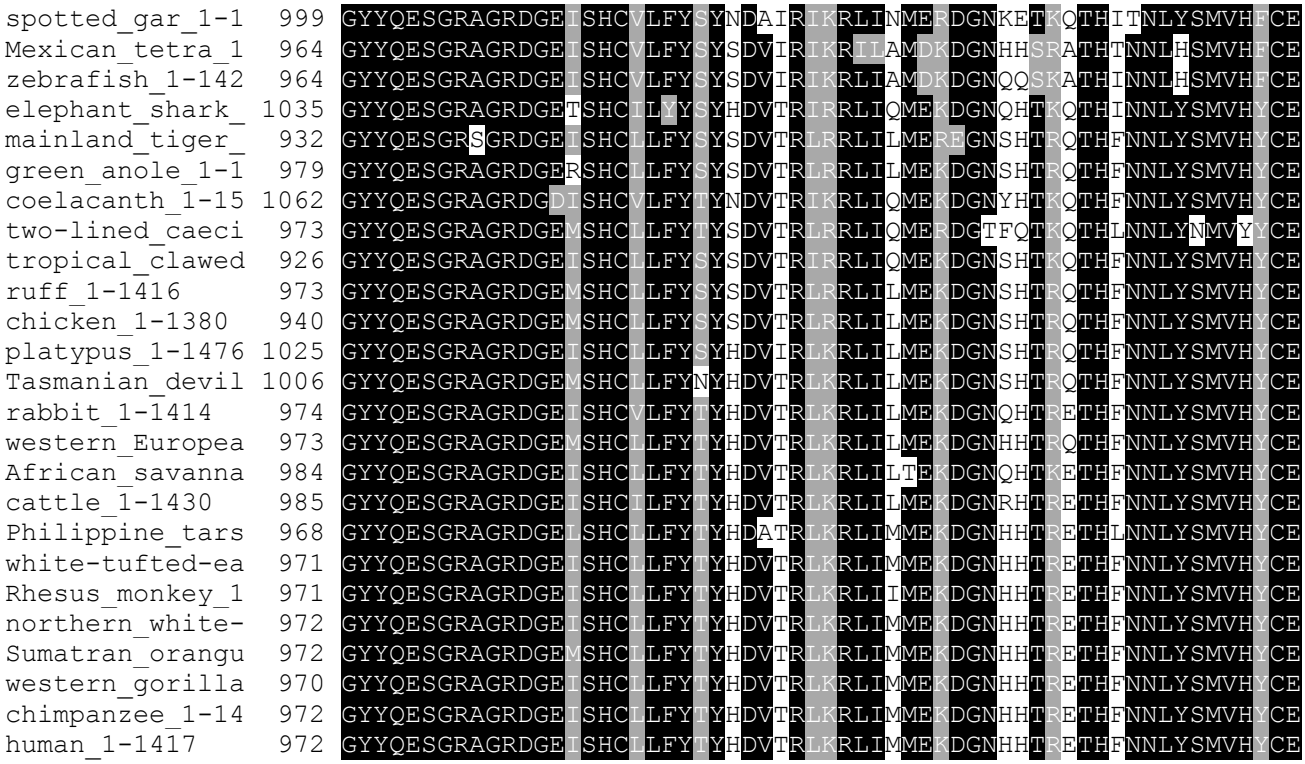

spotted_gar_1-1 645 SAAKPSKPTSECSLPE------PINRNPAHDRFRGCNFPYSQEMMKIFNKRFGLRQFRTN
Mexican_tetra_1 622 ASPEVTY------------------RNPAHDRFRGFSFPHSSEMMKIFHKRFGLHQFRFN
zebrafish_1-142 624 --------------PE------PLYRNPAHDRFRGFNFPHSPEMMKIFHKKFGLHQFRFN
elephant_shark_ 676 AAASVKRSQMGRSPADTLF-ERNLPRNPALERFRGNDFPHSKEMMTVFRKKFGMHQFRTN
mainland_tiger_ 572 NFSFTDLSVSGKSCPVPPLIKDTPSRNAALFPFRGLAFPHSNEMIKIFHKKFGLHHFRTN
green_anole_1-1 620 VLPRANQQVHAKTGLVRMV-ESPSARNPSLERFRGCSFPHSNEMMKVFHKRFGLHHFRTN
coelacanth_1-15 702 NGKKNSLPQCGGVMPNRGFSDKPSFLNPSHDHFRGFQFPHSKEMMKIFHKKFGLHQFRTN
two-lined_caeci 615 PASASNASAPLKHCSDLSL--KKSSLSSVHERFRGFSFPHSKEMMKIFHKKFGLHHFRTN
tropical_clawed 585 L------------------LKPQIVNPALERFKSFNFPHSKEMMKIFHKKFGLHRFRTN
ruff_1-1416 619 PAPKSSFLMATKNHSD------PPATNPALDRFRGTKFSHSEGMMNIFHKKFGLHYFRTN
chicken_1-1380 586 TAPKSGCGISAKSSSE------PLVHNPAHERFRGMKFSHSEEMIKIFHRKFGLHSFRTN
platypus_1-1476 666 AAKNTHFSGSVQNCTDGWV-QNSPSRNAEHERFKRIDFAHSKEMMKIFHKKFGLHHFRTN
Tasmanian_devil 649 ANQNSSLS--IQNFSDKLV-QNKPSRNPEQERFKSLDFPHSKEMMKIFHKKFGLHHFRTN
rabbit_1-1414 615 TAHRRHISESVQNDADKPA-QSLASGNLKHERFQSLNFPHTKDMMKIFHKKFGLHNFRTN
western_Europea 614 TTQSKNFSESIQNYPDKST-QNLASKNLKHEHFQSLSLPHTKEMMKIFHKKFGLHHFRTN
African_savanna 625 AAQNKNFSESTQNYTDKPE-QNFASRNLKHEHFQSLDFPHTKEMMKIFHKKFGLHNFRTN
cattle_1-1430 626 TTQNKNFSESIPNYPDKME-QNLASRNPKHERFQSLNFPHTKEMMKIFHKKFGLHNFRTN
Philippine_tars 609 TAQNINSSESIQSYTDKSA-QNLASRNLKHERFQSLSFPHTKEMMKIFHKKFGLHNFRTN
white-tufted-ea 612 TAQNKNFSESMQNYTDKST-QNLASRNLKHERFQGLSFPHTKEMMKIFHKKFGLHNFRTN
Rhesus_monkey_1 612 TAQNINFSESIQNYTDKSA-QNLASRNLKHERFQSLSFPHTKEMMKIFHKKFGLHNFRTN
northern_white- 613 TAQNINFSESIQNYTNKSA-QNLASRNLKHERFQSLSFPHTKEMMKIFHKKFGLHNFRTN
Sumatran_orangu 613 TAQNINFSESIQNYTDKSA-QNLASRNLKHERFQSLSFPHTKEMMKIFHKKFGLHNFRTN
western_gorilla 611 TAQNINFSESIQNYTDKSA-QNLASRNLKHERFQSLSFPHTKEMMKIFHKKFGLHNFRTN
chimpanzee_1-14 613 TAQNINFSESIQNYTDKSA-QNLASGNLKHERFQSLSFPHTKEMMKIFHKKFGLHNFRTN
human_1-1417 613 TAQNINFSESIQNYTDKSA-QNLASRNLKHERFQSLSFPHTKEMMKIFHKKFGLHNFRTN
spotted_gar_1-1 699 QLEAINATLHCEDTFVLMPTGGGKSLCYQLPACVSPGVTIVISPLRSLIVDQVQKLTTID
Mexican_tetra_1 664 QLEAINATLLGEDTFVLMPTGGGKSLCYQLPACVSAGVTVVISPLRSLIVDQVQKLTTID
zebrafish_1-142 664 QLEAINASLLGEDTFVLMPTGGGKSLCYQLPACVSAGVTVVISPLRSLIVDQVQKLTTID
elephant_shark_ 735 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACLTPGVTVVISPLRSLILDQNQKLALFN
mainland_tiger_ 632 QLEAINAALLGQDCFILMPTGGGKSLCYQLPACVSVGVTIVISPLRSLIVDQVQKLTSMD
green_anole_1-1 679 QQEAINAALLGEDCFVLMPTGGGKSLCYQLPACILVGVTIVISPLRSLIVDQVQKLTSMD
coelacanth_1-15 762 QLEAINCTLLGEDCFILMPTGGGKSLCYQLPACLSPGVTIVVSPLRSLIVDQVQKLTTID
two-lined_caeci 673 QLEAINASVFGEDSFILMPTGGGKSLCYQLPACISPGVTIVISPLRSLIIDQVQKLTSID
tropical_clawed 626 QLEAINACLLGEDCFILMPTGGGKSLCYQLPGCVSPGVTIVISPLRSLIVDQVQKLTSID
ruff_1-1416 673 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACVSAGVTIVISPLRSLIIDQVQKLKTID
chicken_1-1380 640 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACVSAGVTVVISPLRSLIIDQVQKLKTID
platypus_1-1476 725 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACISPGVSIVISPLRSLIVDQVQKLTSID
Tasmanian_devil 706 QLEAINAVLLGEDCFVLMPTGGGKSLCYQLPACVSPGVTIVISPLRSLIIDQVQKLTSID
rabbit_1-1414 674 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACISPGVTIVISPLRSLIVDQVQKLTSID
western_Europea 673 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACVSPGVTIVISPLRSLIVDQVQKLTSID
African_savanna 684 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACIAPGVTIVISPLRSLIVDQVQKLTSID
cattle_1-1430 685 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACVSPGVTIVISPLRSLIVDQVQKLTSLD
Philippine_tars 668 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACISPGVTVVISPLRSLIVDQVQKLTSLD
white-tufted-ea 671 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACVSPGVTIVISPLRSLIVDQVQKLTSID
Rhesus_monkey_1 671 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACVSPGVTIVISPLRSLIVDQVQKLTSID
northern_white- 672 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACVSPGVTIVISPLRSLIVDQVQKLTSID
Sumatran_orangu 672 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACVSPGVTIVISPLRSLIVDQVQKLTSID
western_gorilla 670 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACVSPGVTIVISPLRSLIVDQVQKLTSID
chimpanzee_1-14 672 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACVSPGVTVVISPLRSLIVDQVQKLTSID
human_1-1417 672 QLEAINAALLGEDCFILMPTGGGKSLCYQLPACVSPGVTVVISPLRSLIVDQVQKLTSID

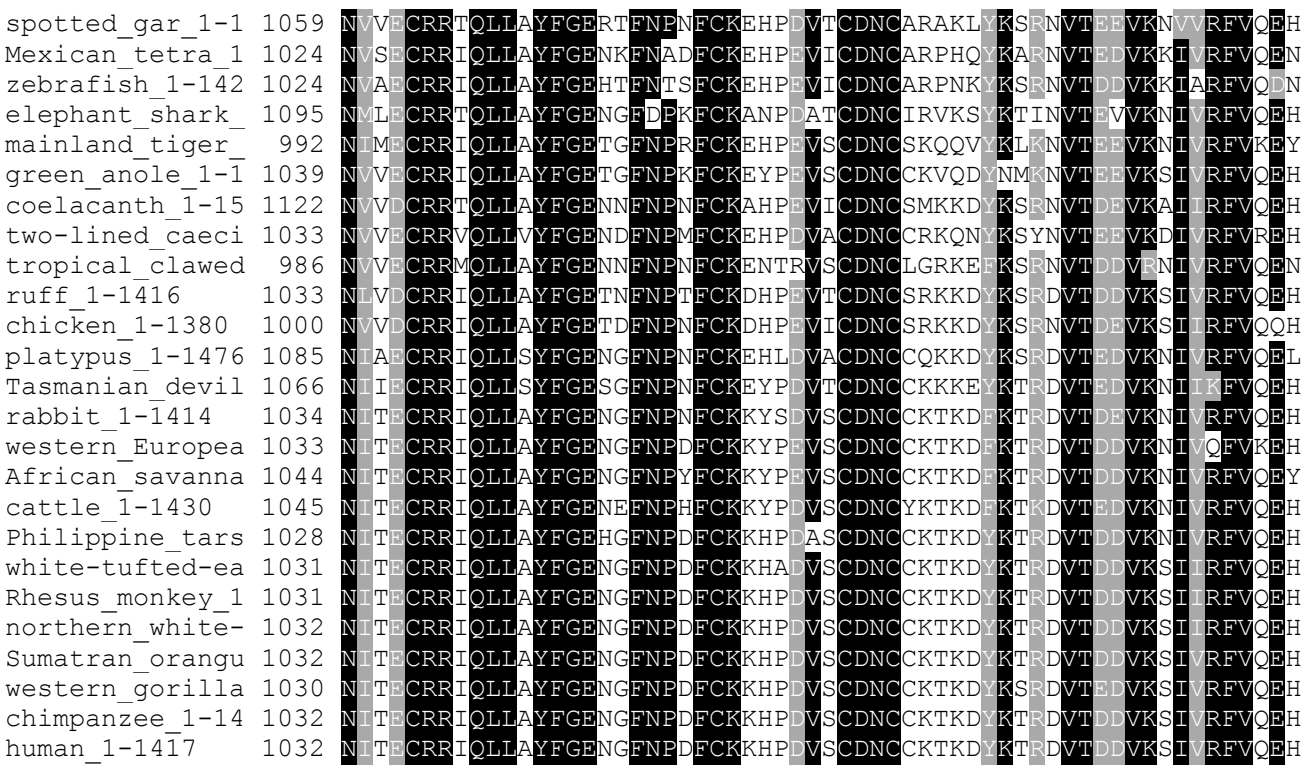

spotted_gar_1-1 759 IPAAGLTGFKSDSESARVYMQLSKKDPIIKLLYCTPEKVCASNKFISALRNLYERSLLSR
Mexican_tetra_1 724 IPATSLSGDKSESEAGRIYMQLSKKDPVIKLLYATPEKVSASGRMISALQNLYERNLLAR
zebrafish_1-142 724 ICATSLSGDKKDSEAARIYMQLSRKDPAIKLLYATPEKVCASGRMISALQNLYERGLLAR
elephant_shark_ 795 IAATYLTGDRSDSDTASIYMQLSKKDPVVKLLYCTPEKVSASNRLISALINLYERKLLAR
mainland_tiger_ 692 IPATYLTGDKTDAEASRIYMQLSKKDPIIKLLYVTPEKVCSSGRLMSALENLYQRQLLAR
green_anole_1-1 739 IPATYLTGDKTDAEASRIYMQLSKKDPIIKLLYVTPEKVCSSGRLMSTLENLYQRQLLAR
coelacanth_1-15 822 IPATYLTGDKTDAEAGSIYMQLSKKDPIVKLLYVTPEKLCASNRLLSTLTNLYERKMLTC
two-lined_caeci 733 IPATYLTGDKTDVEAGSIYMQLAKKDPIIKLLYVTPEKVCASARLISTLENLYDRQLLSR
tropical_clawed 686 IPATYLTGDKTDAEAASIYLQLSKKDPIIKLLYVTPEKVCASTRLISTMENLYERQLLAR
ruff_1-1416 733 IAATYLTGDRTDADASKIYMQLSKKDPVIKLLYVTPEKVCASNRLMSALENLYDRKLLER
chicken_1-1380 700 IASTYLTGDITDADASKTYMQLSKKDPIIKLLYVTPEKVCASNRLLSALENLYNRKLLAR
platypus_1-1476 785 IPATYLTGDKTDSEAASIYLQLSKKDPIIKLLYVTPEKVCASNRLISTLENLYERKLLAR
Tasmanian_devil 766 IPATYLTGDKTDSEAANIYLQLSKKDPIIKLLYVTPEKVCASNRLISTLENLYERKLLAR
rabbit_1-1414 734 IPATYLTGDKTDSEAASIYLQLSKKDPIIKLLYVTPEKVCASNRLISTLENLYERKLLAR
western_Europea 733 IPATYLTGDRTDSEATSIYLQLSKKEPIIKLLYVTPEKVCASNRLISTLENLYERKLLAR
African_savanna 744 IPATYLTGDKTDSEATSIYLQLSKKDPIIKLLYVTPEKVCASNRLISTLENLYQRTLLAR
cattle_1-1430 745 IPATYLTGDKTDSEATSIYLQLSKKDPIIKLLYVTPEKVCASNRLISTLENLYERKLLAR
Philippine_tars 728 IPATYLTGDKTDSEATTIYLQLSKKDPIIKLLYVTPEKVCASNRLISTLENLYERKLLAR
white-tufted-ea 731 IPATYLTGDKTDSEATNIYLQLSKKDPIIKLLYVTPEKICASNRLISTLENLYERRLLAR
Rhesus_monkey_1 731 IPATYLTGDKTDSEATNIYLQLSKKDPIIKLLYVTPEKICASNRLISTLENLYERKLLAR
northern_white- 732 IPATYLTGDKTDSETTNIYLQLSKKDPIIKLLYVTPEKICASNRLISTLENLYERKLLAR
Sumatran_orangu 732 IPATYLTGDKTDSEATNIYLQLSKKDPIIKLLYVTPEKICASNRLVSTLENLYERKLLAR
western_gorilla 730 IPATYLTGDKTDSEATNIYLQLSKKDPIIKLLYVTPEKICASNRLISTLENLYERKLLAR
chimpanzee_1-14 732 IPATYLTGDKTDSEATNIYLQLSKKDPIIKLLYVTPEKICASNRLISTLENLYERKLLAR
human_1-1417 732 IPATYLTGDKTDSEATNIYLQLSKKDPIIKLLYVTPEKICASNRLISTLENLYERKLLAR

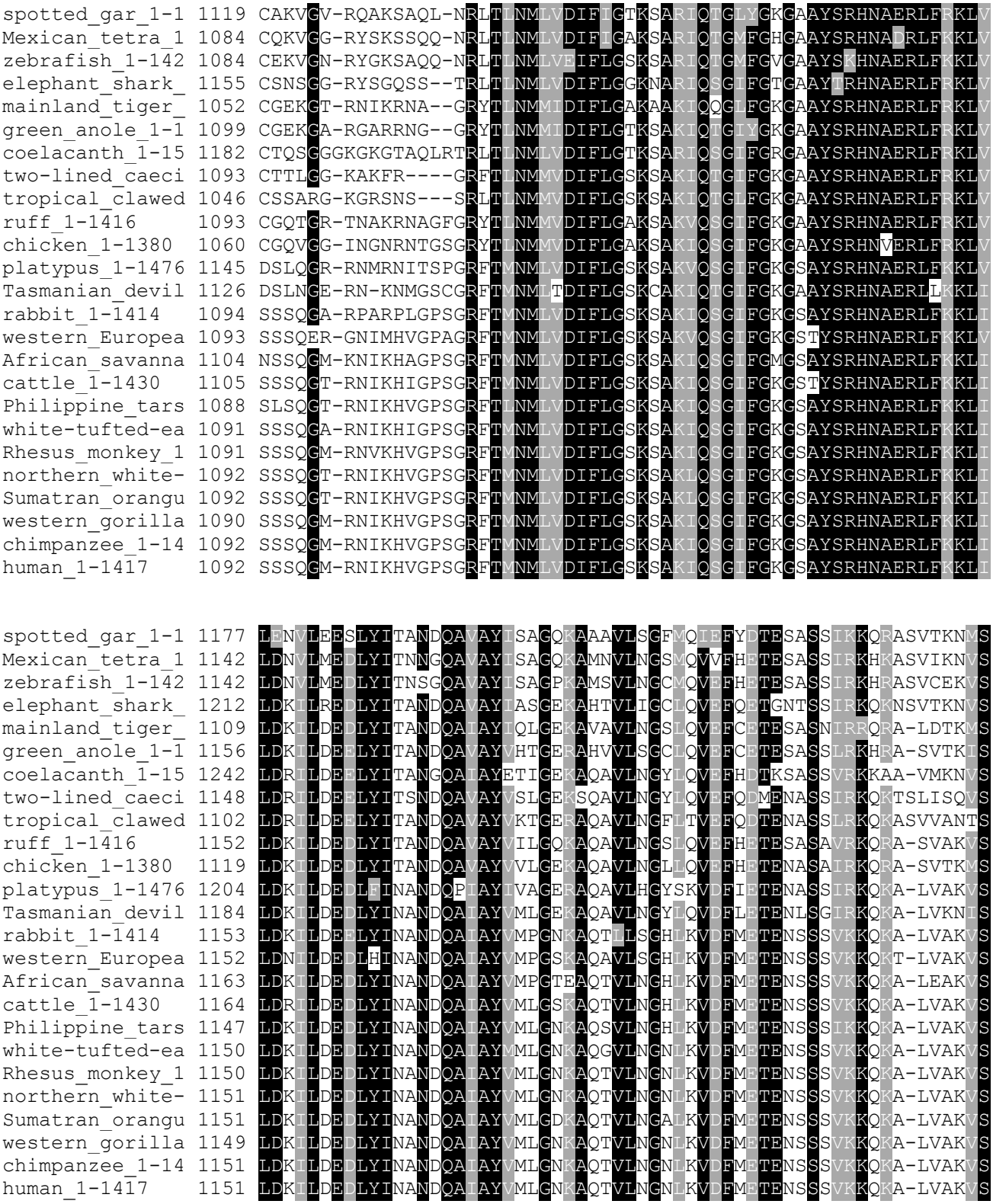

spotted_gar_1-1 819 IVIDEAHCVSQWGHDFRPDYKRLHELRRMFPSVPIMALTATANPRVQKDILNQLQMLKPQ
Mexican_tetra_1 784 FVIDEAHCVSQWGHDFRPDYKRLHELRQKFPNVPIMALTATATPRVQKDILNQMAMNRPQ
zebrafish_1-142 784 LVIDEAHCVSQWGHDFRPDYKRLHELRRMFPNVPIMALTATATPRVQKDILNQLAMTRPQ
elephant_shark_ 855 FVIDEAHCVSQWGHDFRPDYKRLNELRQKFPLVPMMALTATATPRVQKDILNQLKMLNPQ
mainland_tiger_ 752 FVIDEAHCVSQWGHDFRQDYKRLNMLRKKFSLVPMMALTATANPRVQKDILNQLEMHQPQ
green_anole_1-1 799 FVIDEAHCVSQWGHDFRQDYKRLNMLRKKFASVPMMALTATANPRVQKDILNQLEMLKPQ
coelacanth_1-15 882 FVIDEAHCVSQWGHDFRPDYKRLNELRQKFPTVRMMALTATANPRVQKDILNQLKMLRPQ
two-lined_caeci 793 FVIDEAHCVSQWGHDFRPDYKRLNMLRQKFPRVPMMALTATANPRVQKDILNQLKMTKLQ
tropical_clawed 746 FVIDEAHCVSQWGHDFRPDYKRLNMLRQKFQSVPMMALTATANPRVQKDILNQLKMTKPQ
ruff_1-1416 793 FVIDEAHCVSQWGHDFRQDYKRLNMLRRRFRSVPMMALTATANPRVQKDIQNQLEMLKPQ
chicken_1-1380 760 FVIDEAHCVSQWGHDFRKDYKRLNMLRKKFHSVPMMALTATANPRVQKDIQNQLEMLKPQ
platypus_1-1476 845 FVIDEAHCVSQWGHDFRQDYKRMNVLRQKFPSVPVMALTATANPRVQKDILTQLKILKPQ
Tasmanian_devil 826 FVIDEAHCVSQWGHDFRQDYKRLNMLRHKFPSVSMMALTATANPRVQKDILTQLKILNPQ
rabbit_1-1414 794 FVIDEAHCVSQWGHDFRQDYKRMNMLRQKFPSVPVMALTATANPRVQKDILTQLKIVRPQ
western_Europea 793 FVIDEAHCVSQWGHDFRQDYKRMNMLRQKFPSVPVMALTATANPRVQKDILTQLKILTPQ
African_savanna 804 FVIDEAHCVSQWGHDFRQDYKRMNMLRQKFPLVPVMALTATANPRVQKDILTQLKILKPQ
cattle_1-1430 805 FVIDEAHCVSQWGHDFRPDYKRMNILRQKFPSVPVMALTATANPRVQKDILTQLKILRPQ
Philippine_tars 788 FVIDEAHCVSQWGHDFRQDYKRMNMLRQKFPCVPVMALTATANPRVQKDILTQLKILSPQ
white-tufted-ea 791 FVIDEAHCVSQWGHDFRQDYKRMNMLRQKFPSVPVMALTATANPRVQKDILTQLKILRPQ
Rhesus_monkey_1 791 FVIDEAHCVSQWGHDFRQDYKRMNMLRQKFPSVPVMALTATANPRVQKDILTQLKILRPQ
northern_white- 792 FVIDEAHCVSQWGHDFRQDYKRMNILRQKFPSVPVMALTATANPRVQKDILTQLKILRPQ
Sumatran_orangu 792 FVIDEAHCVSQWGHDFRQDYKRMNMLRQKFPSVPVMALTATANPRVQKDILTQLKILRPQ
western_gorilla 790 FVIDEAHCVSQWGHDFRQDYKRMNMLRQKFPSVPVMALTATANPRVQKDILTQLKILRPQ
chimpanzee_1-14 792 FVIDEAHCVSQWGHDFRQDYKRMNMLRQKFPSVPVMALTATANPRVQKDILTQLKILRPQ
human_1-1417 792 FVIDEAHCVSQWGHDFRQDYKRMNMLRQKFPSVPVMALTATANPRVQKDILTQLKILRPQ

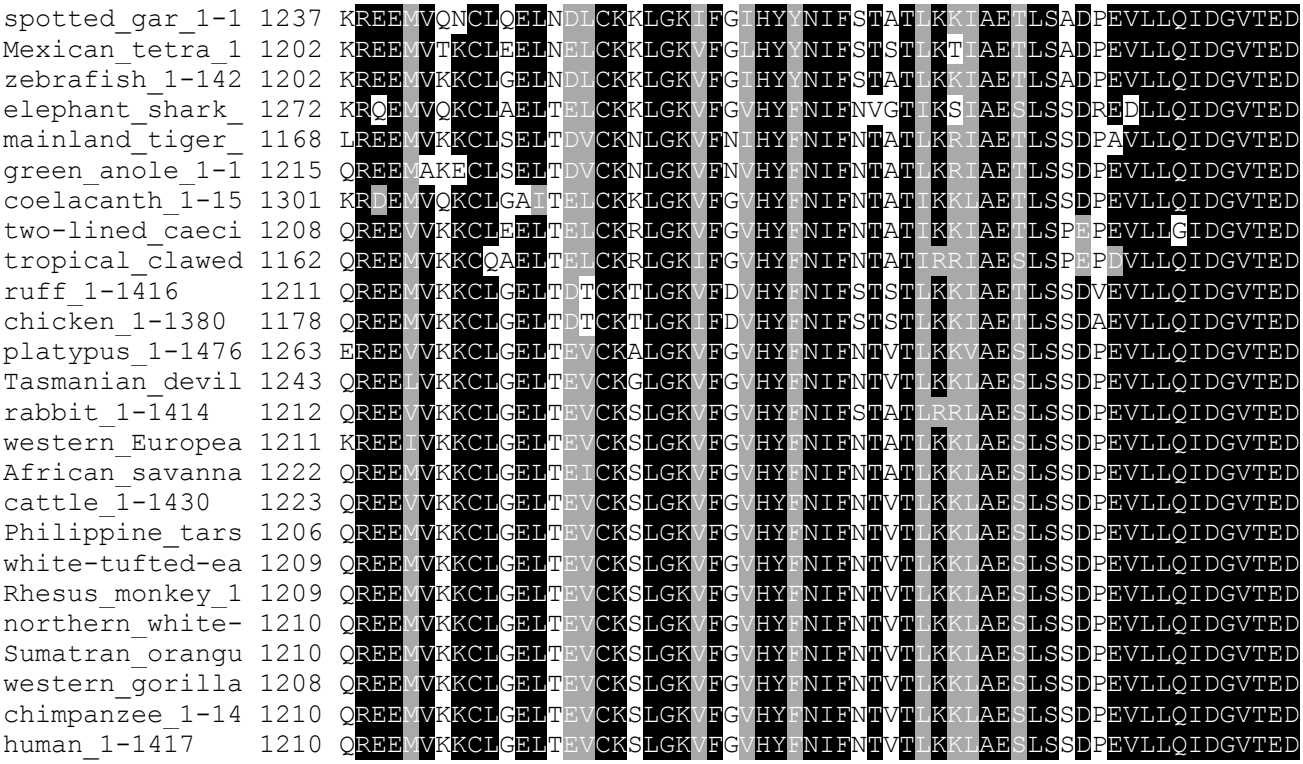

spotted_gar_1-1 879 VFTMSFNRHNLKYAVLPKKPKKIAEDCLEWIKKYHPRDSGIIYCLSRNNCDSMAESLQRG
Mexican_tetra_1 844 VFTMSFNRHNLKYSVLPKKPKKVDEECINWIKKYYPRDSGIIYCLSRNDCDTMADSLQRA
zebrafish_1-142 844 VFTMSFNRNNLKYSVLPKKPKKVDEECIQWIKKYYPRDSGIVYCLSRNDCDTLADSLQRA
elephant_shark_ 915 EFTMSFNRDNLKYEVLPKRPKKIADECTEWIKKNYPRDSGIVYCLSRRECDTVAETLRNR
mainland_tiger_ 812 VFTMSFNRHNLKYYVLPKKPKNVALDCLQWIRKYHPCDSGIIYCLSRYECDSVANALQGG
green_anole_1-1 859 VFSMSFNRHNLKYDVLPKRPKSVALDCLQWIRKYHPYDSGIIYCLSRYECDSMASNLQKA
coelacanth_1-15 942 VFTMSFNRHNLKYEVLPKKPKKVAEDCMEWIKKHYPHDSGIIYCLSRRECDTMADTLKKA
two-lined_caeci 853 VFTMSFNRHNLKYEVLPKKPKKVAQDCIEWIKKYHPNDSGIVYCLSRHECDTIAETLQKA
tropical_clawed 806 IFTMSFNRDNLKYDVLPKKPKRVALDCVEWIKKYHPNDSGIIYCLSRHECDTMADTLQKG
ruff_1-1416 853 VFTMSFNRHNLKYDVLPKKPKKVAMDCLEWIKKYHPHDSGIIYCLSRHECDITAAILQKE
chicken_1-1380 820 VFTMSFNRHNLKYDVLPKKPKKVAMDCLEWIKKYHPHDSGIIYCLSRHECDITAAILQKE
platypus_1-1476 905 VFSMSFNRHNLKYDVLPKKPKKVAFDCLSWIRKHHPYDSGIIYCLSRRECDTMALTLRKD
Tasmanian_devil 886 VFSMSFNRHNLKYSVLPKKPKKVAFDCLEWIRKHHPHDSGIIYCLSRRECDTMADTLQKD
rabbit_1-1414 854 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHHPYDSGIIYCLSRRECDTMADTLQKD
western_Europea 853 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHHPHDSGIIYCLSRRECDTMADTLQKD
African_savanna 864 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHHPYDSGIIYCLSRRECDTMAETLKKN
cattle_1-1430 865 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHHPHDSGIIYCLSRRECDTMAETLQND
Philippine_tars 848 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHHPYDSGIIYCLSRRECDTMADTLQKG
white-tufted-ea 851 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHYPYDSGIIYCLSRRECDTMADTLQRD
Rhesus_monkey_1 851 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHYPYDSGIIYCLSRRECDTMADTLQRD
northern_white- 852 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHHPYDSGIIYCLSRRECDTMADTLQRD
Sumatran_orangu 852 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHHPYDSGIIYCLSRRECDTMADTLQRD
western_gorilla 850 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHHPYDSGIIYCLSRRECDTMADTLQRD
chimpanzee_1-14 852 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHHPYDSGIIYCLSRRECDTMADTLQRD
human_1-1417 852 VFSMSFNRHNLKYYVLPKKPKKVAFDCLEWIRKHHPYDSGIIYCLSRRECDTMADTLQRD

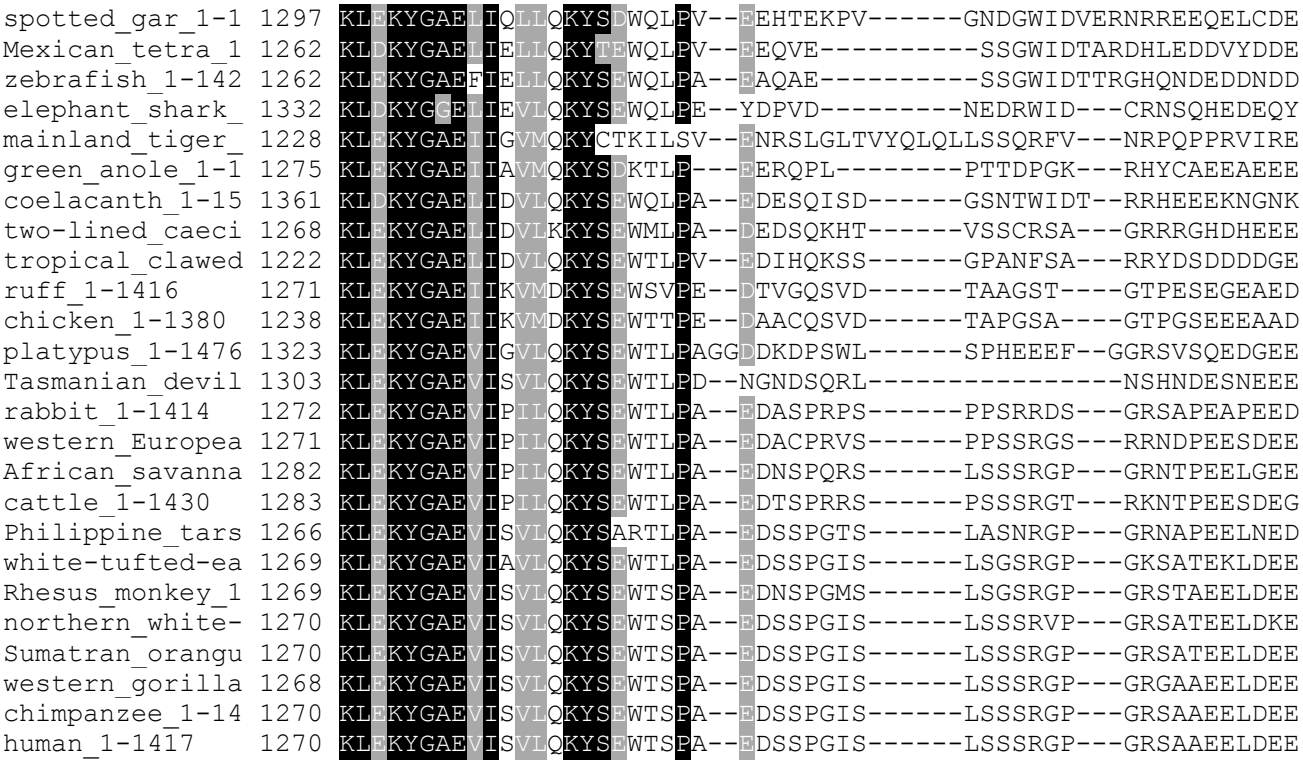

spotted_gar_1-1 939 GITALAYHAGLKDSDRDYVQQKWIEQDECQVMCATIAFGMGIDKPDVRFVIHASLPKSVE
Mexican_tetra_1 904 GIAALAYHAGLNDKDREYVQNKWINQDRCQVMCATIAFGMGIDKPDVRYVIHASLPKSVE
zebrafish_1-142 904 GIAALAYHAGLSDSDREYVQNKWINQDGCQVMCATIAFGMGIDKPDVRYVIHASLPKSVE
elephant_shark_ 975 GLAALAYHAGLKDSNRELVQNKWINQDGCQVICATIAFGMGIDKPDVRYVIHATLPKSVE
mainland_tiger_ 872 GLSALAYHAGLPDGTRDLVQQKWINQDGCQVICATIAFGMGIDKPDVRFVIHASLPKSIE
green_anole_1-1 919 GLSALAYHAGLPDETRDIVQQKWINQDGCQIICATIAFGMGIDKPDVRFVIHASLPKSIE
coelacanth_1-15 1002 GLAALAYHAGLGDSDRDLVQQKWINQDGCQVICATIAFGMGIDKPDVRYVIHASLPKSVE
two-lined_caeci 913 GLSALAYHAGLSDSSRDFVQHKWINQEDCQVICATIAFGMGIDKPDVRYVIHASLPKSME
tropical_clawed 866 GLAALAYHAGLADSNRDYVQHKWINQDDCQVICATIAFGMGIDKPDVRYVIHASLPKSVE
ruff_1-1416 913 GLAALAYHAGLTDSNRDLVQKKWINQEGCQVICATIAFGMGIDKPDVRYVIHASLPKSVE
chicken_1-1380 880 GLAALAYHAGLTDSNRDLVQKKWVNQEGCQVICATIAFGMGIDKPDVRYVIHASLPKSIE
platypus_1-1476 965 GLAALAYHAGLSDSARDEVQHKWINQDGCQVICATIAFGMGIDKPDVRYVIHASLPKSVE
Tasmanian_devil 946 GLAALAYHAGLSDSARDEVQHKWINQDGCQVICATIAFGMGIDKPDVRYVIHSSLPKSVE
rabbit_1-1414 914 GLAALAYHAGLSDSARDEVQHKWINQDGCQVICATIAFGMGIDKPDVRFVIHASLPKSVE
western_Europea 913 GLAALAYHAGLSDAARDEVQHKWINQDGCQVICATIAFGMGIDKPDVRFVIHASLPKSVE
African_savanna 924 GLAALAYHAGLSDCARDEVQHKWINQDGCQVICATVAFGMGIDKPDVRFVIHASLPKSVE
cattle_1-1430 925 GLAALAYHAGLSDSARDEVQHKWINQDGCQVICATIAFGMGIDKPDVRFVIHASLPKSVE
Philippine_tars 908 SLAALAYHAGLTDSARDEVQHKWINQDGCQVICATIAFGMGIDKPDVRFVIHASLPKSVE
white-tufted-ea 911 GLSALAYHAGLSDSARDEIQQKWINQDGCQVICATIAFGMGIDKPDVRFVIHASLPKSME
Rhesus_monkey_1 911 GLAALAYHAGLSDSARDEVQQKWINQDGCQVICATIAFGMGIDKPDVRFVIHASLPKSME
northern_white- 912 GLAALAYHAGLSDSARDEVQQKWINQDGCQVICATIAFGMGIDKPDVRFVIHASLPKSVE
Sumatran_orangu 912 GLAALAYHAGLSDSARDEVQQKWINQDDCQVICATIAFGMGIDKPDVRFVIHASLPKSME
western_gorilla 910 GLAALAYHAGLSDSARDEVQQKWINQDGCQVICATIAFGMGIDKPDVRFVIHASLPKSVE
chimpanzee_1-14 912 GLAALAYHAGLSDSARDEVQQKWINQDGCQVICATIAFGMGIDKPDVRFVIHASLPKSME
human_1-1417 912 GLAALAYHAGLSDSARDEVQQKWINQDGCQVICATIAFGMGIDKPDVRFVIHASLPKSVE

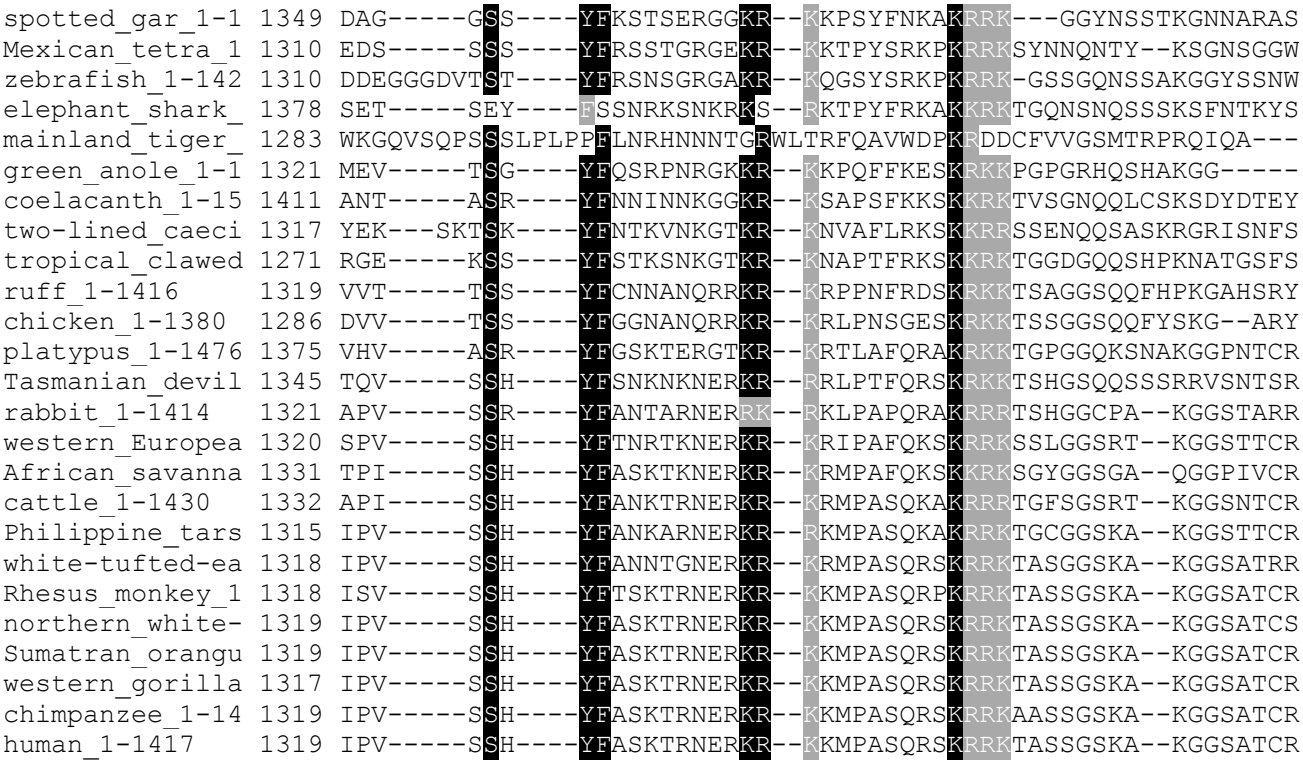

spotted_gar_1-1 999 GYYQESGRAGRDGEISHCVLFYSYNDAIRIKRLINMERDGNKETKQTHITNLYSMVHFCE
Mexican_tetra_1 964 GYYQESGRAGRDGEISHCVLFYSYSDVIRIKRILAMDKDGNHHSRATHTNNLHSMVHFCE
zebrafish_1-142 964 GYYQESGRAGRDGEISHCVLFYSYSDVIRIKRLIAMDKDGNQQSKATHINNLHSMVHFCE
elephant_shark_ 1035 GYYQESGRAGRDGETSHCTLYYSYHDVTRIRRLIQMEKDGNQHTKQTHINNLYSMVHYCE
mainland_tiger_ 932 GYYQESGRSGRDGEISHCLLFYSYSDVTRLRRLILMERDGNSHTRQTHFNNLYSMVHYCE
green_anole_1-1 979 GYYQESGRAGRDGERSHCLLFYSYSDVTRLRRLILMEKDGNSHTRQTHFNNLYSMVHYCE
coelacanth_1-15 1062 GYYQESGRAGRDGDISHCVLFYTYNDVTRIKRLIQMEKDGNYHTKQTHFNNLYSMVHYCE
two-lined_caeci 973 GYYQESGRAGRDGEMSHCLLFYTYSDVTRLRRLIQMERDGTFQTKQTHLNNLYNMVYYCE
tropical_clawed 926 GYYQESGRAGRDGEISHCLLFYSYSDVTRIRRLIQMEKDGNSHTKQTHFNNLYSMVHYCE
ruff_1-1416 973 GYYQESGRAGRDGEMSHCLLFYSYSDVTRLRRLILMEKDGNSHTRQTHFNNLYSMVHYCE
chicken_1-1380 940 GYYQESGRAGRDGEMSHCLLFYSYSDVTRLRRLILMEKDGNSHTRQTHFNNLYSMVHYCE
platypus_1-1476 1025 GYYQESGRAGRDGEISHCLLFYSYHDVIRLKRLILMEKDGNSHTRQTHFNNLYSMVHYCE
Tasmanian_devil 1006 GYYQESGRAGRDGEMSHCLLFYNYHDVTRLKRLILMEKDGNSHTRQTHFNNLYSMVHYCE
rabbit_1-1414 974 GYYQESGRAGRDGEISHCVLFYTYHDVTRLKRLILMEKDGNQHTRETHFNNLYSMVHYCE
western_Europea 973 GYYQESGRAGRDGEMSHCLLFYTYHDVTRLKRLILMEKDGNHHTRQTHFNNLYSMVHYCE
African_savanna 984 GYYQESGRAGRDGEISHCLLFYTYHDVTRLKRLILTEKDGNQHTKETHFNNLYSMVHYCE
cattle_1-1430 985 GYYQESGRAGRDGEISHCTLFYTYHDVTRLKRLILMEKDGNRHTRETHFNNLYSMVHYCE
Philippine_tars 968 GYYQESGRAGRDGELSHCLLFYTYHDATRLKRLIMMEKDGNHHTRETHLNNLYSMVHYCE
white-tufted-ea 971 GYYQESGRAGRDGEISHCLLFYTYHDVTRLKRLIMMEKDGNHHTRETHFNNLYSMVHYCE
Rhesus_monkey_1 971 GYYQESGRAGRDGEISHCLLFYTYHDVTRLKRLIIMEKDGNHHTRETHFNNLYSMVHYCE
northern_white- 972 GYYQESGRAGRDGEISHCLLFYTYHDVTRLKRLIMMEKDGNHHTRETHFNNLYSMVHYCE
Sumatran_orangu 972 GYYQESGRAGRDGEMSHCLLFYTYHDVTRLKRLIMMEKDGNHHTRETHFNNLYSMVHYCE
western_gorilla 970 GYYQESGRAGRDGEISHCLLFYTYHDVTRLKRLIMMEKDGNHHTRETHFNNLYSMVHYCE
chimpanzee_1-14 972 GYYQESGRAGRDGEISHCLLFYTYHDVTRLKRLIMMEKDGNHHTRETHFNNLYSMVHYCE
human_1-1417 972 GYYQESGRAGRDGEISHCLLFYTYHDVTRLKRLIMMEKDGNHHTRETHFNNLYSMVHYCE

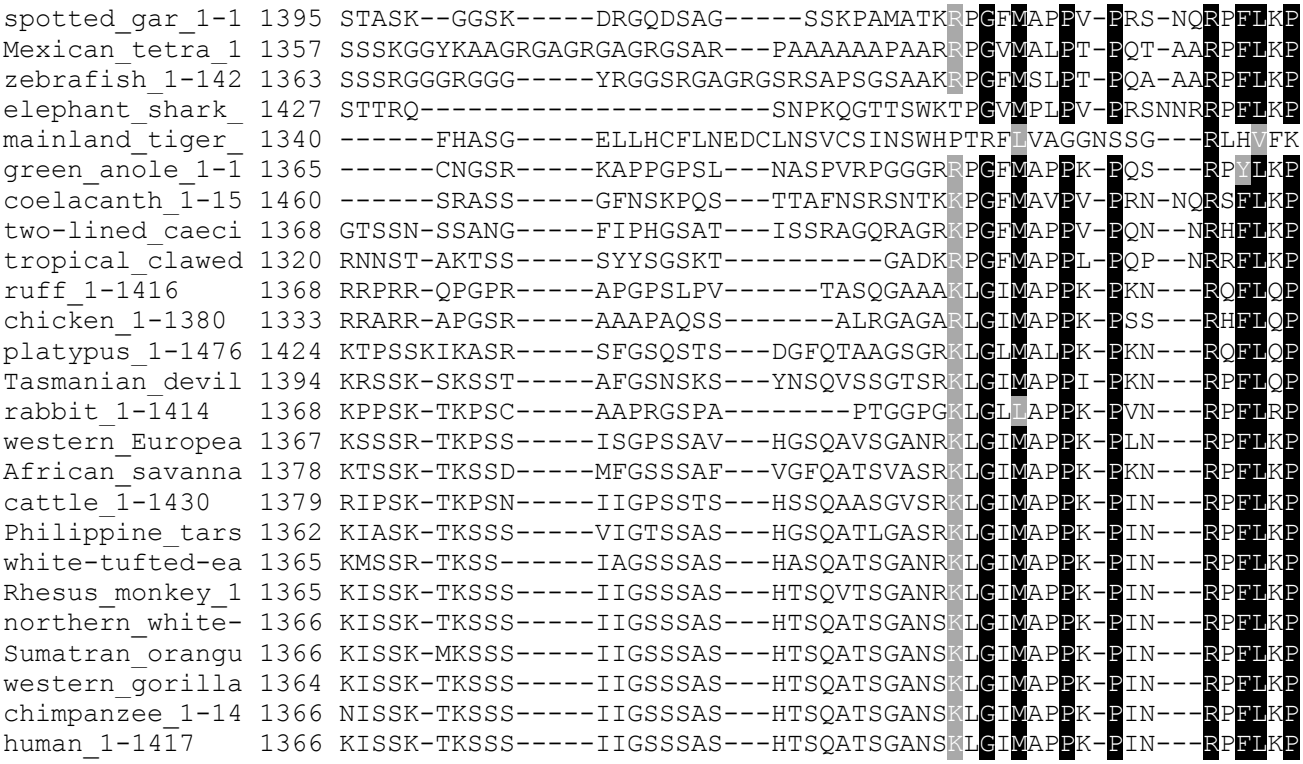

spotted_gar_1-1 1059 NVVECRRTQLLAYFGERTFNPNFCKEHPDVTCDNCARAKLYKSRNVTEEVKNVVRFVQEH
Mexican_tetra_1 1024 NVSECRRIQLLAYFGENKFNADFCKEHPEVICDNCARPHQYKARNVTEDVKKIVRFVQEN
zebrafish_1-142 1024 NVAECRRIQLLAYFGEHTFNTSFCKEHPEVICDNCARPNKYKSRNVTDDVKKIARFVQDN
elephant_shark_ 1095 NMLECRRTQLLAYFGENGFDPKFCKANPDATCDNCIRVKSYKTINVTEVVKNIVRFVQEH
mainland_tiger_ 992 NIMECRRIQLLAYFGETGFNPRFCKEHPEVSCDNCSKQQVYKLKNVTEEVKNIVRFVKEY
green_anole_1-1 1039 NVVECRRIQLLAYFGETGFNPKFCKEYPEVSCDNCCKVQDYNMKNVTEEVKSIVRFVQEH
coelacanth_1-15 1122 NVVDCRRTQLLAYFGENNFNPNFCKAHPEVICDNCSMKKDYKSRNVTDEVKAIIRFVQEH
two-lined_caeci 1033 NVVECRRVQLLVYFGENDFNPMFCKEHPDVACDNCCRKQNYKSYNVTEEVKDIVRFVREH
tropical_clawed 986 NVVECRRMQLLAYFGENNFNPNFCKENTRVSCDNCLGRKEFKSRNVTDDVRNIVRFVQEN
ruff_1-1416 1033 NLVDCRRIQLLAYFGETNFNPTFCKDHPEVTCDNCSRKKDYKSRDVTDDVKSIVRFVQEH
chicken_1-1380 1000 NVVDCRRIQLLAYFGETDFNPNFCKDHPEVICDNCSRKKDYKSRNVTDEVKSIIRFVQQH
platypus_1-1476 1085 NIAECRRIQLLSYFGENGFNPNFCKEHLDVACDNCCQKKDYKSRDVTEDVKNIVRFVQEL
Tasmanian_devil 1066 NIIECRRIQLLSYFGESGFNPNFCKEYPDVTCDNCCKKKEYKTRDVTEDVKNIIKFVQEH
rabbit_1-1414 1034 NITECRRIQLLAYFGENGFNPNFCKKYSDVSCDNCCKTKDFKTRDVTDEVKNIVRFVQEH
western_Europea 1033 NITECRRIQLLAYFGENGFNPDFCKKYPEVSCDNCCKTKDFKTRDVTDDVKNIVQFVKEH
African_savanna 1044 NITECRRIQLLAYFGENGFNPYFCKKYPEVSCDNCCKTKDFKTRDVTDDVKNIVRFVQEY
cattle_1-1430 1045 NITECRRIQLLAYFGENEFNPHFCKKYPDVSCDNCYKTKDFKTKDVTEDVKNIVRFVQEH
Philippine_tars 1028 NITECRRIQLLAYFGEHGFNPDFCKKHPDASCDNCCKTKDYKTRDVTDDVKNIVRFVQEH
white-tufted-ea 1031 NITECRRIQLLAYFGENGFNPDFCKKHADVSCDNCCKTKDYKTRDVTDDVKSIIRFVQEH
Rhesus_monkey_1 1031 NITECRRIQLLAYFGENGFNPDFCKKHPDVSCDNCCKTKDYKTRDVTDDVKSIIRFVQEH
northern_white- 1032 NITECRRIQLLAYFGENGFNPDFCKKHPDVSCDNCCKTKDYKTRDVTDDVKSIIRFVQEH
Sumatran_orangu 1032 NITECRRIQLLAYFGENGFNPDFCKKHPDVSCDNCCKTKDYKTRDVTDDVKSIVRFVQEH
western_gorilla 1030 NITECRRIQLLAYFGENGFNPDFCKKHPDVSCDNCCKTKDYKSRDVTEDVKSIVRFVQEH
chimpanzee_1-14 1032 NITECRRIQLLAYFGENGFNPDFCKKHPDVSCDNCCKTKDYKTRDVTDDVKSIVRFVQEH
human_1-1417 1032 NITECRRIQLLAYFGENGFNPDFCKKHPDVSCDNCCKTKDYKTRDVTDDVKSIVRFVQEH


spotted_gar_1-1 1119 CAKVGV-RQAKSAQL-NRLTLNMLVDIFIGTKSARIQTGLYGKGAAYSRHNAERLFRKLV
Mexican_tetra_1 1084 CQKVGG-RYSKSSQQ-NRLTLNMLVDIFLGAKSARIQTGMFGHGAAYSRHNADRLFKKLV
zebrafish_1-142 1084 CEKVGN-RYGKSAQQ-NRLTLNMLVEIFLGSKSARIQTGMFGVGAAYSKHNAERLFKKLV
elephant_shark_ 1155 CSNSGG-RYSGQSS--TRLTLNMLVDIFLGGKNARIQSGTFGTGAAYTRHNAERLFRKLV
mainland_tiger_ 1052 CGEKGT-RNIKRNA--GRYTLNMMIDIFLGAKAAKIQQGLFGKGAAYSRHNAERLFRKLV
green_anole_1-1 1099 CGEKGA-RGARRNG--GRYTLNMMIDIFLGTKSAKIQTGIYGKGAAYSRHNAERLFRKLV
coelacanth_1-15 1182 CTQSGGGKGKGTAQLRTRLTLNMLVDIFLGTKSARIQSGIFGRGAAYSRHNAERLFRKLV
two-lined_caeci 1093 CTTLGG-KAKFR----GRFTLNMMVDIFLGSKSAKIQIGIFGKGAAYSRHNAERLFRKLV
tropical_clawed 1046 CSSARG-KGRSNS---SRLTLNMMVDIFLGSKSAKIQIGLFGKGAAYSRHNAERLFRKLV
ruff_1-1416 1093 CGQTGR-TNAKRNAGFGRYTLNMMVDIFLGAKSAKVQSGIFGKGAAYSRHNAERLFRKLV
chicken_1-1380 1060 CGQVGG-INGNRNTGSGRYTLNMMVDIFLGAKSAKIQSGIFGKGAAYSRHNVERLFRKLV
platypus_1-1476 1145 DSLQGR-RNMRNITSPGRFTMNMLVDIFLGSKSAKVQSGIFGKGSAYSRHNAERLFKKLV
Tasmanian_devil 1126 DSLNGE-RN-KNMGSCGRFTMNMITDIFLGSKCAKIQTGIFGKGAAYSRHNAERLLKKLI
rabbit_1-1414 1094 SSSQGA-RPARPLGPSGRFTMNMLVDIFLGSKSAKIQSGIFGKGSAYSRHNAERLFKKLI
western_Europea 1093 SSSQER-GNIMHVGPAGRFTMNMLVDIFLGSKSAKVQSGIFGKGSTYSRHNAERLFKKLV
African_savanna 1104 NSSQGM-KNIKHAGPSGRFTMNMLVDIFLGSKSAKIQSGIFGMGSAYSRHNAERLFKKLI
cattle_1-1430 1105 SSSQGT-RNIKHIGPSGRFTMNMLVDIFLGSKSAKIQSGIFGKGSTYSRHNAERLFKKLI
Philippine_tars 1088 SLSQGT-RNIKHVGPSGRFTLNMLVDIFLGSKSAKIQSGIFGKGSAYSRHNAERLFKKLI
white-tufted-ea 1091 SSSQGA-RNIKHIGPSGRFTMNMLVDIFLGSKSAKIQSGIFGKGSAYSRHNAERLFKKLI
Rhesus_monkey_1 1091 SSSQGM-RNVKHVGPSGRFTMNMLVDIFLGSKSAKIQSGIFGKGSAYSRHNAERLFKKLI
northern_white- 1092 SSSQGT-RNIKHVGPSGRFTMNMLVDIFLGSKSAKLQSGIFGKGSAYSRHNAERLFKKLI
Sumatran_orangu 1092 SSSQGT-RNIKHVGPSGRFTMNMLVDIFLGSKSAKLQSGIFGKGSAYSRHNAERLFKKLI
western_gorilla 1090 SSSQGM-RNIKHVGPSGRFTMNMLVDIFLGSKSAKIQSGIFGKGSAYSRHNAERLFKKLI
chimpanzee_1-14 1092 SSSQGM-RNIKHVGPSGRFTMNMLVDIFLGSKSAKIQSGIFGKGSAYSRHNAERLFKKLI
human_1-1417 1092 SSSQGM-RNIKHVGPSGRFTMNMLVDIFLGSKSAKIQSGIFGKGSAYSRHNAERLFKKLI
spotted_gar_1-1 1177 LENVLEESLYITANDQAVAYISAGQKAAAVLSGFMQIEFYDTESASSIKKQRASVTKNMS
Mexican_tetra_1 1142 LDNVLMEDLYITNNGQAVAYISAGQKAMNVLNGSMQVVFHETESASSIRKHKASVIKNVS
zebrafish_1-142 1142 LDNVLMEDLYITNSGQAVAYISAGPKAMSVLNGCMQVEFHETESASSIRKHRASVCEKVS
elephant_shark_ 1212 LDKILREDLYITANDQAVAYIASGEKAHTVLIGCLQVEFQETGNTSSIRKQKNSVTKNVS
mainland_tiger_ 1109 LDKILDEDLYITANDQAIAYIQLGEKAVAVLNGSLQVEFCETESASNIRRQRA-LDTKMS
green_anole_1-1 1156 LDKILDEELYITANDQAVAYVHTGERAHVVLSGCLQVEFCETESASSLRKHRA-SVTKIS
coelacanth_1-15 1242 LDRILDEELYITANGQAIAYETIGEKAQAVLNGYLQVEFHDTKSASSVRKKAA-VMKNVS
two-lined_caeci 1148 LDRILDEELYITSNDQAVAYVSLGEKSQAVLNGYLQVEFQDMENASSIRKQKTSLISQVS
tropical_clawed 1102 LDRILDEELYITANDQAVAYVKTGERAQAVLNGFLTVEFQDTENASSLRKQKASVVANTS
ruff_1-1416 1152 LDKILDEDLYITANDQAVAYVILGQKAQAVLNGSLQVEFHETESASAVRKQRA-SVAKVS
chicken_1-1380 1119 LDKILDEDLYITANDQAVAYVVLGEKAQAVLNGLLQVEFHETENASAIRKQRA-SVTKMS
platypus_1-1476 1204 LDKILDEDLFINANDQPLAYIVAGERAQAVLHGYSKVDFIETENASSIRKQKA-LVAKVS
Tasmanian_devil 1184 LDKILDEDLYINANDQAIAYVMLGEKAQAVLNGYLQVDFLETENLSGIRKQKA-LVKNIS
rabbit_1-1414 1153 LDKILDEELYINANDQAIAYVMPGNKAQTLLSGHLKVDFMETENSSSVKKQKA-LVAKVS
western_Europea 1152 LDNILDEDLHINANDQAIAYVMPGSKAQAVLSGHLKVDFMETENSSSVKKQKT-LVAKVS
African_savanna 1163 LDKILDEDLYINANDQAIAYVMPGTEAQTVLNGHLKVDFMETENSSSVKKQKA-LEAKVS
cattle_1-1430 1164 LDRILDEDLYINANDQAIAYVMLGSKAQTVLNGHLKVDFMETENSSSVKKQKA-LVAKVS
Philippine_tars 1147 LDKILDEDLYINANDQAIAYVMLGNKAQSVLNGHLKVDFMETENSSSIKKQKA-LVAKVS
white-tufted-ea 1150 LDKILDEDLYINANDQAIAYMMLGNKAQGVLNGNLKVDFMETENSSSVKKQKA-LVAKVS
Rhesus_monkey_1 1150 LDKILDEDLYINANDQAIAYVMLGNKAQTVLNGNLKVDFMETENSSSVKKQKA-LVAKVS
northern_white- 1151 LDKILDEDLYINANDQAIAYVMLGNKAQTVLNGNLKVDFMETENSSSVKKQKA-LVAKVS
Sumatran_orangu 1151 LDKILDEDLYINANDQAIAYVMLGDKAQTVLNGALKVDFMETENSSSVKKQKA-LVAKVS
western_gorilla 1149 LDKILDEDLYINANDQAIAYVMLGNKAQTVLNGNLKVDFMETENSSSVKKQKA-LVAKVS
chimpanzee_1-14 1151 LDKILDEDLYINANDQAIAYVMLGNKAQTVLNGNLKVDFMETENSSSVKKQKA-LVAKVS
human_1-1417 1151 LDKILDEDLYINANDQAIAYVMLGNKAQTVLNGNLKVDFMETENSSSVKKQKA-LVAKVS


spotted_gar_1-1 1237 KREEMVQNCLQELNDLCKKLGKIFGIHYYNIFSTATLKKIAETLSADPEVLLQIDGVTED
Mexican_tetra_1 1202 KREEMVTKCLEELNELCKKLGKVFGLHYYNIFSTSTLKTIAETLSADPEVLLQIDGVTED
zebrafish_1-142 1202 KREEMVKKCLGELNDLCKKLGKVFGIHYYNIFSTATLKKLAETLSADPEVLLQIDGVTED
elephant_shark_ 1272 KRQEMVQKCLAELTELCKKLGKVFGVHYFNIFNVGTIKSIAESLSSDREDLLQIDGVTED
mainland_tiger_ 1168 LREEMVKKCLSELTDVCKNLGKVFNIHYFNIFNTATLKRIAETLSSDPAVLLQIDGVTED
green_anole_1-1 1215 QREEMAKECLSELTDVCKNLGKVFNVHYFNIFNTATLKRIAETLSSDPEVLLQIDGVTED
coelacanth_1-15 1301 KRDEMVQKCLGAITELCKKLGKVFGVHYFNIFNTATIKKLAETLSSDPEVLLQIDGVTED
two-lined_caeci 1208 QREEVVKKCLEELTELCKRLGKVFGVHYFNIFNTATIKKIAETLSPEPEVLLGIDGVTED
tropical_clawed 1162 QREEMVKKCQAELTELCKRLGKIFGVHYFNIFNTATIRRIAESLSPEPDVLLQIDGVTED
ruff_1-1416 1211 QREEMVKKCLGELTDTCKTLGKVFDVHYFNIFSTSTLKKIAETLSSDVEVLLQIDGVTED
chicken_1-1380 1178 QREEMVKKCLGELTDTCKTLGKIFDVHYFNIFSTSTLKKIAETLSSDAEVLLQIDGVTED
platypus_1-1476 1263 EREEVVKKCLGELTEVCKALGKVFGVHYFNIFNTVTLKKVAESLSSDPEVLLQIDGVTED
Tasmanian_devil 1243 QREELVKKCLGELTEVCKGLGKVFGVHYFNIFNTVTLKKLAESLSSDPEVLLQIDGVTED
rabbit_1-1414 1212 QREEVVKKCLGELTEVCKSLGKVFGVHYFNIFSTATLRRLAESLSSDPEVLLQIDGVTED
western_Europea 1211 KREEIVKKCLGELTEVCKSLGKVFGVHYFNIFNTATLKKLAESLSSDPEVLLQIDGVTED
African_savanna 1222 QREEMVKKCLGELTEICKSLGKVFGVHYFNIFNTATLKKLAESLSSDPEVLLQIDGVTED
cattle_1-1430 1223 QREEVVKKCLGELTEVCKSLGKVFGVHYFNIFNTVTLKKLAESLSSDPEVLLQIDGVTED
Philippine_tars 1206 QREEMVKKCLGELTEVCKSLGKVFGVHYFNIFNTVTLKKLAESLSSDPEVLLQIDGVTED
white-tufted-ea 1209 QREEMVKKCLGELTEVCKSLGKVFGVHYFNIFNTVTLKKLAESLSSDPEVLLQIDGVTED
Rhesus_monkey_1 1209 QREEMVKKCLGELTEVCKSLGKVFGVHYFNIFNTVTLKKLAESLSSDPEVLLQIDGVTED
northern_white- 1210 QREEMVKKCLGELTEVCKSLGKVFGVHYFNIFNTVTLKKLAESLSSDPEVLLQIDGVTED
Sumatran_orangu 1210 QREEMVKKCLGELTEVCKSLGKVFGVHYFNIFNTVTLKKLAESLSSDPEVLLQIDGVTED
western_gorilla 1208 QREEMVKKCLGELTEVCKSLGKVFGVHYFNIFNTVTLKKLAESLSSDPEVLLQIDGVTED
chimpanzee_1-14 1210 QREEMVKKCLGELTEVCKSLGKVFGVHYFNIFNTVTLKKLAESLSSDPEVLLQIDGVTED
human_1-1417 1210 QREEMVKKCLGELTEVCKSLGKVFGVHYFNIFNTVTLKKLAESLSSDPEVLLQIDGVTED


spotted_gar_1-1 1297 KLEKYGAELIQLLQKYSDWQLPV--EEHTEKPV------GNDGWIDVERNRREEQELCDE
Mexican_tetra_1 1262 KLDKYGAELIELLQKYTEWQLPV--EEQVE----------SSGWIDTARDHLEDDVYDDE
zebrafish_1-142 1262 KLEKYGAEFIELLQKYSEWQLPA--EAQAE----------SSGWIDTTRGHQNDEDDNDD
elephant_shark_ 1332 KLDKYGGELIEVLQKYSEWQLPE--YDPVD---------NEDRWID---CRNSQHEDEQY
mainland_tiger_ 1228 KLEKYGAEIIGVMQKYCTKILSV--ENRSLGLTVYQLQLLSSQRFV---NRPQPPRVIRE
green_anole_1-1 1275 KLEKYGAEIIAVMQKYSDKTLP---EERQPL--------PTTDPGK---RHYCAEEAEEE
coelacanth_1-15 1361 KLDKYGAELIDVLQKYSEWQLPA--EDESQISD------GSNTWIDT--RRHEEEKNGNK
two-lined_caeci 1268 KLEKYGAELIDVLKKYSEWMLPA--DEDSQKHT------VSSCRSA---GRRRGHDHEEE
tropical_clawed 1222 KLEKYGAELIDVLQKYSEWTLPV--EDIHQKSS------GPANFSA---RRYDSDDDDGE
ruff_1-1416 1271 KLEKYGAEIIKVMDKYSEWSVPE--DTVGQSVD------TAAGST----GTPESEGEAED
chicken_1-1380 1238 KLEKYGAEIIKVMDKYSEWTTPE--DAACQSVD------TAPGSA----GTPGSEEEAAD
platypus_1-1476 1323 KLEKYGAEVIGVLQKYSEWTLPAGGDDKDPSWL------SPHEEEF--GGRSVSQEDGEE
Tasmanian_devil 1303 KLEKYGAEVISVLQKYSEWTLPD--NGNDSQRL----------------NSHNDESNEEE
rabbit_1-1414 1272 KLEKYGAEVIPILQKYSEWTLPA--EDASPRPS------PPSRRDS---GRSAPEAPEED
western_Europea 1271 KLEKYGAEVIPILQKYSEWTLPA--EDACPRVS------PPSSRGS---RRNDPEESDEE
African_savanna 1282 KLEKYGAEVIPILQKYSEWTLPA--EDNSPQRS------LSSSRGP---GRNTPEELGEE
cattle_1-1430 1283 KLEKYGAEVIPILQKYSEWTLPA--EDTSPRRS------PSSSRGT---RKNTPEESDEG
Philippine_tars 1266 KLEKYGAEVISVLQKYSARTLPA--EDSSPGTS------LASNRGP---GRNAPEELNED
white-tufted-ea 1269 KLEKYGAEVIAVLQKYSEWTLPA--EDSSPGIS------LSGSRGP---GKSATEKLDEE
Rhesus_monkey_1 1269 KLEKYGAEVISVLQKYSEWTSPA--EDNSPGMS------LSGSRGP---GRSTAEELDEE
northern_white- 1270 KLEKYGAEVISVLQKYSEWTSPA--EDSSPGIS------LSSSRVP---GRSATEELDKE
Sumatran_orangu 1270 KLEKYGAEVISVLQKYSEWTSPA--EDSSPGIS------LSSSRGP---GRSATEELDEE
western_gorilla 1268 KLEKYGAEVISVLQKYSEWTSPA--EDSSPGIS------LSSSRGP---GRGAAEELDEE
chimpanzee_1-14 1270 KLEKYGAEVISVLQKYSEWTSPA--EDSSPGIS------LSSSRGP---GRSAAEELDEE
human_1-1417 1270 KLEKYGAEVISVLQKYSEWTSPA--EDSSPGIS------LSSSRGP---GRSAAEELDEE


spotted_gar_1-1 1349 DAG-----GSS----YFKSTSERGGKR--KKPSYFNKAKRRK---GGYNSSTKGNNARAS
Mexican_tetra_1 1310 EDS-----SSS----YFRSSTGRGEKR--KKTPYSRKPKRRKSYNNQNTY--KSGNSGGW
zebrafish_1-142 1310 DDEGGGDVTST----YFRSNSGRGAKR--KQGSYSRKPKRRK-GSSGQNSSAKGGYSSNW
elephant_shark_ 1378 SET-----SEY----FSSNRKSNKRKS--RKTPYFRKAKKRKTGQNSNQSSSKSFNTKYS
mainland_tiger_ 1283 WKGQVSQPSSSLPLPPFLNRHNNNTGRWLTRFQAVWDPKRDDCFVVGSMTRPRQIQA---
green_anole_1-1 1321 MEV-----TSG----YFQSRPNRGKKR--KKPQFFKESKRKKPGPGRHQSHAKGG-----
coelacanth_1-15 1411 ANT-----ASR----YFNNINNKGGKR--KSAPSFKKSKKRKTVSGNQQLCSKSDYDTEY
two-lined_caeci 1317 YEK---SKTSK----YFNTKVNKGTKR--KNVAFLRKSKKRRSSENQQSASKRGRISNFS
tropical_clawed 1271 RGE-----KSS----YFSTKSNKGTKR--KNAPTFRKSKKRKTGGDGQQSHPKNATGSFS
ruff_1-1416   1319 VVT-----TSS----YFCNNANQRRKR--KRPPNFRDSKRKKTSAGGSQQFHPKGAHSRY
chicken_1-1380 1286 DVV-----TSS----YFGGNANQRRKR--KRLPNSGESKRKKTSSGGSQQFYSKG--ARY
platypus_1-1476 1375 VHV-----ASR----YFGSKTERGTKR--KRTLAFQRAKRKKTGPGGQKSNAKGGPNTCR
Tasmanian_devil 1345 TQV-----SSH----YFSNKNKNERKR--RRLPTFQRSKRKKTSHGSQQSSSRRVSNTSR
rabbit_1-1414   1321 APV-----SSR----YFANTARNERRK--RKLPAPQRAKRRRTSHGGCPA--KGGSTARR
western_Europea 1320 SPV-----SSH----YFTNRTKNERKR--KRIPAFQKSKRRKSSLGGSRT--KGGSTTCR
African_savanna 1331 TPI-----SSH----YFASKTKNERKR--KRMPAFQKSKKRKSGYGGSGA--QGGPIVCR
cattle_1-1430   1332 API-----SSH----YFANKTRNERKR--KRMPASQKAKRRRTGFSGSRT--KGGSNTCR
Philippine_tars 1315 IPV-----SSH----YFANKARNERKR--RKMPASQKAKRRKTGCGGSKA--KGGSTTCR
white-tufted-ea 1318 IPV-----SSH----YFANNTGNERKR--KRMPASQRSKRRKTASGGSKA--KGGSATRR
Rhesus_monkey_1 1318 ISV-----SSH----YFTSKTRNERKR--KKMPASQRPKRRKTASSGSKA--KGGSATCR
northern_white- 1319 IPV-----SSH----YFASKTRNERKR--KKMPASQRSKRRKTASSGSKA--KGGSATCS
Sumatran_orangu 1319 IPV-----SSH----YFASKTRNERKR--KKMPASQRSKRRKTASSGSKA--KGGSATCR
western_gorilla 1317 IPV-----SSH----YFASKTRNERKR--KKMPASQRSKRRKTASSGSKA--KGGSATCR
chimpanzee_1-14 1319 IPV-----SSH----YFASKTRNERKR--KKMPASQRSKRRKAASSGSKA--KGGSATCR
human_1-1417   1319 IPV-----SSH----YFASKTRNERKR--KKMPASQRSKRRKTASSGSKA--KGGSATCR


spotted_gar_1-1 1395 STASK--GGSK-----DRGQDSAG-----SSKPAMATKRPGFMAPPV-PRS-NQRPFLKP
Mexican_tetra_1 1357 SSSKGGYKAAGRGAGRGAGRGSAR---PAAAAAAPAARRPGVMALPT-PQT-AARPFLKP
zebrafish_1-142 1363 SSSRGGGRGGG-----YRGGSRGAGRGSRSAPSGSAAKRPGFMSLPT-PQA-AARPFLKP
elephant_shark_ 1427 STTRQ----------------------SNPKQGTTSWKTPGVMPLPV-PRSNNRRPFLKP
mainland_tiger_ 1340 ------FHASG-----ELLHCFLNEDCLNSVCSINSWHPTRFLVAGGNSSG---RLHVFK
green_anole_1-1 1365 ------CNGSR-----KAPPGPSL---NASPVRPGGGRRPGFMAPPK-PQS---RPYLKP
coelacanth_1-15 1460 ------SRASS-----GFNSKPQS---TTAFNSRSNTKKPGFMAVPV-PRN-NQRSFLKP
two-lined_caeci 1368 GTSSN-SSANG-----FIPHGSAT---ISSRAGQRAGRKPGFMAPPV-PQN--NRHFLKP
tropical_clawed 1320 RNNST-AKTSS-----SYYSGSKT----------GADKRPGFMAPPL-PQP--NRRFLKP
ruff_1-1416   1368 RRPRR-QPGPR-----APGPSLPV------TASQGAAAKLGIMAPPK-PKN---RQFLQP
chicken_1-1380 1333 RRARR-APGSR-----AAAPAQSS-------ALRGAGARLGIMAPPK-PSS---RHFLQP
platypus_1-1476 1424 KTPSSKIKASR-----SFGSQSTS---DGFQTAAGSGRKLGLMALPK-PKN---RQFLQP
Tasmanian_devil 1394 KRSSK-SKSST-----AFGSNSKS---YNSQVSSGTSRKLGIMAPPI-PKN---RPFLQP
rabbit_1-1414   1368 KPPSK-TKPSC-----AAPRGSPA--------PTGGPGKLGLLAPPK-PVN---RPFLRP
western_Europea 1367 KSSSR-TKPSS-----ISGPSSAV---HGSQAVSGANRKLGIMAPPK-PLN---RPFLKP
African_savanna 1378 KTSSK-TKSSD-----MFGSSSAF---VGFQATSVASRKLGIMAPPK-PKN---RPFLKP
cattle_1-1430   1379 RIPSK-TKPSN-----IIGPSSTS---HSSQAASGVSRKLGIMAPPK-PIN---RPFLKP
Philippine_tars 1362 KIASK-TKSSS-----VIGTSSAS---HGSQATLGASRKLGIMAPPK-PIN---RPFLKP
white-tufted-ea 1365 KMSSR-TKSS------IAGSSSAS---HASQATSGANRKLGIMAPPK-PIN---RPFLKP
Rhesus_monkey_1 1365 KISSK-TKSSS-----IIGSSSAS---HTSQVTSGANRKLGIMAPPK-PIN---RPFLKP
northern_white- 1366 KISSK-TKSSS-----IIGSSSAS---HTSQATSGANSKLGIMAPPK-PIN---RPFLKP
Sumatran_orangu 1366 KISSK-MKSSS-----IIGSSSAS---HTSQATSGANSKLGIMAPPK-PIN---RPFLKP
western_gorilla 1364 KISSK-TKSSS-----IIGSSSAS---HTSQATSGANSKLGIMAPPK-PIN---RPFLKP
chimpanzee_1-14 1366 NISSK-TKSSS-----IIGSSSAS---HTSQATSGANSKLGIMAPPK-PIN---RPFLKP
human_1-1417   1366 KISSK-TKSSS-----IIGSSSAS---HTSQATSGANSKLGIMAPPK-PIN---RPFLKP

```
spotted_gar_1-1 1441 AFSHM-
Mexican_tetra_1 1412 AFSHLG
zebrafish_1-142 1416 SFSHL-
elephant_shark_ 1464 SYSFG-
mainland_tiger_ 1386 EW----
green_anole_1-1 1407 SYSLL-
coelacanth_1-15 1504 SFSLL-
two-lined_caeci 1416 SYAFL-
tropical_clawed 1361 SYTLF-
ruff_1-1416     1412 SYSVL-
chicken_1-1380  1376 SYAVL-
platypus_1-1476 1472 SYTFL-
Tasmanian_devil 1441 SYSFL-
rabbit_1-1414   1410 SFAVC-
western_Europea 1414 SFTLL-
African_savanna 1425 SYAFS-
cattle_1-1430   1426 SYAFS-
Philippine_tars 1409 SYAFS-
white-tufted-ea 1411 SYAFA-
Rhesus_monkey_1 1412 SYAFS-
northern_white- 1413 SYAFS-
Sumatran_orangu 1413 SYAFS-
western_gorilla 1411 SYAFS-
chimpanzee_1-14 1413 SYAFS-
human_1-1417    1413 SYAFS-
```

8.7. List of 120 DDR Proteins used as an Input in the SLiMCon Tool

Table 8.7.1 List of 120 DDR proteins used as an input in the SLiMCon tool					
ABRAXAS1	DCLRE1A	FANCI	PAXX	RADX	SMARCAD1
APLF	DCLRE1B	FANCM	PMS1	RBBP8	SMARCAL1
ATM	DCLRE1C	FBH1	POLD3	RECQL4	SMC6
ATR	EME1	GEN1	POLH	RECQL5	SPIDR
ATRIP	ERCC3	HELB	POLI	REV1	SPRTN
ATRX	ERCC4	HELQ	POLK	REV3L	TIMELESS
AUNIP	ERCC5	HMCES	POLL	RHNO1	TIPIN
BARD1	ERCC6	HROB	POLN	RIF1	TOPBP1
BLM	ERCC6L	MAD2L2	POLQ	RMI1	TP53BP1
BRCA1	ERCC6L2	MCPH1	PRIMPOL	RMI2	TRAIP
BRCA2	ESCO1	MDC1	RAD17	RNF168	USP1
BRIP1	ESCO2	MLH1	RAD18	RNF8	UVSSA
CDC25A	ETAA1	MRNIP	RAD23A	RTEL1	WAPL
CDC25C	EXD2	MSH3	RAD51AP1	SCAI	WRN
CDKL5	EXO1	MUS81	RAD51B	SETX	XPC
CLSPN	FAAP100	NBN	RAD52	SHLD1	XRCC1
CTC1	FAAP20	NEIL3	RAD54B	SHLD2	XRCC2
CTCF	FANCA	NHEJ1	RAD54L	SLF2	XRCC3
CYREN	FANCF	NIPBL	RAD54L2	SLX4	XRCC4
DAXX	FANCG	PALB2	RAD9A	SLX4IP	ZRANB3

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-020-20818-5 OPEN

The Bloom syndrome complex senses RPA-coated single-stranded DNA to restart stalled replication forks

Ann-Marie K. Shorrocks[1,2,6], Samuel E. Jones[1,2,6], Kaima Tsukada[1,2,3,6], Carl A. Morrow[1,2,6], Zoulikha Belblidia[1,2], Johanna Shen[1,2,4], Iolanda Vendrell[2,5], Roman Fischer[5], Benedikt M. Kessler[5] & Andrew N. Blackford[1,2✉]

The Bloom syndrome helicase BLM interacts with topoisomerase IIIα (TOP3A), RMI1 and RMI2 to form the BTR complex, which dissolves double Holliday junctions to produce non-crossover homologous recombination (HR) products. BLM also promotes DNA-end resection, restart of stalled replication forks, and processing of ultra-fine DNA bridges in mitosis. How these activities of the BTR complex are regulated in cells is still unclear. Here, we identify multiple conserved motifs within the BTR complex that interact cooperatively with the single-stranded DNA (ssDNA)-binding protein RPA. Furthermore, we demonstrate that RPA-binding is required for stable BLM recruitment to sites of DNA replication stress and for fork restart, but not for its roles in HR or mitosis. Our findings suggest a model in which the BTR complex contains the intrinsic ability to sense levels of RPA-ssDNA at replication forks, which controls BLM recruitment and activation in response to replication stress.

[1] MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK. [2] MRC Oxford Institute for Radiation Oncology, Department of Oncology, University of Oxford, Oxford OX3 7DQ, UK. [3] Department of Transdisciplinary Science and Engineering, School of Environment and Society, Tokyo Institute of Technology, Tokyo 152-8550, Japan. [4] Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520, USA. [5] Target Discovery Institute, Centre for Medicines Discovery, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK. [6] These authors contributed equally: Ann-Marie K. Shorrocks, Samuel E. Jones, Kaima Tsukada, Carl A. Morrow. ✉email: andrew.blackford@imm.ox.ac.uk

Bloom syndrome is a rare hereditary chromosomal instability disorder caused by mutations in the gene encoding the BLM helicase[1]. It is characterized by growth retardation, immunodeficiency, hypersensitivity to sunlight, and high cancer predisposition[2]. Remarkably, Bloom syndrome patients are predisposed to the full spectrum of malignancies found in the general population, in contrast to other inherited cancer-predisposing disorders[3]. Cells from these patients display multiple signatures of genome instability, in particular a large increase in sister chromatid exchanges (SCEs)[4], which consequently serves as a diagnostic test for Bloom syndrome[5].

BLM functions in cells with TOP3A, RMI1, and RMI2, which together form the BTR "dissolvasome" complex that can process HR intermediates to prevent genetic crossovers[6]. Accordingly, mutations in *TOP3A*, *RMI1*, and *RMI2* have been identified in patients with conditions similar to Bloom syndrome[7,8]. BLM has been proposed to act as a tumor suppressor by preventing crossovers between homologous chromosomes, which could lead to loss-of-heterozygosity[9,10]. However, more recent data from genome-wide analyses suggest that loss-of-heterozygosity events in BLM-deficient cells are extremely rare[11], indicating that other cellular functions of BLM may contribute to the increased cancer predisposition in Bloom syndrome patients. Such functions include an early-stage role in the HR repair process by stimulating DNA-end resection[12], processing of ultra-fine DNA bridges (UFBs) between sister chromatids in anaphase[13], and promoting restart of stalled replication forks[14]. Currently, no separation-of-function mutations have been identified that could distinguish the roles of BLM in double Holliday junction (dHJ) dissolution, DNA-end resection, UFB processing, and stalled fork restart, to prove which contribute to tumor suppression in a suitable animal model. Furthermore, very little is understood as to how the activities of the BTR complex are regulated in response to genotoxic stress, and how BLM can tell the difference between normal DNA metabolic processes and DNA damage or stalled replication forks.

RPA is an abundant, high-affinity ssDNA-binding complex that is thought to recognize ssDNA rapidly in cells whether it is generated during DNA replication or as a result of DNA damage[15]. It consists of three subunits, RPA1, RPA2, and RPA3 (also referred to as RPA70, RPA32 and RPA14 respectively, based on how they migrate during sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)). All three subunits contain oligonucleotide (OB)-binding folds that either bind to ssDNA or mediate protein–protein interactions, with the N-terminal OB-fold of RPA1 and the C-terminal winged-helix domain of RPA2 being the major interaction sites recognized by proteins that respond to DNA damage or replication stress.

BLM complexes isolated from cell extracts contain RPA[16], and in early studies the N-terminal 447 residues of BLM were suggested to bind directly to the RPA1 subunit[17,18]. However, later work suggested that RMI1 rather than BLM mediates RPA association with the BTR complex[19]. Thus, the RPA-binding sites within the BTR complex have not yet been precisely mapped and the physiological relevance of its association with RPA has not yet been defined.

Here, we demonstrate that the Bloom syndrome complex contains three RPA-binding motifs, two in BLM and one in RMI1. We show that all three motifs contribute to RPA-binding of the BTR complex, and that they all interact with the same N-terminal domain of RPA1, indicating that the BTR complex can associate with at least three discrete RPA complexes. Using a combination of CRISPR-Cas9, RNA interference, and ectopic expression of recombinant proteins delivered by lentivirus, we generated a system to produce cells expressing a BTR complex that cannot bind RPA. Characterization of these cells revealed that RPA-binding is required for stable recruitment of BLM to DNA damage sites, stalled replication fork restart, and cellular resistance to replication stress, but not for the roles of BLM in HR or UFB processing.

Results

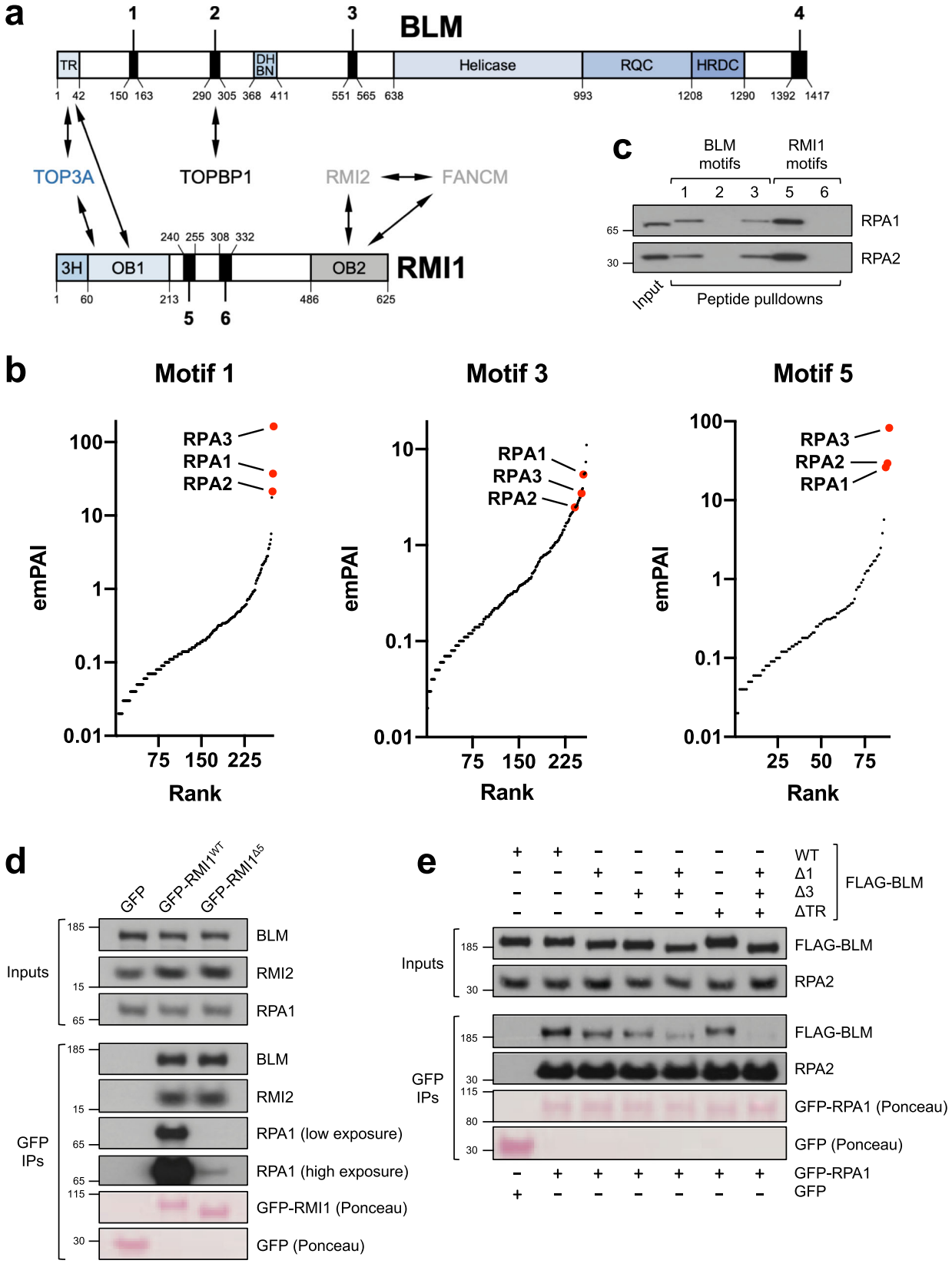
Identification of conserved RPA-binding motifs in the BTR complex. The biochemical activities of the BTR dissolvasome complex have been extensively characterized in earlier studies (reviewed in ref. [6,20]). However, very little is known about how the functions of this protein complex are regulated in cells. To address this question, we focused our attention on the intrinsically disordered regions of the BTR complex, the functions of which are still poorly understood. Regions of intrinsic structural disorder are hotspots for post-translational modifications and often contain short evolutionarily conserved linear peptide motifs that can act as binding sites for other proteins[21].

To identify such motifs in the BTR complex, we performed amino-acid sequence alignments of all four BTR subunits (BLM, TOP3A, RMI1, and RMI2) from representative vertebrate orthologues (Supplementary Fig. 1a–d). These revealed the presence of six peptide motifs in the complex that have been highly conserved throughout evolution (four in BLM and two in RMI1; Fig. 1a and Supplementary Fig. 1a, b). We did not find any putative peptide motifs in TOP3A as this protein appears to consist only of its topoisomerase domain and a series of structured zinc-finger motifs[22]; nor did we identify peptide motifs in RMI2, as this protein is relatively small and made up of a single oligonucleotide-binding (OB)-fold domain[23,24] (Supplementary Fig. 1c, d).

We hypothesized that the conserved peptide motifs we identified in BLM and RMI1 mediate interactions between the BTR complex and other proteins. In support of this idea, motif 2 has previously been shown to bind to the DNA damage response mediator protein TOPBP1 when phosphorylated on Ser304[25,26]. We therefore designed synthetic biotinylated peptides corresponding to residues in all six motifs for use as baits in pulldown experiments. In addition, we examined these motifs for the presence of confirmed phosphorylation sites (defined as having been identified in at least two separate studies collated at https://www.phosphosite.org), but only motif 2 contained such a modification in a conserved residue (Ser304, as mentioned above). We therefore synthesized both native and phospho-Ser304 versions of motif 2 for use in subsequent experiments.

Biotinylated peptides were conjugated to streptavidin-coupled beads and incubated with HeLa nuclear extracts, and potential interacting partners were identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS; Supplementary Data 1–7). Importantly, TOPBP1 was identified as one of the top hits in the pulldown with motif 2, but only when the peptide derived from this motif was phosphorylated on Ser304 (Supplementary Data 2, 3). In contrast, peptides corresponding to motifs 4 and 6 did not interact with any proteins other than common contaminants[27] (Supplementary Data 5, 7). Strikingly, the top hits from the pulldowns with peptides based on motifs 1 and 3 (from BLM), and motif 5 (from RMI1), were all three subunits of RPA (Fig. 1b and Supplementary Data 1, 4, and 6). We confirmed these results by immunoblotting (Fig. 1c), leading us to speculate that the BTR complex contains at least three separate interaction sites for RPA: two in BLM, and one in RMI1.

To test this, we investigated whether mutation of these motifs in the context of full-length BLM and RMI1 affected their interaction with RPA. First, 293FT cells were transfected with plasmids expressing GFP-tagged wild-type (WT) RMI1 or mutant RMI1 in which motif 5 was deleted. GFP immunoprecipitations (IPs) from lysates from these cells revealed that the mutant RMI1



protein was severely deficient in its ability to interact with RPA (Fig. 1d). Conversely, its interactions with BLM and RMI2, which bind to other regions of RMI1[28,29], were unaffected, indicating that loss of motif 5 does not cause general misfolding of RMI1. Notably, at higher exposures of the RPA1 blot in this experiment, it was clear that some residual RPA still associated with the mutant GFP-RMI1 protein above the level observed in the GFP-only lane, probably indirectly via RPA-binding motifs in BLM (see below).

Next, we deleted motifs 1 and 3 alone or in combination in constructs expressing FLAG-tagged BLM, transfected them into 293FT cells along with GFP or GFP-tagged RPA1, and performed

Fig. 1 Identification of conserved RPA-binding motifs in the BTR complex. a Schematic of proteins, domains, and motifs in the BTR complex. *TR* TOP3A-RMI1-binding domain, *DHBN* dimerization helical bundle in N-terminal, *RQC* RecQ C-terminal, *HRDC* helicase and RNaseD C-terminal; *3H* 3-helix bundle, *OB* oligonucleotide-binding. **b** Plots of MS hits from peptide pulldowns with motifs 1, 3, and 5, ranked by emPAI score[116] (number of unique peptide sequences adjusted for protein size). RPA subunits are highlighted in red. **c** Validation of MS results by western blotting, showing that RPA from HeLa nuclear extracts is specifically pulled down by biotinylated peptides based on motifs 1, 3, and 5 but not motifs 2 and 6. **d** GFP-pulldowns from 293FT cells transfected with constructs expressing GFP or the indicated GFP-tagged RMI1 variants, showing that loss of motif 5 disrupts RPA-binding. **e** GFP-pulldowns from 293FT cells transfected with constructs expressing either GFP or GFP-RPA1, and the indicated FLAG-tagged BLM variants, showing how loss of motifs 1 and/or 3, and the N-terminal TOP3A-RMI1 (TR) binding domain of BLM impacts on RPA-binding.

GFP IPs from cell lysates. Interestingly, loss of either motif 1 or motif 3 caused a marginal reduction in binding of BLM to RPA, whereas deletion of both BLM motifs caused a far more pronounced defect in RPA binding (Fig. 1e).

Given that some residual BLM-RPA binding was still apparent even in absence of both motifs 1 and 3, we considered the possibility that the presence of endogenous RMI1 (with its motif 5 that also binds RPA) in the BTR complex might be responsible for this. We therefore needed to design a BLM construct missing key residues in its TOP3A/RMI1 association domain. It has previously been shown that the first 133 residues of BLM are required for it to interact with the rest of the BTR complex[25,30]. However, motif 1 is located in close proximity to this region, raising the possibility that deletion of BLM residues 1–133 might interfere with RPA-binding regardless of BLM interaction with RMI1. To narrow down the binding site for BTR components on BLM, we examined the human BLM sequence for the presence of conserved residues known to disrupt binding of Sgs1, the budding yeast orthologue of BLM, to Top3 and Rmi1[31] (Supplementary Fig. 2a). This led us to predict that deletion of a region spanning residues 4–37 in the TOP3A/RMI1-binding ("TR") domain of BLM would disrupt its association with BTR complex components, and indeed this was the case (Supplementary Fig. 2b). We next examined the effect of this mutation (ΔTR) on BLM-RPA binding, in constructs either with or without motifs 1 and 3. Mutation of the TR domain of BLM caused a small decrease in RPA binding, indicating that stable BLM association with RPA depends to some extent on the presence of the RPA-binding motif in RMI1 (Fig. 1e). In support of this, mutation of the TR domain in combination with loss of motifs 1 and 3 abrogated almost all RPA binding to BLM. Taken together, these results demonstrate that all three RPA-binding motifs in the BTR complex contribute to maximize the association of BLM with RPA.

The RPA-binding motifs in BLM and RMI1 interact with the RPA1 N-terminal OB-fold. Given the presence of three RPA-binding motifs in BLM and RMI1, we wondered which subunit(s) of RPA these motifs interact with. To answer this question, we carried out further pulldown experiments with peptides derived from motifs 1, 3, and 5 in which one aromatic residue was replaced with the synthetic photoreactive amino-acid *p*-benzoyl-L-phenylalanine (BPA). UV-crosslinking and subsequent denaturing washes revealed that only the RPA1 subunit could be retained by these peptides (Fig. 2a), indicating that all three motifs interact directly with RPA1, but not RPA2 or RPA3.

The N-terminus of RPA1 contains an OB-fold domain with a basic binding cleft that mediates interactions with multiple DNA damage checkpoint and repair factors including p53, MRE11, ATRIP, RAD9, DNA2, ETAA1, and PRIMPOL[32–39]. Interestingly, when we compared the sequences of BLM/RMI1 motifs 1, 3, and 5 with the motifs from these proteins, we noticed significant sequence similarity (Fig. 2b). We thus speculated that BLM and RMI1 might also interact with the same site on RPA1. To test this, we in vitro translated the RPA1-RPA2-RPA3 complex with either WT or a mutant version of RPA1 in which key residues in the N-terminal basic binding cleft of RPA1 were mutated (R41E/Y42F)[40]. We then incubated the protein products with biotinylated peptides derived from the three BLM/RMI1 motifs for subsequent peptide pulldowns. Results from these experiments demonstrated that while all three peptides could bind WT RPA, the R41E/Y42F mutation in RPA1 abolished these interactions (Fig. 2c).

To verify the importance of the N-terminal OB-fold domain of RPA1 in the context of full-length proteins in cells, we carried out pulldowns from 293FT cells transfected with constructs expressing GFP-tagged WT or R41E/Y42F RPA1. Results from this experiment showed that mutation of the RPA1 N-terminal OB-fold inhibited MRE11 binding as expected from previous work[33], and led to a complete loss of detectable binding of RPA1 to the BTR complex (Fig. 2d). Importantly, the RPA1-R41E/Y42F mutant was still proficient in binding to XPA, which associates with RPA via the central region of RPA1 and the C-terminus of RPA2[41,42].

We conclude from these data and those above that the BTR complex contains three acidic peptide motifs (two in BLM and one in RMI1) that interact with the same basic binding cleft in the N-terminal OB-fold of RPA1, raising the possibility that at least three separate RPA complexes can associate simultaneously with the BTR complex.

Generation of a cell-based system to test the physiological relevance of BTR-RPA binding. To investigate whether RPA-binding contributes to some or all of the functions of the BTR complex, we needed to generate cells expressing mutants of BLM and RMI1 in which all three RPA-binding motifs were deleted. To do this, we first generated *BLM* knockout RPE-1 cells using CRISPR-Cas9 (Fig. 3a). We also tried to obtain *RMI1* single and *BLM/RMI1* double knockout cells but these attempts were unsuccessful, in line with previous studies indicating that complete loss of RMI1 or its stable binding partner TOP3A is lethal in vertebrate cells, regardless of the presence of p53[43–46].

We selected three independent $BLM^{-/-}$ RPE-1 clones in which no BLM protein could be detected (Fig. 3a), for further characterization. Sequencing of the targeted alleles in all three clones confirmed unique indels resulting in frameshift mutations and premature stop codons in the *BLM* locus (Supplementary Data 8). BLM-deficient cells typically display a large increase in SCE levels, so we first analyzed these in our CRISPR-generated cell lines. We found that $BLM^{-/-}$ RPE-1 cells had a > 10-fold increase in SCEs (Supplementary Fig. 3a), similar to Bloom syndrome patient cells[4], $Blm^{-/-}$ mouse embryonic fibroblasts[47] and $BLM^{-/-}$ chicken DT40 cells[48], thus confirming that these RPE-1 clones are true BLM knockout cells.

We also wished to know how our $BLM^{-/-}$ RPE-1 cells behaved in assays measuring other reported roles for BLM, in order to reveal the cellular role of RPA-binding in BTR complex functions. BLM has a key role in processing UFBs, and its absence leads to an increase in these structures in anaphase cells[13]. We therefore analyzed UFB numbers in WT and $BLM^{-/-}$ clones and found that as expected, our $BLM^{-/-}$ knockouts displayed around a threefold increase in UFBs (Supplementary Fig. 3b).

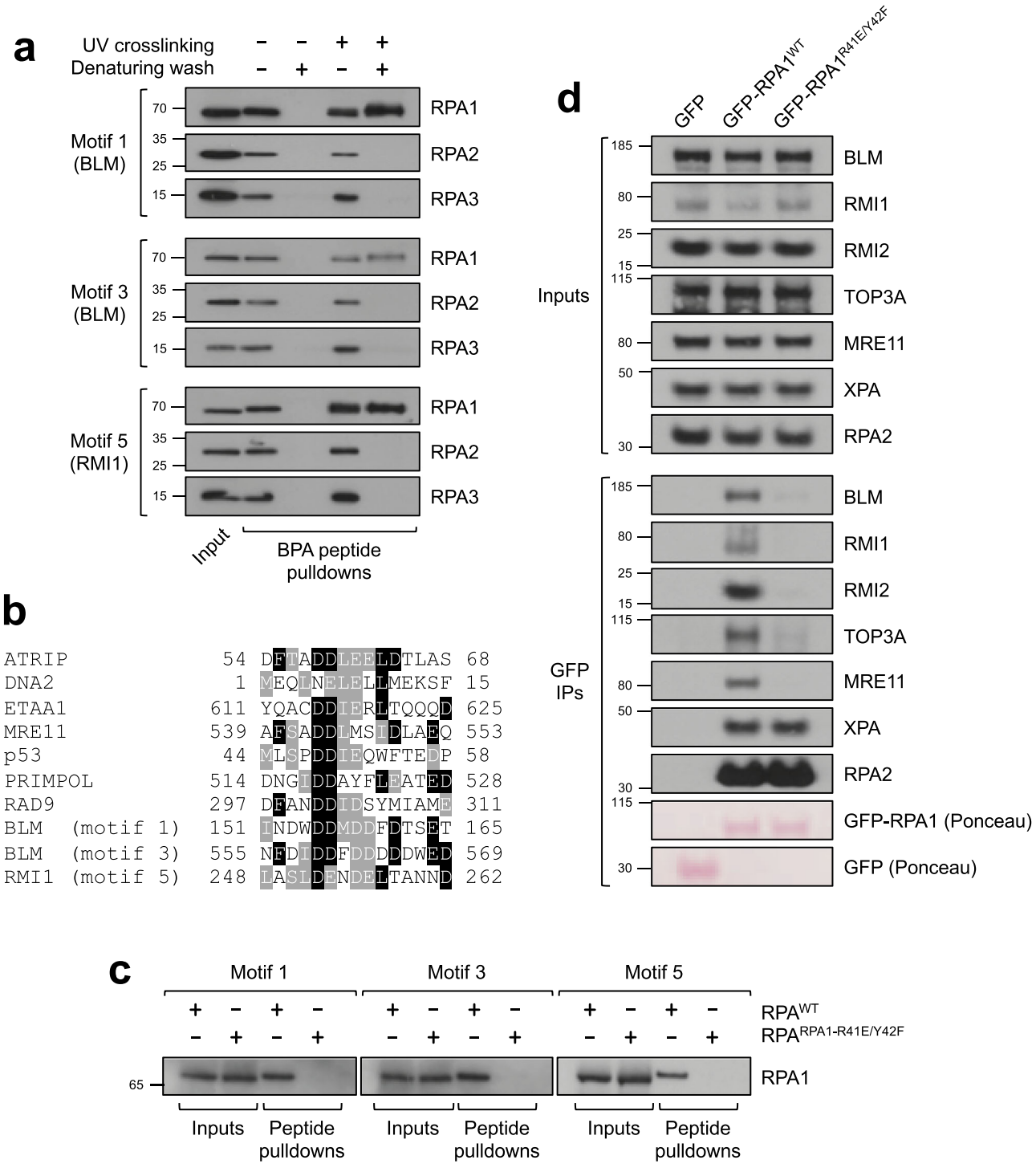


Fig. 2 The RPA-binding motifs in BLM and RMI1 interact with the RPA1 N-terminal OB-fold. a BPA-modified peptide pulldowns from HeLa nuclear extracts with UV-crosslinking and denaturing washes where indicated, demonstrating that motifs 1, 3, and 5 interact directly with the RPA1 subunit of RPA. **b** Sequence alignments showing the similarity of RPA1-binding motifs in DNA damage response proteins. **c** Peptide pulldowns from rabbit reticulocyte lysates containing in vitro translated heterotrimeric RPA complexes incorporating either WT or R41E/Y42F RPA1, showing that all three BLM/RMI1 motifs cannot interact with the mutant RPA complex. **d** GFP-pulldowns from 293FT cells transfected with constructs expressing GFP or the indicated GFP-tagged RPA1 variants, showing that mutation of the N-terminal OB-fold of RPA1 (R41E/Y42F) disrupts binding to the BTR complex and MRE11. XPA is a negative control because it binds to RPA via different domains.

BLM has also been implicated in promoting DNA-end resection in some studies[12,49–51], but not in others[52,53]. To test how BLM loss affects resection in RPE-1 cells, we used a fluorescence-activated cell-sorting (FACS)-based assay to measure RPA recruitment to chromatin in response to camptothecin (CPT) treatment in $BLM^{+/+}$ and $BLM^{-/-}$ cells[54]. Results from this assay demonstrated that $BLM^{-/-}$ cells are significantly impaired in DNA-end resection, but not by >~50% (Supplementary Fig. 3c). Thus, in our hands, BLM does promote resection but is not essential for this process. This is in line with previous

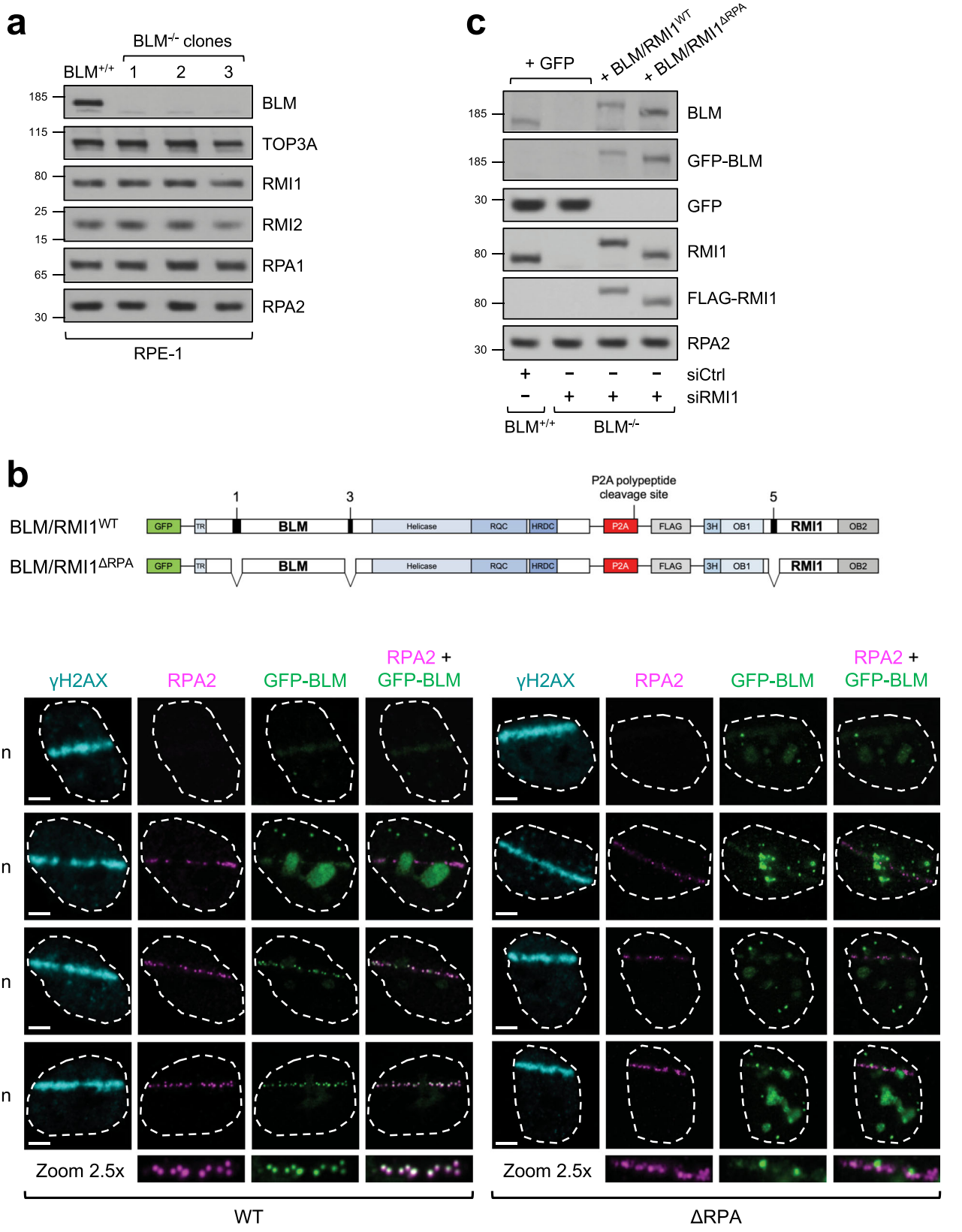


Fig. 3 Generation and complementation of *BLM*$^{-/-}$ RPE-1 cells with RPA-binding mutants of BLM and RMI1. a Western blots showing that BLM expression is undetectable in three separate *BLM* knockout RPE-1 clones. **b** Schematic showing the complementation strategy for adding back WT BLM/RMI1 proteins or mutant ones lacking RPA-binding motifs to *BLM*$^{-/-}$ cells treated with siRMI1. **c** Western blots showing the levels of exogenous GFP and BLM/RMI1 proteins delivered by lentivirus in WT or *BLM*$^{-/-}$ cells treated with siCtrl or siRMI1 as indicated. **d** Airyscan high-resolution confocal microscopy image time-course, showing the kinetics of recruitment of γH2AX, RPA2, and WT/mutant GFP-BLM proteins to laser-induced DSBs in representative cells. Zoomed-in images of laser lines are from the 60-min time-point images. Scale bars, 5 μm.

studies suggesting that BLM is functionally redundant with other RecQ helicases such as WRN and a parallel long-range resection pathway involving EXO1[12,50].

The role of BLM in maintenance of genome stability may be most important during S phase. Some BLM-deficient cell lines have been reported to be hypersensitive to genotoxic agents such as hydroxyurea (HU) that perturb DNA replication[55–58], although not all studies have found this[59,60]. We therefore investigated the requirement for BLM in response to DNA replication stress in our knockout cells, and found that all three of our *BLM*$^{-/-}$ RPE-1 clones were hypersensitive to HU treatment (Supplementary Fig. 3d). Using DNA fiber assays, BLM-deficient cells have been shown to be defective in replication fork restart after exposure to HU and other agents that interfere with DNA replication[14,61]. We confirmed that *BLM*$^{-/-}$ RPE-1 cells have a replication restart defect after HU treatment using similar analyses (Supplementary Fig. 3e, f). The fork restart defect was not due to increased RAD51 accumulation at stalled forks (as observed in budding yeast cells deficient in Sgs1[62]), because RAD51 foci formation after exposure to HU was no different regardless of the presence of BLM (Supplementary Fig. 3g, h).

To establish a complementation system to test the cellular requirement for RPA association with the BTR complex in these assays, we designed a lentiviral vector to express GFP-tagged BLM and FLAG-tagged RMI1 from the same promoter, separated by a P2A self-cleaving peptide sequence to produce two separate polypeptides (Fig. 3b). Mutations were then introduced to delete the two RPA-binding motifs in BLM and the one in RMI1. *BLM*$^{-/-}$ cells were then infected with viruses to express either WT GFP-BLM and FLAG-RMI1, or RPA-binding mutant versions of these proteins (BLM/RMI1$^{\Delta RPA}$), or GFP alone as a control. By subsequently depleting RMI1 transiently using siRNA, we were able to establish cells that do not detectably express BLM or RMI1, and cells expressing only RPA-binding mutant versions of these proteins for functional assays (Fig. 3c).

RPA-binding stabilizes BLM at ssDNA damage sites but is not essential for BLM recruitment to DSBs. Given that RPA is a platform for recruitment of multiple DNA damage response proteins in response to genotoxic stress[15], we considered the possibility that RPA-binding promotes recruitment of BLM to DNA damage sites. We therefore used laser microirradiation combined with 5-bromo-2'-deoxyuridine (BrdU) presensitization to induce such lesions, and examined GFP-BLM^{WT} and GFP-BLM$^{\Delta RPA}$ recruitment in a time-course experiment (Fig. 3d). Interestingly, both proteins accumulated rapidly to laser lines (within 5 min), although at relatively low levels and with a diffuse staining pattern. Such early recruitment to DSBs has been described previously for BLM[63], and requires its C-terminus rather than the N-terminal region containing the RMI1 and RPA-binding sites. Given also that RPA recruitment was not detectable until after 5 min, it made sense to us that there was no difference between WT and ΔRPA GFP-BLM recruitment at this early time point. By 15 min, RPA microfoci appeared along the path of the laser line in ~50% of irradiated cells, in keeping with previous reports showing that such RPA-coated ssDNA tracts are produced as a result of DNA-end resection in S and G2 cells only[64,65], with a ~700 s half-life of recruitment[66]. Interestingly, by 20 min WT GFP-BLM also formed bright microfoci along the path of the laser line, which colocalized with RPA. In contrast, the GFP-BLM$^{\Delta RPA}$ protein instead remained diffuse and at a low level at this time-point, resembling its early localization pattern. Indeed, even though at later times (>60 min) the mutant protein formed a small number of discernible foci along the path of the laser line, these rarely colocalized with RPA. Therefore, we conclude that at least one major function of RPA-binding is to stably recruit the BTR complex to ssDNA damage sites in cells, although BLM can still localize to DSBs through other mechanisms.

RPA-binding is required for the BTR complex to restart stalled replication forks but not to suppress SCEs, process UFBs, or promote DNA-end resection. Given that BLM recruitment to RPA-ssDNA sites was defective in absence of the RPA-binding motifs in BLM and RMI1, we next examined whether any of the biological processes that depend on BLM was defective in cells expressing a BTR complex defective in RPA binding. We first examined the level of SCEs in cells expressing BLM and RMI1 that cannot bind RPA. Although *BLM*$^{-/-}$ cells depleted of RMI1 and complemented with GFP alone displayed the expected 10-fold excess of SCEs, cells complemented with either exogenous WT or the RPA-binding mutant BLM/RMI1 proteins had levels of SCEs similar to WT RPE-1 cells (Fig. 4a). Thus, RPA association with the BTR complex is not required for suppression of SCEs. This is in line with the fact that dHJs do not contain ssDNA and our observation that BLM can still localize to DSB sites in the absence of the RPA-binding motifs in BLM and RMI1, although at a reduced level (Fig. 3d).

We then checked whether BLM-RPA binding is required for UFB processing. RPA is not observed on all bridges, and although RPA recruitment to UFBs is BLM-dependent, BLM and RPA do not colocalize on these structures[67,68]. Furthermore, recent biochemical evidence suggests that RPA may actually exclude BLM once deposited on UFBs[69]. In line with these studies, in cells expressing RPA-binding BLM/RMI1 mutants, GFP-BLM was still able to localize to UFBs (Fig. 4b), and the number of UFBs was similar to that observed in cells expressing WT BLM (Fig. 4c), indicating that BLM-RPA binding is not required for the role of the BTR complex in UFB processing.

Next, we examined whether BTR-RPA binding is required for DNA-end resection. Similar to the situation with SCEs and UFBs, we found that the RPA-binding mutant BLM/RMI1 proteins were able to complement the mild resection defect of BLM/RMI1-deficient cells (Fig. 4d), indicating that RPA-binding is not required for the BTR complex to stimulate DNA-end resection. This is in line with the observation that a small amount of BLM localizes to DSBs prior to the appearance of RPA foci (Fig. 3d).

We then examined whether RPA-binding is required for the role of the BTR complex in promoting replication fork restart. Using DNA fiber analyses as described above (Supplementary Fig. 3e), we found that in contrast to WT proteins, which could complement BLM/RMI1-deficient cells, cells expressing the BLM/RMI1$^{\Delta RPA}$ mutants were as defective in fork restart and displayed a similar increase in stalled forks after HU as cells lacking BLM/RMI1 expression altogether (Fig. 5a and Supplementary Fig. 4). In line with this, the mutant proteins were also unable to complement the HU hypersensitivity of BLM/RMI1-deficient cells (Fig. 5b). Thus, we conclude that RPA-binding is specifically required for the role of the BTR complex in promoting replication fork restart, rather than its roles in suppression of SCEs, UFB processing, or DNA-end resection.

RPA-binding is required for BLM recruitment to stalled replication forks. Given that RPA-binding promotes stable BLM recruitment to ssDNA foci at DSB sites (Fig. 3d), we considered the possibility that RPA-binding might be important for recruitment of BLM to stalled replication forks in response to HU, as nucleotide depletion by this agent leads to helicase-polymerase uncoupling and accumulation of RPA-ssDNA[64]. To test this, we examined the colocalization of WT and RPA-binding

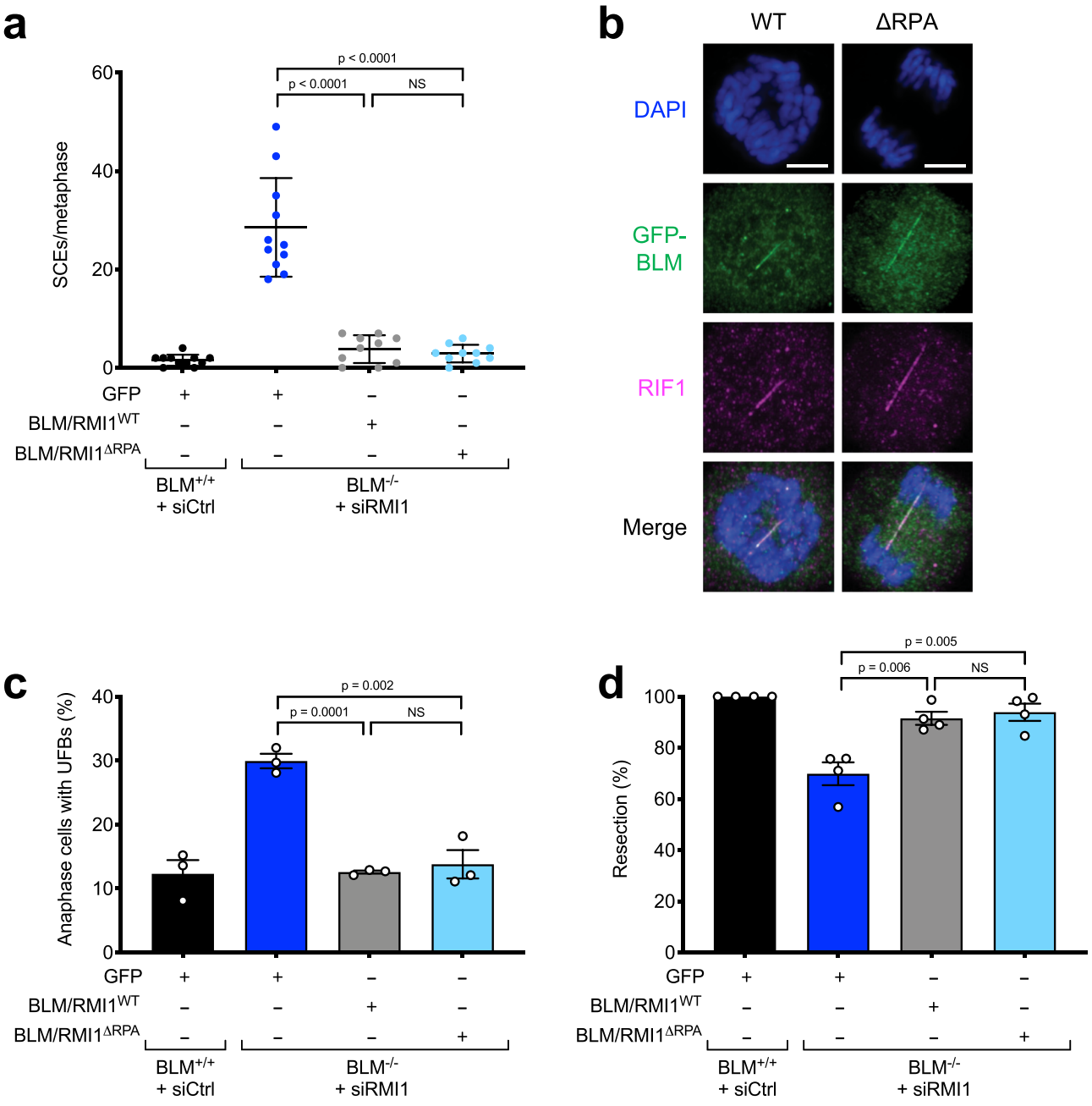


Fig. 4 RPA-binding is not required for the BTR complex to suppress SCEs, process UFBs, or promote DNA-end resection. a RPA-binding by BLM and RMI1 is not required for suppression of SCEs ($n = 3$, one representative experiment shown). At least 10 metaphases were scored per sample. Error bars denote standard deviation (SD) from the mean. Significance was determined using the two-sided Mann–Whitney *U* test. **b** Airyscan high-resolution confocal microscopy images showing that both WT and RPA-binding (ΔRPA) mutant BLM proteins localize to UFBs (marked by RIF1, which localizes independently of BLM to UFBs[68]). Scale bars, 5 μm. **c** RPA-binding by BLM and RMI1 is not required for suppression of UFBs. Error bars denote standard error of the mean (SEM). Significance was determined by two-sided unpaired *t* test ($n = 3$). **d** RPA-binding by BLM and RMI1 is not required for DNA-end resection. Error bars denote SEM. Significance was determined by two-sided unpaired *t* test ($n = 4$).

mutant GFP-BLM proteins with RPA foci formed in response to HU treatment by immunofluorescence microscopy. Results from these analyses indicated that there was a significant defect in the ability of the GFP-BLM$^{\Delta RPA}$ mutant protein to form foci at RPA-ssDNA sites in response to replication stress (Fig. 5c, d).

Thus, we conclude that RPA-binding promotes stable recruitment of BLM to stalled replication forks, where it is required for replication fork restart and genome stability after replication stress. Furthermore, given that the BTR complex contains three discrete binding sites for RPA1, our data suggest a model in which BTR contains the intrinsic ability to measure the level of RPA-ssDNA at replication forks to control BLM recruitment and activation in response to replication stress (Fig. 6a, b).

Discussion

In this report, we have precisely delineated how the Bloom syndrome complex interacts with RPA by carrying out unbiased proteomic analyses of conserved peptide motifs. We found that BLM contains two RPA-binding motifs, and RMI1 contains a third (Fig. 1b, c). Interestingly, these motifs all recognize the same N-terminal domain on the RPA1 subunit of the RPA

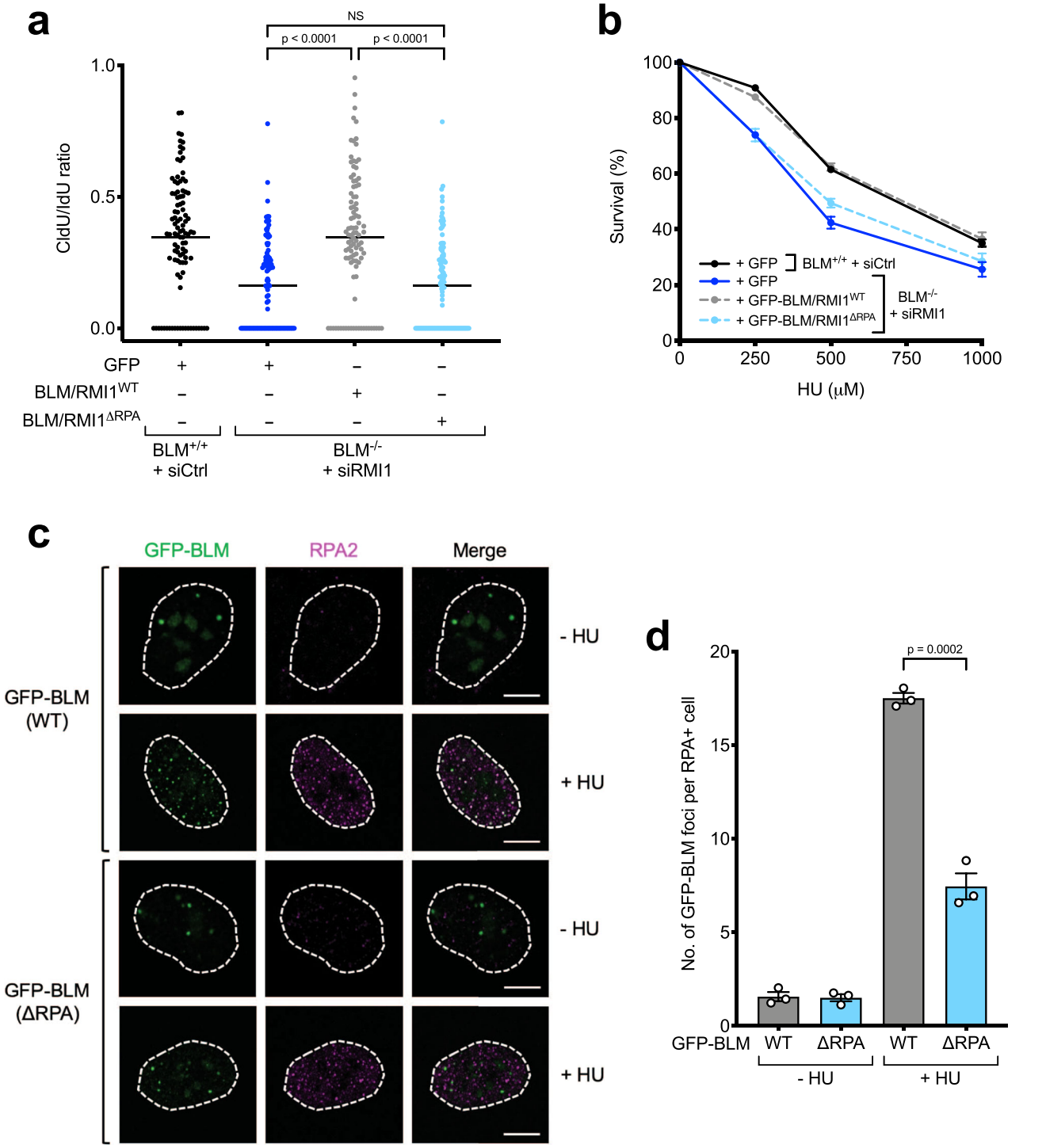


Fig. 5 RPA-binding is required for the role of the BTR complex in response to DNA replication stress. a Quantification of CldU/IdU tract length ratios in DNA fiber assays showing that RPA-binding by BLM and RMI1 is required for replication fork restart ($n = 3$, one representative experiment shown). At least 100 DNA fibers were analyzed per sample. Significance was determined using the two-sided Mann–Whitney *U* test. **b** Colony survival assays showing RPA-binding by BLM and RMI1 is required for HU resistance ($n = 3$, error bars denote SEM). Cells were treated with the indicated doses of HU for 24 h. **c** Immunofluorescence confocal microscopy images of the indicated proteins, demonstrating that the RPA-binding BLM mutant (ΔRPA) is defective in its ability to form foci that colocalize with RPA2 in response to replication stress. Cells were treated where indicated with 2 mM HU for 24 h. **d** Quantification of the number of WT and ΔRPA GFP-BLM foci in the experiment shown in **c**). Significance was determined using the two-sided unpaired *t* test ($n = 3$, error bars denote SEM).

heterotrimer (Fig. 2a–d). Crucially, all three motifs contribute to RPA-binding in the context of full-length BLM and RMI1 proteins (Fig. 1d, e), indicating that they are not functionally redundant. Instead, our data suggest that the BTR complex requires interaction with three RPA complexes for stable recruitment to ssDNA sites via motifs in both BLM and RMI1 (Figs. 3d and 5c, d), which explains why BLM recruitment to DNA damage foci depends on RMI1 and vice versa[65,70]. Notably, given that it is still unclear what the stoichiometry is of the various subunits within the BTR complex (BLM for example can

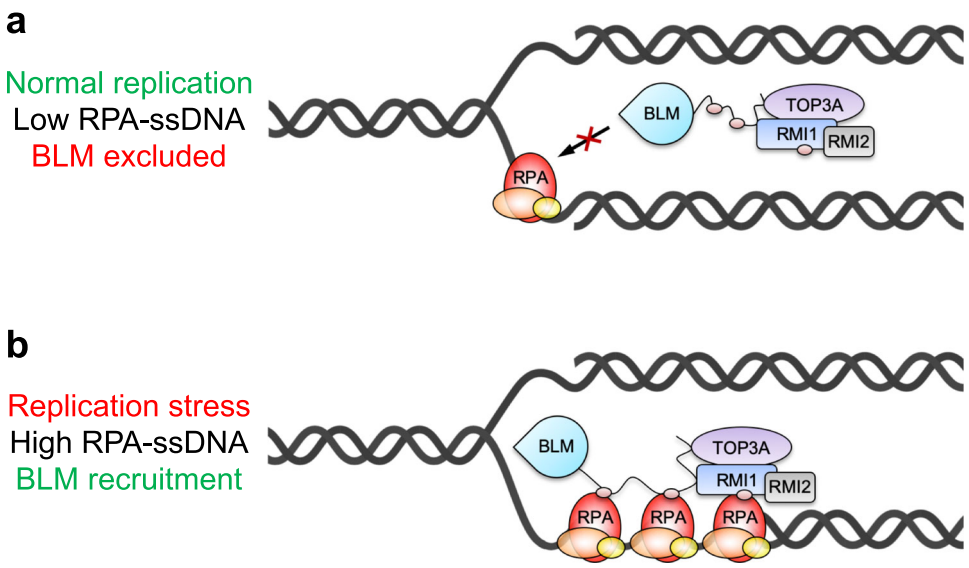


Fig. 6 Model for the role of RPA-binding in specifically recruiting the BTR complex to stalled replication forks. a During normal DNA replication, the BTR complex is not recruited to ongoing replication forks because not enough RPA-coated ssDNA complexes are present. **b** When a replication fork stalls, increased RPA-ssDNA leads to stable recruitment of the BTR complex and replication fork restart. Gray spheres represent the three RPA-binding motifs in BLM and RMI1.

dimerize via its DHBN domain[71]), the BTR complex in cells could potentially associate with more than three RPA complexes.

BLM was originally shown to interact with RPA using purified recombinant proteins in vitro[17]. Subsequently, it was demonstrated that RPA is present in BTR complexes immunoprecipitated from cells[16], and that the N-terminus of BLM is required for RPA-binding[18]. It was recently suggested that acidic peptides from BLM (similar to motifs 1 and 2 in our study; Fig. 1a) can interact with a recombinant N-terminal fragment of RPA1[72]. However, motif 2 is in fact the TOPBP1-binding site on BLM[25], and in line with this we did not detect any interaction of this motif with RPA (Fig. 1c), nor did deletion of this motif from the full-length BLM protein have any effect on RPA binding (Supplementary Fig. 5a). A different group could not detect any interaction between recombinant BLM and RPA, and instead mapped a binding site for RPA on RMI1[19]. Interestingly, that site corresponds to what we designated as motif 6 in our current study (Fig. 1a). However, in our hands the peptide corresponding to this motif does not interact with RPA (Fig. 1c); furthermore, deletion of this motif in full-length RMI1 does not impact on RPA binding, unlike deletion of motif 5, which almost abrogates the RMI1–RPA interaction (Fig. 1d and Supplementary Fig. 5b).

Some biochemical studies have suggested that RPA-binding stimulates the dHJ dissolution reaction by the BTR complex[19,73], whereas others have contradicted this[69,74]. We found that in human cells, RPA-binding by the BTR complex is probably not required for dHJ dissolution because cells expressing mutant BLM and RMI1 that cannot bind RPA do not display increased SCEs (Fig. 4a). This made sense to us, given that dHJ structures do not contain ssDNA and would therefore not be expected to require RPA-binding to recruit the BTR complex to dissolve them. Likewise, we found that direct binding of RPA to the BTR complex is not required for DNA-end resection (Fig. 4d). In vitro, BLM can promote resection by two nucleases, EXO1 and DNA2[75,76]. Genetically, however, loss of BLM is epistatic with DNA2 depletion and additive with EXO1 deficiency[12,50], indicating that in cells BLM cooperates with DNA2 to promote resection. This is in line with studies in budding yeast, where Sgs1 cooperates with Dna2, and Exo1 works in a parallel resection pathway[12,77,78]. Interestingly, an interaction between DNA2 and RPA is conserved throughout evolution and stimulates DNA2 nuclease activity[35,79–83]. Thus, it may be that RPA-binding by DNA2 is sufficient to promote resection by the BLM/DNA2 resectosome in cells even if RPA-binding by the BTR complex is lost. Finally, although the BTR complex plays a key role in UFB processing in anaphase, we found that RPA-binding does not recruit the complex to UFBs (Fig. 4b), nor does it appear to play a role in UFB processing (Fig. 4c). This is in line with recent evidence suggesting that RPA actively excludes BLM from ssDNA in mitosis[69], and that BTR recruitment to UFBs instead depends on the DNA translocase PICH[84].

Whereas RPA-binding is not required for all cellular roles of the BTR complex, we found that it does promote replication fork restart and cellular resistance to HU (Fig. 5a, b). The precise mechanism as to why BLM is required for fork restart is still unclear, but might for example involve fork remodeling by promoting fork regression or unwinding of aberrant DNA structures such as G-quadruplexes[85,86], which could be stimulated by RPA-binding[18]. In addition, our cell-based experiments indicate that a major role of RPA-binding is to stably recruit the BTR complex to stalled replication forks (Fig. 5c, d). This finding also suggests a mechanism whereby cells can "measure" the amount of RPA-coated ssDNA present at a stalled replication fork and engage the BTR complex to rescue it, while preventing precocious recruitment and activation of BLM at ongoing replication forks where it is not required (Fig. 6a, b). Our data also suggest that BLM can still localize to DNA damage sites through other mechanisms, where it is able to promote DNA-end resection at early times upon DSB induction (prior to RPA accumulation), and dissolution of dHJs (that do not contain ssDNA) at later times. One mechanism may involve recruitment of the BTR complex via FANCM, which can cooperate with BLM to suppress SCE and error-prone long-tract gene conversion[87–89].

It is interesting to note that interaction between a RecQ helicase and a ssDNA-binding protein has been highly conserved throughout evolution, possibly in order to respond to stalled replication forks. For example, immunodepletion of RPA from *Xenopus* egg extracts prevents the association of BLM with chromatin[90]. Furthermore, Sgs1 has been shown to interact with RPA via an acidic region in the N-terminus[91], and the fission

yeast ortholog of BLM, Rqh1, also forms a complex with RPA and has been suggested to be recruited to stalled forks via this interaction[92]. Even in bacteria, it has been shown that RecQ interacts with SSB, and that this interaction is required for localization of the helicase to the replication factory[93].

Our data suggest that the fork restart defect observed in absence of BLM is not caused by excessive accumulation of RAD51 (as seen in Sgs1 mutant budding yeast[62]), because we did not detect any difference in RAD51 foci between WT and *BLM*$^{-/-}$ cells (Supplementary Fig. 3g, h). This is in line with previous cell-based studies demonstrating that BLM deficiency does not result in increased RAD51 recruitment at stalled forks or DSBs induced by ionizing radiation, even though the loss of BLM in certain HR-deficient genetic backgrounds can restore RAD51 foci to some extent through an unknown mechanism[52,94]. Although BLM has been shown to disrupt RAD51-ssDNA filaments in vitro, unlike other known anti-recombinases it can only do so when RAD51 is in its recombination-inactive, ADP-bound form[95]. In contrast, Sgs1 can readily dismantle Rad51-ssDNA filaments in vitro via a mechanism that is not coupled to the Rad51 ATP hydrolysis cycle[96], a function shared with mammalian RECQL5[97]. This suggests that BLM may have lost the anti-recombinogenic function of Sgs1 in disassembling RAD51-ssDNA filaments at some point during evolution, while RECQL5 and possibly other RecQ helicases have maintained it. The BTR complex may nonetheless have one or more anti-recombinogenic functions downstream of RAD51-mediated strand invasion. Although BLM can dissolve D-loops in vitro, it can only do so when they are protein-free[98,99], a situation that is unlikely to occur in cells. In contrast however, purified TOP3A-RMI1-RMI2 can readily dissolve D-loops either in the presence or absence of RAD51, RAD54, and RPA[100], and Sgs1-Top3-Rmi1 disrupts D-loops in vivo[101]. Future studies will be required to establish whether the BTR complex similarly has a role in disrupting D-loops in human cells, or whether its role in the late stages of HR is limited to promoting dissolution of late recombination intermediates into non-crossovers (as suggested previously[102]), rather than actually suppressing recombination per se. If the BTR complex does disrupt D-loops in cells, it will be interesting to establish whether RPA-binding by BLM/RMI1 is required for this putative function.

Bloom syndrome is a heterogeneous condition with many associated pathologies. *BLM* mutations found in patients either result in loss of BLM protein or inhibit its helicase activity, which is required for all its known functions in human cells. It is thus unclear which cellular roles of BLM are required to prevent the various symptoms of the disease. We have found that loss of RPA association with the Bloom syndrome complex constitutes a clear separation-of-function mutation with regard to the cellular functions of BLM, given that RPA-binding is required for replication fork restart but not for other roles in DNA repair. Interestingly, mutation of Thr99 in the N-terminus of human BLM (a TQ site that is a preferred substrate for the apical DNA damage response kinases ATM and ATR[103]) has also been reported to cause replication fork restart defects but not increased SCEs[14]; however, as Thr99 is not conserved beyond primates (Supplementary Fig. 1a), it cannot easily be used as a separation-of-function mutation to investigate the role of BLM in vivo. In contrast, future studies in animal models in which the RPA-binding motifs in BLM and RMI1 are mutated may be informative in elucidating which function(s) of BLM are required to prevent the various pathologies with Bloom syndrome, with possible future therapeutic consequences for patients.

Methods

Sequence alignments. The following NCBI reference sequences were used to generate alignments in Supplementary Fig. 1a–d: BLM, NP_000048.1 (human), NP_031576.4 (mouse), XP_003434427.1 (dog), NP_001007088.2 (chicken), NP_001079095.1 (frog), XP_017207539.1 (zebrafish); RMI1, NP_079221.2 (human), NP_083180.3 (mouse), XP_013972800.1 (dog), NP_001026783.1 (chicken), NP_001086875.1 (frog), NP_956474.1 (zebrafish); RMI2, NP_689521.1 (human), NP_001156404.1 (mouse), XP_852348.2 (dog), NP_001006174.1 (chicken), XP_018091789.1 (frog), XP_691674.2 (zebrafish); TOP3A, NP_004609.1 (human), NP_033436.1 (mouse), XP_546656.2 (dog), NP_001025807.1 (chicken), NP_001085699.1 (frog), XP_688695.4 (zebrafish). Sequences were aligned using T-Coffee[104], with conserved and similar residues highlighted using Boxshade (https://embnet.vital-it.ch/software/BOX_form.html).

Peptide pulldowns. Biotinylated peptides (GenScript) were bound to streptavidin-coupled Dynabeads M-280 (Thermo Fisher Scientific). HeLa nuclear extracts (Ipracell) were diluted 1:1 with dilution buffer (300 mM NaCl, 10 mM NaF, 0.2 mM ethylenediaminetetraacetic acid (EDTA), 0.2% Igepal CA-630, 20 mM HEPES-KOH, pH 7.4) supplemented with cOmplete protease inhibitor cocktail (Roche), and cleared by centrifugation. Dynabead-conjugated peptides were incubated with clarified extracts with end-to-end mixing at 4 °C for 2 h. Beads were washed with peptide pulldown buffer (150 mM NaCl, 50 mM KCl, 5 mM NaF, 0.2 mM EDTA, 0.1% Igepal CA-630, 10% glycerol, 20 mM HEPES-KOH, pH 7.4) supplemented with cOmplete protease inhibitor cocktail, before elution in 2× SDS sample buffer (4% SDS, 20% glycerol, 50 mM TCEP, 0.002% bromophenol blue, 125 mM Tris-HCl, pH 6.8) for mass spectrometry or SDS-PAGE. Peptides were biotinylated on the N-terminus, with the following sequences: motif 1 (SGSG-KKLEFSSSPDSLSTINDWDDMDDFDTS), motif 2 (SGSG-DYDTDFVPPSPEEII and SGSG-DYDTDFVPP-[pSer]-PEEII), motif 3 (SGSG-DIDNF-DIDDFDDDDDWE), motif 4 (SGSG-ANSKLGIMAPPKPINRPFLKPSYAFS), motif 5 (SGSG-LGPSDEELLASLDENDELTAN), and motif 6 (SGSG-DGELDDFSLEEALLEETVQKE).

For pulldowns with UV-crosslinking, biotinylated peptides containing BPA (Genosphere Biotechnologies) were incubated with clarified HeLa nuclear extracts as above. Where indicated, samples were exposed to light from a Stratalinker UV Crosslinker 2400 (Stratagene) for 1 h on ice, followed by washing with peptide pulldown buffer at 4 °C or denaturing wash buffer (137 mM NaCl, 3 mM KCl, 0.1% Tween 20, 1% SDS, 25 mM Tris-HCl, pH 7.5) at room temperature. Samples were eluted in 2× SDS sample buffer for SDS-PAGE. BPA-modified peptides were biotinylated on the N-terminus, with the following sequences: motif 1 (SGSG-KKLEFSSSPDSLSTIND-[BPA]-DDMDDFDTS), motif 3 (SGSG-DIDNFDIDD-[BPA]-DDDDDWE) and motif 5 (SGSG-LGPSDEELLAS-[BPA]-DENDELTAN).

For pulldowns with in vitro translated heterotrimeric RPA (incorporating either RPA1^{WT} or RPA1$^{R41E/Y42F}$), RPA was first transcribed and translated using the TNT Quick Coupled Transcription/Translation System (Promega) according to the manufacturer's instructions. In vitro translated complexes were diluted in modified (50 mM NaCl instead of 150 mM) peptide pulldown buffer supplemented with 1 mM $MgCl_2$, cOmplete protease inhibitor cocktail and 25 U/mL Benzonase (Novagen). After nuclease digestion, NaCl and EDTA concentrations were adjusted to 200 mM and 2 mM, respectively, and samples were cleared by centrifugation before being used for pulldowns with biotinylated peptides as above.

Mass spectrometry. Peptide pulldown eluates in 2× SDS sample buffer were subjected to two rounds of chloroform-methanol precipitation[105], followed by in-solution digestion with trypsin[106]. Desalted peptides were injected into a LC-MS/MS platform consisting of a Thermo Scientific UltiMate 3000 UHPLC and Q Exactive mass spectrometer. Samples were separated with 250 nl/min flow on a 50-cm EASY-Spray column with a gradient of 2–35% acetonitrile in 5% dimethyl sulfoxide (DMSO) and 0.1% formic acid. MS1 data were acquired with a resolution of 70,000 and an AGC target of 3E6. Selected precursors were fragmented with 28% normalized collision energy and fragment masses acquired (MS2) with a resolution of 17,500 and an AGC target of 1E5. Selected precursors were excluded for 27 s. Mass spectrometry data were converted into mgf files by msConvert[107], and analyzed with Mascot 2.5 (Matrix Science). The *Homo sapiens* Swissprot/TrEMBL database (retrieved 15/10/2014) was used for protein identification. Precursor mass tolerance was set to 10 ppm and fragment tolerance to 0.05 Da. Trypsin was selected as enzyme (one missed cleavage), carbamidomethyl (C) as fixed modification, and oxidation (M) and deamidation (NQ) as variable modification. Keratins and peptide identifications with an ion score of <20 were dismissed, and protein identification confidence was set to 1% false-discovery rate at identity or homology threshold.

SDS-PAGE and western blotting. SDS-PAGE and western blotting were performed using 7% Tris-Bicine gels with the SE400 and TE42 systems from Hoefer, or Bolt 4–12% Bis-Tris Plus gels from Thermo Fisher Scientific. The following antibodies were used at the indicated dilutions: BLM (A300-110A, Bethyl Laboratories, 1/2000), FLAG (F1804, Sigma-Aldrich, 1/2000), GFP (11814460001, Roche, 1/5000), MRE11 (PC388, Merck, 1/500) RMI1 (NB100-1720, Novus Biologicals, 1/1000), RMI2 (NBP1-89962, Novus Biologicals, 1/4000), RPA1 (NA13, Merck, 1/200), RPA2 (ab10359, Abcam, 1/10,000), RPA3 (HPA005708, Sigma-Aldrich, 1/1000), TOP3A (14525-1-1AP, Proteintech, 1/1000), TOPBP1 (A300-111A, Bethyl Laboratories, 1/2000), XPA (ab65963, Abcam, 1/2000). Uncropped and unprocessed scans of all blots can be found in the Source Data file associated with this paper.

Cell lines, cell culture conditions, and RNA interference. All cells were grown in humidified incubators supplied with 5% CO_2 and maintained at 37 °C, with regular testing to verify mycoplasma-free status using a LookOut Mycoplasma PCR Detection Kit (Sigma-Aldrich). RPE-1 FRT/TR cells[108] were purchased from Ximbio, and cultured in Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific) supplemented with 10% fetal calf serum (FCS; Thermo Fisher Scientific) and 100 µ/ml penicillin/100 µg/ml streptomycin (Lonza). 293FT cells were obtained from Thermo Fisher Scientific and cultured in DMEM supplemented with 10% FCS, 2 mM glutamine (Lonza), 1% MEM non-essential amino acids (Thermo Fisher Scientific) and 500 µg/ml Geneticin (G418; Thermo Fisher Scientific). siRNAs were transfected using Lipofectamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's instructions, with the following sequences: siCtrl (targeting firefly luciferase), 5′-CGUACGCGGAAUACUUCGA-3′; siRMI1, 5′-AGCCUUCACGAAUGUUGAU-3′.

Plasmids and cloning. All plasmids were maxiprepped using GenElute HP Endotoxin-Free Plasmid Maxiprep Kits (Sigma-Aldrich) according to the manufacturer's instructions, and verified by Sanger sequencing (Source BioScience). Plasmids expressing GFP-RMI1 (note the abbreviation "GFP" is used to include the various different *A. victoria* green fluorescent protein variants used in this study with similar excitation and emission wavelengths, including EGFP and Clover variants[109]) were constructed by subcloning human RMI1 cDNA (a gift from Andrew Deans) into mClover2-C1 (a gift from Michael Davidson; Addgene plasmid # 54577). Deletions of motifs 5 (residues 240–260) and 6 (residues 308–329) to generate plasmids expressing GFP-RMI1$^{\Delta 5}$ and GFP-RMI1$^{\Delta 6}$, respectively, were carried out by site-directed mutagenesis using QuikChange (Agilent Technologies).

Plasmids expressing GFP-tagged or FLAG-tagged BLM[25] were mutagenized to produce constructs with deletions in the TR domain (Δ4–37), motif 1 (Δ137–163), motif 2 (Δ292–309), and/or motif 3 (Δ552–568). Vectors expressing WT or R41E/Y42F RPA1 were a gift from Marc Wold[40]. Primers used in this study are listed in Supplementary Data 9.

Immunoprecipitations. Plasmids were transfected into 293FT cells using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's instructions. For preparation of lysates for IPs, cells were washed in phosphate-buffered saline (PBS), and lysed in IP buffer (100 mM NaCl, 0.2% Igepal CA-630, 1 mM $MgCl_2$, 10% glycerol, 5 mM NaF, 50 mM Tris-HCl, pH 7.5), supplemented with cOmplete EDTA-free protease inhibitor cocktail and 25 U/mL Benzonase. After nuclease digestion, NaCl and EDTA concentrations were adjusted to 200 mM and 2 mM, respectively, and lysates were cleared by centrifugation. Lysates were then incubated with 15 µl of GFP-Trap magnetic agarose beads (ChromoTek) for 2 h with end-to-end mixing at 4 °C. Immunoglobulin–antigen complexes were washed five times with IP buffer before elution in 2× SDS sample buffer for SDS-PAGE.

Generation of BLM knockout RPE-1 cells. Small guide RNAs (sgRNAs) were designed to target exon 4 of *BLM* using CRISPR Design Tool (https://crispr.mit.edu), targeting the following sequences: sgRNA-1 (anti-sense), CCACCTTCTC-CAGAAGAAATTAT; sgRNA-2 (sense), CTTCCTCTTCAAAATGCCTTAGG. sgRNA oligonucleotides (Sigma-Aldrich) were cloned into AIO-GFP (a gift from Steve Jackson[110]), to produce a vector expressing Cas9^{D10A} nickase, the two sgRNAs, and GFP for sorting successfully transfected cells by FACS. Plasmids were transfected into RPE-1 FRT/TR cells using Lipofectamine 3000 (Thermo Fisher Scientific), according to the manufacturer's instructions. Cells were allowed to recover for 3 days before sorting GFP-positive cells. Colonies were grown from single cells and screened for BLM expression by western blotting.

SCE assays. RPE-1 cells were treated with 10 µM BrdU for two cell cycles, and blocked in mitosis by addition of 750 nM colchicine for 2 h before harvesting. Cells were resuspended in 75 mM KCl and incubated at 37 °C for 30 min, followed by fixation by addition of ice-cold 3:1 methanol/acetic acid solution. Following gentle inversion and centrifugation, pellets were resuspended in fixative and spun down, with this procedure being repeated twice. Pellets were then resuspended in fixative, and cells were dropped onto glass slides that had been pre-treated with 0.432% HCl for 1 h, washed six times in water, and stored in ethanol. Cells were left to dry at room temperature for up to 3 days in the dark, before staining by immersion in 15 µg/ml Hoechst diluted in 100 mM phosphate buffer (51 mM NaH_2PO_4, 49 mM Na_2HPO_4, pH 6.8) for 30 min. Cells were washed in PBS, exposed to a 365-nm UV light source for 2 h, and washed again in PBS. This was followed by incubation in 2× saline-sodium citrate buffer (300 mM NaCl, 30 mM trisodium citrate-HCl, pH 7) for 1 h at 65 °C and four washes in water. Cells were stained with Leishman's stain (Sigma-Aldrich) for 90 s and left in the dark for 30 min to dry. Cells were then washed once in water, left to dry overnight in the dark before mounting in VECTASHIELD (Vector Laboratories) with glass coverslips (Appleton Woods). Images were acquired using a ZEISS Axio Observer Z1 widefield microscope fitted with an Axiocam MRm camera and a Plan-Apochromat ×100/1.4 oil objective, running on ZEN Pro 2012 software.

DNA-end resection assay. Cells were incubated with 10 µM 5-ethynyl-2'-deoxyuridine (EdU) and 1 µM CPT for 1 h as indicated. Cells were harvested by trypsinization, pre-extracted in 0.2% Triton X-100 in PBS for 10 min on ice, and fixed using BD Cytofix/Cytoperm solution (BD Biosciences) according to the manufacturer's instructions. Cells were labeled with RPA2 antibodies (ab2175, Abcam, 1/200) diluted in BD Perm/Wash buffer (BD Biosciences) for 1 h at room temperature, followed by incubation with Alexa Fluor 647 goat anti-mouse IgG1 secondary antibodies (A21240, Thermo Fisher Scientific, 1/200) diluted in BD Perm/Wash buffer for 1 h at room temperature in the dark. EdU was labeled with Alexa Fluor 555 using a Click-iT EdU Cell Proliferation Kit for Imaging (Thermo Fisher Scientific) according to the manufacturer's instructions. After washing in BD Perm/Wash buffer, cells were resuspended in PBS containing 250 µg/ml RNase A (Sigma-Aldrich), 0.02% sodium azide and 2 µg/ml DAPI for 30 min at 37 °C in the dark. Samples were resuspended in BD Perm/Wash buffer for analysis. Cell cycle profiles were generated using DAPI content and EdU intensity. DNA-end resection was quantified by subtracting the RPA2 intensity of untreated S phase cells from the RPA2 intensity of CPT-treated S phase cells, and results normalized to WT RPE-1 cells treated with siCtrl. See Supplementary Fig. 6 for further details of gating strategy.

Clonogenic survival assays. Cells were plated at low densities and treated with the indicated doses of HU for 24 h the following day. Cells were washed three times in PBS before addition of fresh medium, and left for up to 14 days to allow colonies to develop. Colonies were washed in PBS, stained in 0.1% Coomassie Brilliant Blue R 250, 7% acetic acid, 50% methanol at room temperature for 30 min, and washed in water before counting.

DNA fiber analyses. Cells were treated with 25 µM 5-iodo-2'-deoxyuridine (IdU) for 30 min, washed three times in PBS, before addition of 2 mM HU for 2 h. Cells were then washed three times in PBS before addition of 250 µM 5-chloro-2'-deoxyuridine (CldU) for 30 min. Cells were washed in PBS and harvested using Accutase (Thermo Fisher Scientific), before being embedded in low melting point agarose (Bio-Rad). Agarose plugs were solidified at 4 °C for 30 min, then incubated overnight at 50 °C in 20 mM NaCl, 100 mM EDTA, 0.5% SDS, 10 mM Tris-HCl pH 8 solution containing 1 mg/ml proteinase K (Sigma-Aldrich). Digested proteins were removed by two washes in 10 mM Tris-HCl, 1 mM EDTA (pH 8). Plugs were used immediately or stored at 4 °C in 500 µM EDTA (pH 8). Plugs were melted at 67 °C for 30 min, and cooled to 42 °C before addition of β-Agarase (New England Biolabs) according to the manufacturer's instructions. Plugs were left to digest overnight and the following day, MES buffer (pH 5.4) was added. DNA solution was heated to 65 °C for 30 min, and samples were left to cool to room temperature before transfer to the Teflon reservoir of a FiberComb Molecular Combing System (Genomic Vision). Silanized coverslips were dipped into the reservoir for 15 min and pulled out at a constant speed of 300 µm/s. Coverslips were baked overnight at 60 °C, and then incubated in 12 M HCl for 45 min with gentle shaking to denature the DNA. Coverslips were washed three times in PBS, before incubation in blocking buffer (1% bovine serum albumin, 0.2% Tween 20 in PBS) for 15 min. Samples were then incubated in the following primary antibodies diluted in blocking buffer for 1 h: anti-BrdU (347580, BD Biosciences, 1/25) to recognize IdU, and anti-BrdU (ab6326, Abcam, 1/400) to recognize CldU. Samples were washed three times in 0.2% Tween 20 in PBS, before incubation in the following secondary antibodies from Thermo Fisher Scientific diluted 1/500 in blocking buffer for 45 min: goat anti-rat IgG Alexa Fluor 488 (A11006) and donkey anti-mouse IgG Alexa Fluor 568 (A10037). Samples were washed three times in 0.2% Tween 20 in PBS and once in PBS alone, before mounting in ProLong Gold Antifade Mountant (Thermo Fisher Scientific). DNA fibers were analyzed using Fiji[111].

Lentiviral transductions. Generation of lentiviral vectors for complementation of *BLM*$^{-/-}$ RPE-1 cells with GFP-BLM and FLAG-RMI1 was done as follows. First, DNA encoding mClover3 followed by a multiple cloning site was synthesized (Thermo Fisher Scientific) and cloned into pcDNA5/FRT/TO-neo (a gift from Jon Pines; Addgene plasmid # 41000) to generate pcDNA5/FRT/TO-neo-mClover3. cDNAs encoding WT BLM or BLM lacking motifs 1 and 3 were subcloned into this vector to generate pcDNA5/FRT/TO-neo-mClover3-BLM^{WT} and pcDNA5/FRT/TO-neo-mClover3-BLM$^{\Delta RPA}$. DNA encoding P2A-myc-3xFLAG-RMI1 was synthesized with silent mutations to render it siRMI1-resistant (Thermo Fisher Scientific) and cloned into pcDNA5/FRT/TO-neo-mClover3-BLM^{WT} to generate pcDNA5/FRT/TO-neo-mClover3-BLM^{WT}-P2A-myc-3xFLAG-RMI1^{WT}, and into pcDNA5/FRT/TO-neo-mClover3-BLM$^{\Delta RPA}$ to generate pcDNA5/FRT/TO-neo-mClover3-BLM$^{\Delta RPA}$-P2A-myc-3xFLAG-RMI1^{WT}. The latter construct was mutagenized to delete motif 5 in RMI1 (residues 240–260) and generate pcDNA5/FRT/TO-neo-mClover3-BLM$^{\Delta RPA}$-P2A-myc-3xFLAG-RMI1$^{\Delta RPA}$. The WT and ΔRPA mClover3-BLM-P2A-myc-3xFLAG-RMI1 sequences were amplified by PCR, cloned into pDONR221 using Gateway Technology (Thermo Fisher Scientific), and then into pLenti PGK Neo DEST (a gift from Eric Campeau and Paul Kaufman[112]; Addgene plasmid # 19067) according to the manufacturer's instructions.

To assemble lentiviruses, 293FT cells were transfected with pLenti PGK Neo DEST plasmids encoding WT or mutant GFP-BLM and FLAG-RMI1 proteins or

GFP alone, along with pHDM-tat1b, pHDM-G, pRC/CMV-rev1b, and pHDM-Hgpm2 lentiviral assembly plasmids (all gifts from Ross Chapman[113]). Viral supernatants were harvested at 24 and 48 h, pooled, passed through a 0.45-µm filter, and stored at −80 °C.

For lentiviral infections, RPE-1 cells were incubated in viral media for 48 h, before selection in fresh medium containing 500 µg/ml G418 for up to 7 days. GFP-positive cells were sorted by FACS to generate stable lines for further study, followed by verification by western blotting.

Immunofluorescence and confocal microscopy. For analyses of BLM foci, RAD51 foci or UFBs, cells were grown on 18 × 18 mm #1.5H glass coverslips (Marienfeld Superior). For laser microirradiation, cells were grown in 35-mm #1.5H glass bottom dishes (Thistle Scientific). Cells were washed once in PBS before fixing for 10 min in 4% paraformaldehyde (PFA) in PBS at room temperature. When imaging UFBs, cells were instead fixed in 4% PFA in 250 mM HEPES (pH 7.0-7.6) and 0.1 % Triton X-100 for 20 min at 4 °C. PBS was added simultaneously during PFA removal to prevent samples from drying due to PFA evaporation. To visualize BLM recruitment, a mild pre-extraction step (ice-cold 0.2% Triton X-100 in PBS for 1 min followed by a wash in PBS) was carried out between the first PBS wash and the fixation step. After fixation, cells were permeabilized in 0.2% Triton X-100 in PBS for 5 min (or 15 min when imaging UFBs), and then washed once more in PBS before incubation with primary antibodies diluted in antibody buffer (DMEM supplemented with 10% FCS and 0.05% sodium azide; filtered) for 1 h. Where indicated, EdU was labeled with Alexa Fluor 647 using a Click-iT EdU Cell Proliferation Kit for Imaging (Thermo Fisher Scientific) according to the manufacturer's instructions prior to antibody incubations. The following primary antibodies were used at the indicated concentrations: Cyclin A (611268, BD Biosciences, 1/200), GFP (PABG1, ChromoTek, 1/500), γH2AX (05-636, Merck, 1/500), γH2AX (2577, Cell Signaling Technology, 1/500), RAD51 (70-002, BioAcademia, 1/1000), RIF1 (sc-515573, Santa Cruz Biotechnology, 1/50), RPA2 (ab2175, Abcam, 1/300). Cells were washed three times in 0.2% Tween 20 in PBS, before incubation with secondary antibodies diluted in antibody buffer containing 0.5 µg/mL DAPI in the dark for 30 min, followed by five washes in 0.2% Tween 20 in PBS. The following secondary antibodies were from Thermo Fisher Scientific and diluted 1/1000: goat anti-rabbit IgG Alexa Fluor 488 (A11034), donkey anti-mouse IgG Alexa Fluor 568 (A10037), donkey anti-rabbit IgG Alexa Fluor 568 (A10042) and goat anti-mouse IgG1 Alexa Fluor 647 (A21240). Cells in glass bottom dishes were imaged directly in PBS, whereas cells on coverslips were washed once in water and mounted in VECTASHIELD (unless imaging samples with Alexa Fluor 647, in which case SlowFade Diamond Antifade Mountant (Thermo Fisher Scientific) was used to avoid fluorescence quenching[114]).

All images were acquired using an LSM 880 Airyscan inverted microscope (ZEISS) equipped with a ×63/1.4 NA Plan-Apochromat objective and an Airyscan 32-pinhole detector unit. Foci were quantified using Fiji. For Airyscan imaging, raw data were processed using Airyscan processing with Wiener Filter strength 6 using ZEN Black software version 2.1, yielding 16-bit images with approximately 180-nm lateral resolution. Display of images was adjusted for intensity for optimal display of structures of interest, and false-colored in magenta or cyan where indicated.

For laser microirradiation, cells were grown for 24 h in the presence of 10 µM BrdU, before being exposed to the 405-nm diode laser from the same microscope used for imaging. Laser energy output was determined by biological calibration.

Statistics and reproducibility. All experiments in the paper were performed independently at least twice, with similar results.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Datasets generated for this study are available from the corresponding author upon reasonable request. Proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository[115] with the data set identifier ["PXD018322" http://www.ebi.ac.uk/pride/archive/projects/PXD018322]. Source data are provided with this paper.

Received: 10 April 2020; Accepted: 22 December 2020;
Published online: 26 January 2021

References

1. Ellis, N. A. et al. The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* **83**, 655–666 (1995).
2. German, J. Bloom syndrome: a mendelian prototype of somatic mutational disease. *Med. (Baltim.)* **72**, 393–406 (1993).
3. Chu, W. K. & Hickson, I. D. RecQ helicases: multifunctional genome caretakers. *Nat. Rev. Cancer* **9**, 644–654 (2009).
4. Chaganti, R. S., Schonberg, S. & German, J. A manyfold increase in sister chromatid exchanges in Bloom's syndrome lymphocytes. *Proc. Natl. Acad. Sci. USA* **71**, 4508–4512 (1974).
5. Cunniff, C., Bassetti, J. A. & Ellis, N. A. Bloom's syndrome: clinical spectrum, molecular pathogenesis, and cancer predisposition. *Mol. Syndromol.* **8**, 4–23 (2017).
6. Bizard, A. H. & Hickson, I. D. The dissolution of double Holliday junctions. *Cold Spring Harb. Perspect. Biol.* **6**, a016477 (2014).
7. Hudson, D. F. et al. Loss of RMI2 increases genome instability and causes a bloom-like syndrome. *PLoS Genet.* **12**, e1006483 (2016).
8. Martin, C.-A. et al. Mutations in TOP3A cause a Bloom syndrome-like disorder. *Am. J. Hum. Genet.* **103**, 221–231 (2018).
9. Groden, J., Nakamura, Y. & German, J. Molecular evidence that homologous recombination occurs in proliferating human somatic cells. *Proc. Natl. Acad. Sci. USA* **87**, 4315–4319 (1990).
10. Luo, G. et al. Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nat. Genet.* **26**, 424–429 (2000).
11. van Wietmarschen, N. et al. BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes. *Nat. Commun.* **9**, 271 (2018).
12. Gravel, S., Chapman, J. R., Magill, C. & Jackson, S. P. DNA helicases Sgs1 and BLM promote DNA double-strand break resection. *Genes Dev.* **22**, 2767–2772 (2008).
13. Chan, K.-L., North, P. S. & Hickson, I. D. BLM is required for faithful chromosome segregation and its localization defines a class of ultrafine anaphase bridges. *EMBO J.* **26**, 3397–3409 (2007).
14. Davies, S. L., North, P. S. & Hickson, I. D. Role for BLM in replication-fork restart and suppression of origin firing after replicative stress. *Nat. Struct. Mol. Biol.* **14**, 677–679 (2007).
15. Bhat, K. P. & Cortez, D. RPA and RAD51: fork reversal, fork protection, and genome stability. *Nat. Struct. Mol. Biol.* **25**, 446–453 (2018).
16. Meetei, A. R. et al. A multiprotein nuclear complex connects Fanconi anemia and Bloom syndrome. *Mol. Cell Biol.* **23**, 3417–3426 (2003).
17. Brosh, R. M. et al. Replication protein A physically interacts with the Bloom's syndrome protein and stimulates its helicase activity. *J. Biol. Chem.* **275**, 23500–23508 (2000).
18. Doherty, K. M. et al. Physical and functional mapping of the replication protein a interaction domain of the werner and bloom syndrome helicases. *J. Biol. Chem.* **280**, 29494–29505 (2005).
19. Xue, X., Raynard, S., Busygina, V., Singh, A. K. & Sung, P. Role of replication protein A in double holliday junction dissolution mediated by the BLM-Topo IIIα-RMI1-RMI2 protein complex. *J. Biol. Chem.* **288**, 14221–14227 (2013).
20. Manthei, K. A. & Keck, J. L. The BLM dissolvasome in DNA replication and repair. *Cell Mol. Life Sci.* **70**, 4067–4084 (2013).
21. Tompa, P., Davey, N. E., Gibson, T. J. & Babu, M. M. A million peptide motifs for the molecular biologist. *Mol. Cell* **55**, 161–169 (2014).
22. Bocquet, N. et al. Structural and mechanistic insight into Holliday-junction dissolution by topoisomerase IIIα and RMI1. *Nat. Struct. Mol. Biol.* **21**, 261–268 (2014).
23. Hoadley, K. A. et al. Structure and cellular roles of the RMI core complex from the bloom syndrome dissolvasome. *Structure* **18**, 1149–1158 (2010).
24. Wang, F. et al. Crystal structures of RMI1 and RMI2, two OB-fold regulatory subunits of the BLM complex. *Structure* **18**, 1159–1170 (2010).
25. Blackford, A. N. et al. TopBP1 interacts with BLM to maintain genome stability but is dispensable for preventing BLM degradation. *Mol. Cell* **57**, 1133–1141 (2015).
26. Sun, L. et al. Structural Insight into BLM Recognition by TopBP1. *Structure* **25**, 1582–1588 (2017). e3.
27. Mellacheruvu, D. et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* **10**, 730–736 (2013).
28. Raynard, S. et al. Functional role of BLAP75 in BLM-topoisomerase IIIalpha-dependent holliday junction processing. *J. Biol. Chem.* **283**, 15701–15708 (2008).
29. Xu, D. et al. RMI, a new OB-fold complex essential for Bloom syndrome protein to maintain genome stability. *Genes Dev.* **22**, 2843–2855 (2008).
30. Hu, P. et al. Evidence for BLM and Topoisomerase IIIalpha interaction in genomic stability. *Hum. Mol. Genet.* **10**, 1287–1298 (2001).
31. Kennedy, J. A., Daughdrill, G. W. & Schmidt, K. H. A transient α-helical molecular recognition element in the disordered N-terminus of the Sgs1 helicase is critical for chromosome stability and binding of Top3/Rmi1. *Nucleic Acids Res.* **41**, 10215–10227 (2013).
32. Bochkareva, E. et al. Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proc. Natl. Acad. Sci. USA* **102**, 15412–15417 (2005).
33. Olson, E. et al. The Mre11 complex mediates the S-phase checkpoint through an interaction with replication protein A. *Mol. Cell Biol.* **27**, 6053–6067 (2007).

34. Xu, X. et al. The basic cleft of RPA70N binds multiple checkpoint proteins, including RAD9, to regulate ATR signaling. *Mol. Cell Biol.* **28**, 7345–7353 (2008).
35. Zhou, C., Pourmal, S. & Pavletich, N. P. Dna2 nuclease-helicase structure, mechanism and regulation by Rpa. *eLife* **4**, 213 (2015).
36. Feng, S. et al. Ewing tumor-associated antigen 1 interacts with replication protein a to promote restart of stalled replication forks. *J. Biol. Chem.* **291**, 21956–21962 (2016).
37. Bass, T. E. et al. ETAA1 acts at stalled replication forks to maintain genome integrity. *Nat. Cell Biol.* **18**, 1185–1195 (2016).
38. Haahr, P. et al. Activation of the ATR kinase by the RPA-binding protein ETAA1. *Nat. Cell Biol.* **18**, 1196–1207 (2016).
39. Guilliam, T. A. et al. Molecular basis for PrimPol recruitment to replication forks by RPA. *Nat. Commun.* **8**, 15222 (2017).
40. Haring, S. J., Mason, A. C., Binz, S. K. & Wold, M. S. Cellular functions of human RPA1. Multiple roles of domains in replication, repair, and checkpoints. *J. Biol. Chem.* **283**, 19095–19111 (2008).
41. Stigger, E., Drissi, R. & Lee, S. H. Functional analysis of human replication protein A in nucleotide excision repair. *J. Biol. Chem.* **273**, 9337–9343 (1998).
42. Walther, A. P., Gomes, X. V., Lao, Y., Lee, C. G. & Wold, M. S. Replication protein A interactions with DNA. 1. Functions of the DNA-binding and zinc-finger domains of the 70-kDa subunit. *Biochemistry* **38**, 3963–3973 (1999).
43. Li, W. & Wang, J. C. Mammalian DNA topoisomerase IIIalpha is essential in early embryogenesis. *Proc. Natl. Acad. Sci. USA* **95**, 1010–1013 (1998).
44. Seki, M. et al. Bloom helicase and DNA topoisomerase IIIalpha are involved in the dissolution of sister chromatids. *Mol. Cell Biol.* **26**, 6299–6307 (2006).
45. Mönnich, M. et al. Developing T lymphocytes are uniquely sensitive to a lack of topoisomerase III alpha. *Eur. J. Immunol.* **40**, 2379–2384 (2010).
46. Chen, H., You, M. J., Jiang, Y., Wang, W. & Li, L. RMI1 attenuates tumor development and is essential for early embryonic survival. *Mol. Carcinog.* **50**, 80–88 (2011).
47. Chester, N., Kuo, F., Kozak, C., O'Hara, C. D. & Leder, P. Stage-specific apoptosis, developmental delay, and embryonic lethality in mice homozygous for a targeted disruption in the murine Bloom's syndrome gene. *Genes Dev.* **12**, 3382–3393 (1998).
48. Wang, W. et al. Possible association of BLM in decreasing DNA double strand breaks during DNA replication. *EMBO J.* **19**, 3428–3435 (2000).
49. Tomimatsu, N. et al. Exo1 plays a major role in DNA end resection in humans and influences double-strand break repair and damage signaling decisions. *DNA Repair (Amst.)* **11**, 441–448 (2012).
50. Sturzenegger, A. et al. DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *J. Biol. Chem.* **289**, 27314–27326 (2014).
51. Lu, H. et al. RECQL4 promotes DNA end resection in repair of DNA double-strand breaks. *CellReports* **16**, 161–173 (2016).
52. Patel, D. S., Misenko, S. M., Her, J. & Bunting, S. F. BLM helicase regulates DNA repair by counteracting RAD51 loading at DNA double-strand break sites. *J. Cell Biol.* **216**, 3521–3534 (2017).
53. Mendez-Dorantes, C., Tsai, L. J., Jahanshir, E., Lopezcolorado, F. W. & Stark, J. M. BLM has contrary effects on repeat-mediated deletions, based on the distance of DNA DSBs to a repeat and repeat divergence. *CellReports* **30**, 1342–1357 (2020).
54. Forment, J. V., Walker, R. V. & Jackson, S. P. A high-throughput, flow cytometry-based method to quantify DNA-end resection in mammalian cells. *Cytom. A* **81**, 922–928 (2012).
55. Davies, S. L., North, P. S., Dart, A., Lakin, N. D. & Hickson, I. D. Phosphorylation of the Bloom's syndrome helicase and its role in recovery from S-phase arrest. *Mol. Cell Biol.* **24**, 1279–1291 (2004).
56. Fukao, T. et al. Disruption of the BLM gene in ATM-null DT40 cells does not exacerbate either phenotype. *Oncogene* **23**, 1498–1506 (2004).
57. Pichierri, P., Franchitto, A. & Rosselli, F. BLM and the FANC proteins collaborate in a common pathway in response to stalled replication forks. *EMBO J.* **23**, 3154–3163 (2004).
58. So, S., Adachi, N., Lieber, M. R. & Koyama, H. Genetic interactions between BLM and DNA ligase IV in human cells. *J. Biol. Chem.* **279**, 55433–55442 (2004).
59. Imamura, O. et al. Bloom helicase is involved in DNA surveillance in early S phase in vertebrate cells. *Oncogene* **20**, 1143–1151 (2001).
60. Lahkim Bennani-Belhaj, K., Buhagiar-Labarchède, G., Jmari, N., Onclercq-Delic, R. & Amor-Guéret, M. BLM distinct replication fork protection pathway connects fanconi anemia tumor suppresess. *J. Nucleic Acids* **2010**, 1–8 (2010).
61. Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* **22**, 106–116 (2012).
62. Liberi, G. et al. Rad51-dependent DNA structures accumulate at damaged replication forks in sgs1 mutants defective in the yeast ortholog of BLM RecQ helicase. *Genes Dev.* **19**, 339–350 (2005).
63. Karmakar, P. et al. BLM is an early responder to DNA double-strand breaks. *Biochem. Biophys. Res. Commun.* **348**, 62–69 (2006).
64. Byun, T. S., Pacek, M., Yee, M.-C., Walter, J. C. & Cimprich, K. A. Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev.* **19**, 1040–1052 (2005).
65. Yin, J. et al. BLAP75, an essential component of Bloom's syndrome protein complexes that maintain genome integrity. *EMBO J.* **24**, 1465–1476 (2005).
66. Aleksandrov, R. et al. Protein dynamics in complex DNA lesions. *Mol. Cell* **69**, 1046–1061 (2018). e5.
67. Chan, K. L. & Hickson, I. D. On the origins of ultra-fine anaphase bridges. *Cell Cycle* **8**, 3065–3066 (2009).
68. Hengeveld, R. C. C. et al. Rif1 is required for resolution of ultrafine DNA bridges in anaphase to ensure genomic stability. *Dev. Cell* **34**, 466–474 (2015).
69. Sarlós, K. et al. Reconstitution of anaphase DNA bridge recognition and disjunction. *Nat. Struct. Mol. Biol.* **25**, 868–876 (2018).
70. Yang, J., O'Donnell, L., Durocher, D. & Brown, G. W. RMI1 Promotes DNA Replication Fork Progression and Recovery from Replication Fork Stress. *Mol. Cell Biol.* **32**, 3054–3064 (2012).
71. Shi, J. et al. A helical bundle in the N-terminal domain of the BLM helicase mediates dimer and potentially hexamer formation. *J. Biol. Chem.* **292**, 5909–5920 (2017).
72. Kang, D. et al. Interaction of replication protein A with two acidic peptides from human Bloom syndrome protein. *FEBS Lett.* **592**, 547–558 (2018).
73. Plank, J. L., Wu, J. & Hsieh, T.-S. Topoisomerase IIIalpha and Bloom's helicase can resolve a mobile double Holliday junction substrate through convergent branch migration. *Proc. Natl Acad. Sci. USA* **103**, 11118–11123 (2006).
74. Cejka, P., Plank, J. L., Bachrati, C. Z., Hickson, I. D. & Kowalczykowski, S. C. Rmi1 stimulates decatenation of double Holliday junctions during dissolution by Sgs1-Top3. *Nat. Struct. Mol. Biol.* **17**, 1377–1382 (2010).
75. Nimonkar, A. V. et al. BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes Dev.* **25**, 350–362 (2011).
76. Soniat, M. M., Myler, L. R., Kuo, H.-C., Paull, T. T. & Finkelstein, I. J. RPA Phosphorylation Inhibits DNA Resection. *Mol. Cell* **75**, 145–153 (2019).
77. Zhu, Z., Chung, W.-H., Shim, E. Y., Lee, S. E. & Ira, G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* **134**, 981–994 (2008).
78. Mimitou, E. P. & Symington, L. S. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* **455**, 770–774 (2008).
79. Bae, K.-H. et al. Bimodal interaction between replication-protein A and Dna2 is critical for Dna2 function both in vivo and in vitro. *Nucleic Acids Res.* **31**, 3006–3015 (2003).
80. Kim, D.-H. et al. Enzymatic properties of the Caenorhabditis elegans Dna2 endonuclease/helicase and a species-specific interaction between RPA and Dna2. *Nucleic Acids Res.* **33**, 1372–1383 (2005).
81. Cejka, P. et al. DNA end resection by Dna2-Sgs1-RPA and its stimulation by Top3-Rmi1 and Mre11-Rad50-Xrs2. *Nature* **467**, 112–116 (2010).
82. Niu, H. et al. Mechanism of the ATP-dependent DNA end-resection machinery from Saccharomyces cerevisiae. *Nature* **467**, 108–111 (2010).
83. Tammaro, M., Liao, S., McCane, J. & Yan, H. The N-terminus of RPA large subunit and its spatial position are important for the 5'->3' resection of DNA double-strand breaks. *Nucleic Acids Res.* **43**, 8790–8800 (2015).
84. Ke, Y. et al. PICH and BLM limit histone association with anaphase centromeric DNA threads and promote their resolution. *EMBO J.* **30**, 3309–3321 (2011).
85. Sun, H., Karow, J. K., Hickson, I. D. & Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.* **273**, 27587–27592 (1998).
86. Ralf, C., Hickson, I. D. & Wu, L. The Bloom's syndrome helicase can promote the regression of a model replication fork. *J. Biol. Chem.* **281**, 22839–22846 (2006).
87. Rosado, I. V., Niedzwiedz, W., Alpi, A. F. & Patel, K. J. The Walker B motif in avian FANCM is required to limit sister chromatid exchanges but is dispensable for DNA crosslink repair. *Nucleic Acids Res.* **37**, 4360–4370 (2009).
88. Deans, A. J. & West, S. C. FANCM connects the genome instability disorders Bloom's Syndrome and Fanconi Anemia. *Mol. Cell* **36**, 943–953 (2009).
89. Willis, N. A. et al. Mechanism of tandem duplication formation in BRCA1-mutant cells. *Nature* **551**, 590–595 (2017).
90. Li, W., Kim, S.-M., Lee, J. & Dunphy, W. G. Absence of BLM leads to accumulation of chromosomal DNA breaks during both unperturbed and disrupted S phases. *J. Cell Biol.* **165**, 801–812 (2004).
91. Hegnauer, A. M. et al. An N-terminal acidic region of Sgs1 interacts with Rpa70 and recruits Rad53 kinase to stalled forks. *EMBO J.* **31**, 3768–3783 (2012).
92. Ahamad, N., Khan, S. & Xu, Y.-J. RecQ DNA helicase Rqh1 promotes Rad3ATR kinase signaling in the DNA replication checkpoint pathway of fission yeast. *Mol. Cell Biol.* **40**, (2020).

93. Lecointe, F. et al. Anticipating chromosomal replication fork arrest: SSB targets repair DNA helicases to active forks. *EMBO J.* **26**, 4239–4251 (2007).
94. Higgs, M. R. et al. BOD1L is required to suppress deleterious resection of stressed replication forks. *Mol. Cell* **59**, 462–477 (2015).
95. Bugreev, D. V., Yu, X., Egelman, E. H. & Mazin, A. V. Novel pro- and anti-recombination activities of the Bloom's syndrome helicase. *Genes Dev.* **21**, 3085–3094 (2007).
96. Crickard, J. B. et al. The RecQ helicase Sgs1 drives ATP-dependent disruption of Rad51 filaments. *Nucleic Acids Res.* **47**, 4694–4706 (2019).
97. Hu, Y. et al. RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. *Genes Dev.* **21**, 3073–3084 (2007).
98. van Brabant, A. J. et al. Binding and melting of D-loops by the Bloom syndrome helicase. *Biochemistry* **39**, 14617–14625 (2000).
99. Bachrati, C. Z., Borts, R. H. & Hickson, I. D. Mobile D-loops are a preferred substrate for the Bloom's syndrome helicase. *Nucleic Acids Res.* **34**, 2269–2279 (2006).
100. Fasching, C. L., Cejka, P., Kowalczykowski, S. C. & Heyer, W.-D. Top3-Rmi1 dissolve Rad51-mediated D loops by a topoisomerase-based mechanism. *Mol. Cell* **57**, 595–606 (2015).
101. Piazza, A. et al. Dynamic processing of displacement loops during recombinational DNA repair. *Mol. Cell* **73**, 1255–1266 (2019). e4.
102. LaRocque, J. R. et al. Interhomolog recombination and loss of heterozygosity in wild-type and Bloom syndrome helicase (BLM)-deficient mammalian cells. *Proc. Natl Acad. Sci. USA* **108**, 11971–11976 (2011).
103. Blackford, A. N. & Jackson, S. P. ATM, ATR, and DNA-PK: the trinity at the heart of the DNA daage response. *Mol. Cell* **66**, 801–817 (2017).
104. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
105. Wessel, D. & Flügge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143 (1984).
106. Fischer, R. & Kessler, B. M. Gel-aided sample preparation (GASP)–a simplified method for gel-assisted proteomic sample generation from protein extracts and intact cells. *Proteomics* **15**, 1224–1229 (2015).
107. Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
108. Mansfeld, J., Collin, P., Collins, M. O., Choudhary, J. S. & Pines, J. APC15 drives the turnover of MCC-CDC20 to make the spindle assembly checkpoint responsive to kinetochore attachment. *Nat. Cell Biol.* **13**, 1234–1243 (2011).
109. Bajar, B. T. et al. Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci. Rep.* **6**, 20889 (2016).
110. Chiang, T.-W. W., le Sage, C., Larrieu, D., Demir, M. & Jackson, S. P. CRISPR-Cas9(D10A) nickase-based genotypic and phenotypic screening to enhance genome editing. *Sci. Rep.* **6**, 24356 (2016).
111. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
112. Campeau, E. et al. A versatile viral system for expression and depletion of proteins in mammalian cells. *PLoS ONE* **4**, e6529 (2009).
113. Becker, J. R. et al. The ASCIZ-DYNLL1 axis promotes 53BP1-dependent non-homologous end joining and PARP inhibitor sensitivity. *Nat. Commun.* **9**, 5406 (2018).
114. Arsić, A., Stajković, N., Spiegel, R. & Nikić-Spiegel, I. Effect of Vectashield-induced fluorescence quenching on conventional and super-resolution microscopy. *Sci. Rep.* **10**, 6441 (2020).
115. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).
116. Ishihama, Y. et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell Proteom.* **4**, 1265–1272 (2005).

Acknowledgements

We thank Ross Chapman, Michael Davidson, Andrew Deans, Steve Jackson, Jon Pines, Marc Wold, Eric Campeau, and Paul Kaufman for providing valuable reagents. We are grateful to David Clynes, Ester Hammond, and Steve Jackson for critical reading of the manuscript and helpful discussions. We thank Jill Brown, Rosa Camarillo, and Pablo Huertas for help setting up DNA combing, Nausica Arnoult and Alan Saghatelian for providing BPA peptide pulldown advice, Rebecca Konietzny for help with mass spectrometry, and Jill Brown and Wojciech Niedzwiedz for assistance with SCE assays. We acknowledge Kevin Clark, Sally-Ann Clark, Paul Sopp, and Craig Waugh in the MRC Weatherall Institute of Molecular Medicine (WIMM) Flow Cytometry Facility for providing cell-sorting services and technical expertize; Dominic Waithe, Silvia Galliani, Jana Koth, and Christoffer Lagerholm in the Wolfson Imaging Centre for help with microscopy; and Ryan Beveridge in the MRC WIMM Virus Screening Facility for help in generation of lentiviral preparations. These facilities are supported by the MRC Molecular Hematology Unit (MC_UU_12009), the MRC Human Immunology Unit (MC_UU_12010), the Wolfson Foundation (grant 18272), the Wellcome Trust (Micron 107457/Z/15Z), the National Institute for Health Research Oxford Biomedical Research Centre (IS-BRC-1215-20008), the Cancer Research UK (CRUK) Oxford Centre, the Kay Kendall Leukemia Fund (KKL1057), the John Fell Fund (131/030 and 101/517), the EPA fund (CF182 and CF170) and the MRC WIMM Strategic Alliance (G0902418 and MC_UU_12025). The Blackford lab is supported by a CRUK Career Development Fellowship (C29215/A20772) and an Against Breast Cancer/Oriel College Research Fellowship to A.N.B., and an MRC WIMM DPhil Prize Studentship (MR/N013468/1) to A.-M.K.S. K.T. is supported by the Tokyo Tech Academy for Co-creative Education of Environment and Energy Science, the Tokyo Tech Academy for Leadership, and a Grant-in-Aid Fellowship from the Japan Society for the Promotion of Science (JP20J13601). J.S. is supported by a Tetelman Fellowship for International Research in the Sciences from Yale College.

Author contributions

The project was conceived and supervised by A.N.B. Experiments were performed by A.-M.K.S., S.E.J., K.T., Z.B., J.S., and A.N.B. Reagents were generated by S.E.J., K.T., C.A.M., and A.N.B. Mass spectrometry data were contributed by I.V., R.F., and B.M.K. The paper was written by A.N.B. with contributions from A.-M.K.S., S.E.J., and K.T.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-20818-5.

Correspondence and requests for materials should be addressed to A.N.B.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

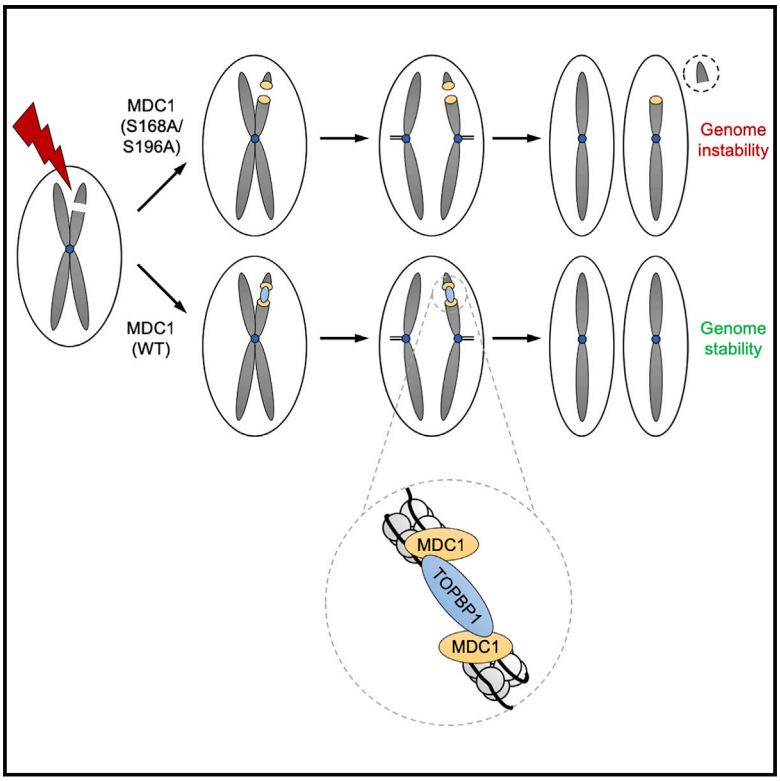
© The Author(s) 2021

Article

Molecular Cell

MDC1 Interacts with TOPBP1 to Maintain Chromosomal Stability during Mitosis

Graphical Abstract



Authors

Pia-Amata Leimbacher,
Samuel E. Jones,
Ann-Marie K. Shorrocks, ...,
Laurence H. Pearl,
Andrew N. Blackford, Manuel Stucki

Correspondence

andrew.blackford@oncology.ox.ac.uk (A.N.B.),
manuel.stucki@uzh.ch (M.S.)

In Brief

Leimbacher et al. reveal that two DNA damage-response mediator proteins, MDC1 and TOPBP1, form a complex required for maintaining chromosome stability when DNA double-strand breaks occur during mitosis, possibly by tethering broken chromosomes together until DNA-repair pathways are reactivated in the following G1 phase.

Highlights

- MDC1 recruits TOPBP1 to DNA double-strand breaks specifically in mitosis
- MDC1, when phosphorylated on S168 and S196, directly binds TOPBP1 BRCT domains 1 and 2
- Loss of MDC1-TOPBP1 binding causes mitotic radiosensitivity and genome instability
- TOPBP1 forms filamentous assemblies that can bridge MDC1 foci in mitosis

Leimbacher et al., 2019, Molecular Cell *74*, 571–583
May 2, 2019 © 2019 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.molcel.2019.02.014

CellPress

Molecular Cell
Article

MDC1 Interacts with TOPBP1 to Maintain Chromosomal Stability during Mitosis

Pia-Amata Leimbacher,[1,6] Samuel E. Jones,[2,3,6] Ann-Marie K. Shorrocks,[2,3] Mara de Marco Zompit,[1] Matthew Day,[4] Jordy Blaauwendraad,[1] Diana Bundschuh,[1] Sarah Bonham,[5] Roman Fischer,[5] Daniel Fink,[1] Benedikt M. Kessler,[5] Antony W. Oliver,[4] Laurence H. Pearl,[4] Andrew N. Blackford,[2,3,*] and Manuel Stucki[1,7,*]

[1]Department of Gynecology, University Hospital and University of Zurich, Wagistrasse 14, 8952 Schlieren, Switzerland
[2]Department of Oncology, Medical Research Council Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK
[3]Cancer Research UK/Medical Research Council Oxford Institute for Radiation Oncology, University of Oxford, Oxford OX3 7DQ, UK
[4]Cancer Research UK DNA Repair Enzymes Group, Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Falmer BN1 9RQ, UK
[5]Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK
[6]These authors contributed equally
[7]Lead Contact
*Correspondence: andrew.blackford@oncology.ox.ac.uk (A.N.B.), manuel.stucki@uzh.ch (M.S.)
https://doi.org/10.1016/j.molcel.2019.02.014

SUMMARY

In mitosis, cells inactivate DNA double-strand break (DSB) repair pathways to preserve genome stability. However, some early signaling events still occur, such as recruitment of the scaffold protein MDC1 to phosphorylated histone H2AX at DSBs. Yet, it remains unclear whether these events are important for maintaining genome stability during mitosis. Here, we identify a highly conserved protein-interaction surface in MDC1 that is phosphorylated by CK2 and recognized by the DNA-damage response mediator protein TOPBP1. Disruption of MDC1-TOPBP1 binding causes a specific loss of TOPBP1 recruitment to DSBs in mitotic but not interphase cells, accompanied by mitotic radiosensitivity, increased micronuclei, and chromosomal instability. Mechanistically, we find that TOPBP1 forms filamentous structures capable of bridging MDC1 foci in mitosis, indicating that MDC1-TOPBP1 complexes tether DSBs until repair is reactivated in the following G1 phase. Thus, we reveal an important, hitherto-unnoticed cooperation between MDC1 and TOPBP1 in maintaining genome stability during cell division.

INTRODUCTION

DNA double-strand breaks (DSBs) are particularly toxic DNA lesions that must be repaired accurately in order to avoid genome instability, cell death, or cancer (Jackson and Bartek, 2009). Interphase cells respond to DSBs by triggering a signaling cascade to activate cell-cycle checkpoints and DNA repair. In contrast, in mitotic cells there is no DNA damage checkpoint after prophase (Rieder and Cole, 1998), and DSBs are transmitted into the following G1 phase for repair to avoid chromosomal instability (Lee et al., 2014; Orthwein et al., 2014).

The cellular response to DSBs is regulated by three related protein kinases, ataxia-telangiectasia mutated (ATM), ATM and Rad3-related (ATR), and DNA-dependent protein kinase (DNA-PK) (Blackford and Jackson, 2017). Upon DNA damage, one of the earliest substrates of these kinases is the histone variant H2AX, which is phosphorylated at DSB sites on Ser139 and then referred to as γH2AX (Rogakou et al., 1999). γH2AX is recognized by MDC1 (Stucki et al., 2005), a scaffold protein that acts as a platform for recruitment of various DNA damage response factors to mediate DNA repair. One of these is the MRE11-RAD50-NBS1 (MRN) complex, which binds to MDC1 via a direct interaction between the NBS1 subunit of MRN and multiple acidic sequence motifs near the N terminus of MDC1 (Chapman and Jackson, 2008; Hari et al., 2010; Melander et al., 2008; Spycher et al., 2008; Wu et al., 2008). Another is RNF8, an E3 ubiquitin ligase with an FHA domain that binds to a cluster of conserved threonine residues in MDC1 that are phosphorylated by ATM in response to DSBs to promote chromatin ubiquitylation events required for recruitment of DNA damage response mediator proteins such as 53BP1 and BRCA1 (Huen et al., 2007; Kolas et al., 2007; Mailand et al., 2007). Recruitment of these factors to chromatin-flanking DSB sites channels DNA repair into either the non-homologous end-joining pathway or homology-directed repair via mechanisms that are still not completely understood (Hustedt and Durocher, 2016).

γH2AX and MDC1 form foci at DSBs throughout the cell cycle, but recruitment of downstream factors such as RNF8 and 53BP1 is blocked during mitosis (Giunta et al., 2010; Nakamura et al., 2010; Nelson et al., 2009; van Vugt et al., 2010; Lee et al., 2014; Orthwein et al., 2014). However, given that inhibition of ATM and DNA-PK activity in mitosis causes radiosensitivity, it is possible that DNA damage signaling as well as recruitment of MDC1 and potentially some of its downstream factors, play an as-yet unidentified role in dealing with DNA damage in this cell-cycle phase.

Here, we identify two highly conserved motifs in MDC1 and show that they are phosphorylated by casein kinase 2 (CK2).

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

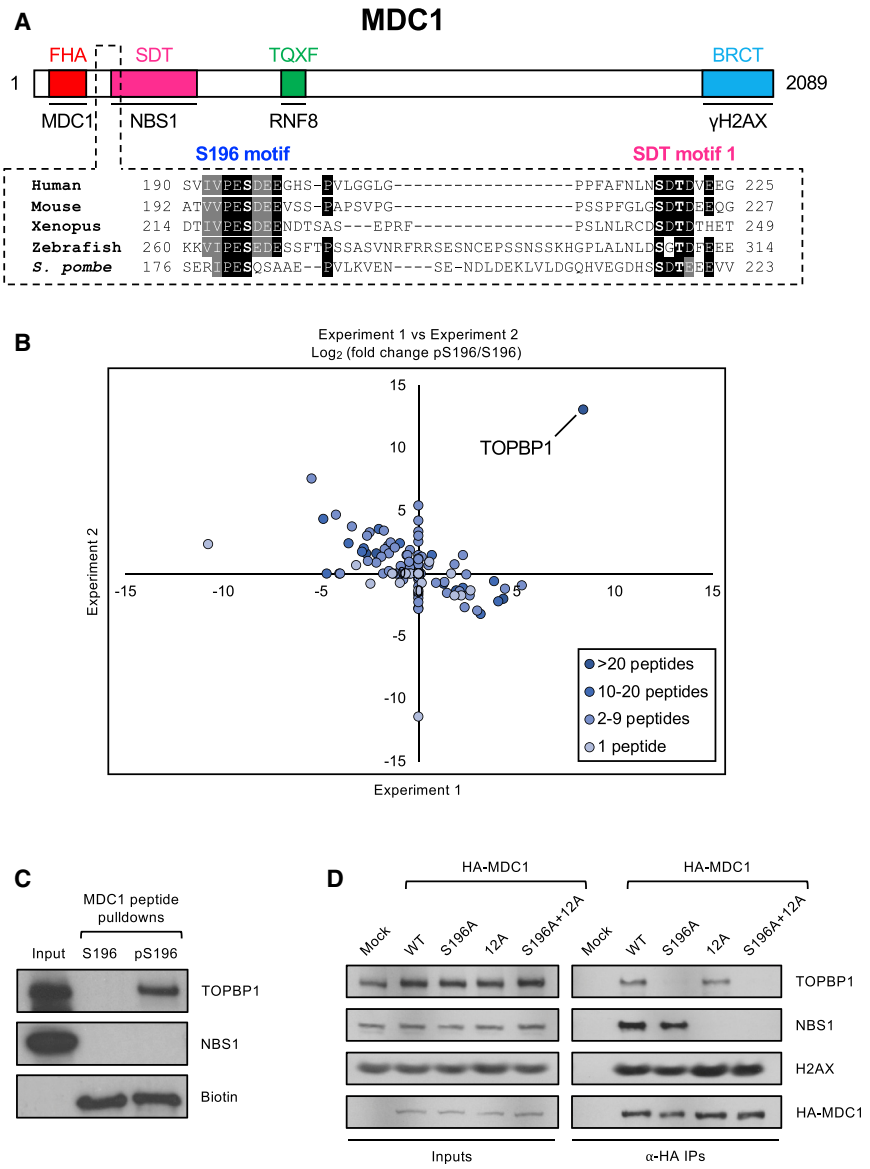


Figure 1. A Conserved Acidic Sequence Motif near the N Terminus of MDC1 Binds to TOPBP1

(A) Schematic showing the layout of conserved domains and motifs in MDC1. Names of the known MDC1 binding partners, NBS1, RNF8, and H2AX, are shown below the motifs with which they interact. The FHA domain promotes MDC1 dimerization; hence, its binding partner is MDC1. Key phosphorylated residues are highlighted in bold.

(B) Identification of TOPBP1 as a specific interactor for the MDC1-S196 phosphopeptide by LC-MS/MS and label-free quantification. Scatterplot depicts $\log_2$ fold enrichment of MDC1-S196 versus MDC1-pS196 peptide-binding proteins from 2 independent experiments. See Table S1 for raw MS data.

(C) Peptide pull-downs from HeLa nuclear extracts using biotinylated peptides corresponding to residues surrounding MDC1-S196, either native (S196) or phosphorylated (pS196).

(D) HA immunoprecipitations from 293FT cells transfected with the indicated HA-tagged MDC1 variants.

We identify the DNA damage response mediator protein TOPBP1 as the binding partner for these motifs and demonstrate that the MDC1-TOPBP1 interaction is specifically required for TOPBP1 recruitment to DSBs in mitosis. Loss of MDC1-TOPBP1 binding leads to radiosensitivity in mitotic cells, as well as increased micronuclei formation, chromosome/chromatid breaks, and chromosome end-to-end fusions.

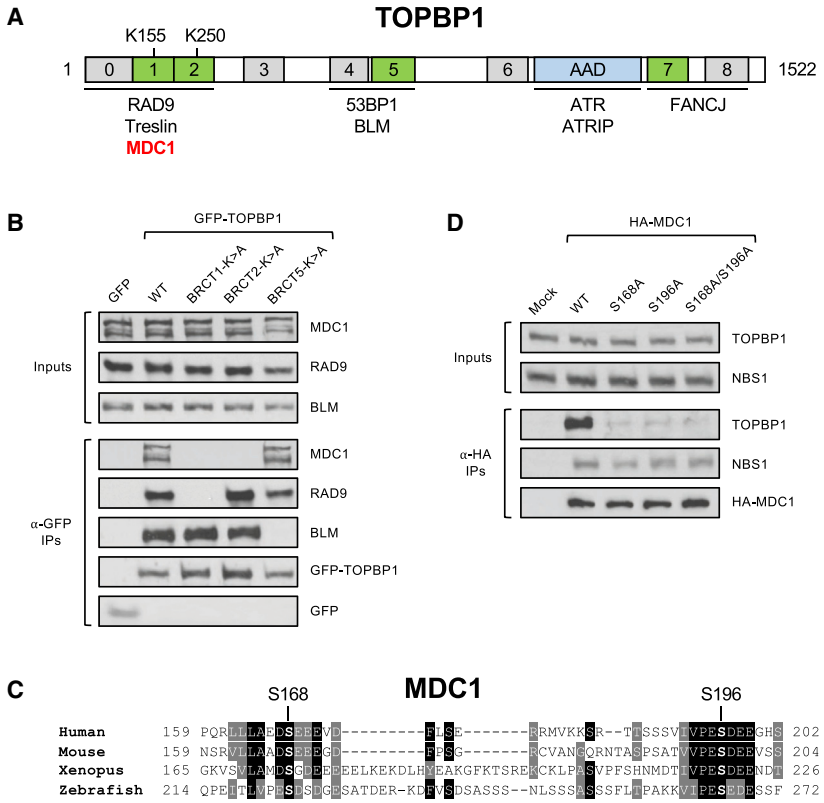
RESULTS

A Conserved Acidic Sequence Motif near the N Terminus of MDC1 Binds to TOPBP1

Previously, we and others identified six conserved acidic sequence motifs near the N terminus of MDC1 that directly interact with NBS1 and are required for MRN foci formation at sites of DSBs (Chapman and Jackson, 2008; Melander et al., 2008; Spycher et al., 2008; Wu et al., 2008). These motifs contain Ser-Asp-Thr residues embedded in an acidic sequence environment; hence, they have been named SDT motifs (Figure 1A). The serine and threonine residues within the SDT motifs are constitutively phosphorylated by CK2, and the unique bivalent N-terminal FHA/BRCT2 phospho-binding module of NBS1 specifically recognizes and binds to the doubly phosphorylated MDC1 SDT motifs (Hari et al., 2010; Lloyd et al., 2009).

As part of an ongoing screen to identify and characterize short evolutionarily conserved linear peptide motifs mediating protein-protein interactions in the DNA damage response, we noticed one such potentially novel sequence in MDC1. Centered around S196, it is located downstream of the FHA domain and just upstream of the first SDT repeat (Figure 1A). This motif is highly conserved in MDC1 orthologs and is composed of a Val/Ile-Pro-Glu-Ser consensus followed by a stretch of acidic amino acids in vertebrate MDC1. Given that S196 is phosphorylated in human cells (Yi et al., 2014), we considered the possibility that the S196 motif constitutes a previously uncharacterized interaction surface for a DNA damage response factor with phosphopeptide-binding activity.

To test this possibility, we designed a synthetic biotinylated peptide corresponding to human MDC1 residues 180–205, bearing a phospho-serine at the S196 position. This phosphopeptide and its native equivalent were conjugated to streptavidin-coupled beads and incubated in HeLa nuclear extracts, and potential interacting partners were identified from two independent experiments by liquid chromatography-tandem mass spectrometry (LC-MS/MS; Table S1). When we examined proteins present only in the phospho-MDC1 peptide pull-downs,



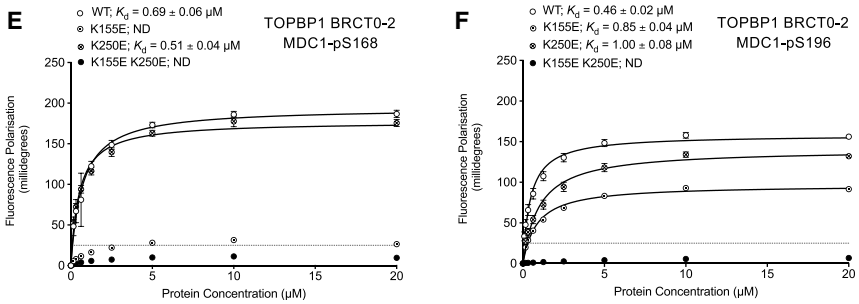


Figure 2. BRCT Domains 1 and 2 of TOPBP1 Interact with MDC1 via Two Phosphorylated Residues, Ser168 and Ser196

(A) Schematic showing the layout of conserved domains and motifs in TOPBP1. Numbered boxes represent BRCT domains, with phosphopeptide-binding domains in green and domains lacking phosphopeptide-binding activity in gray. Names of known TOPBP1 binding partners are shown below the domains they interact with. AAD = ATR-activation domain.

(B) GFP pull-downs from 293FT cells transfected with the indicated GFP-tagged TOPBP1 constructs.

(C) Sequence alignment showing the conservation of Ser168, Ser196, and surrounding residues in MDC1 in vertebrates. Key phospho-serines are highlighted in bold.

(D) HA-immunoprecipitations from 293FT cells transfected with the indicated HA-tagged MDC1 variants.

(E) Fluorescence polarization with recombinant TOPBP1 BRCT domains 0–2 and MDC1-pS168 phosphopeptide. K155E is a mutation in TOPBP1 BRCT domain 1; K250E is a mutation in BRCT domain 2. ND = not determined. Dotted line indicates threshold for specific protein-protein interactions.

(F) Fluorescence polarization with recombinant TOPBP1 BRCT domains 0–2 and MDC1-pS196 phosphopeptide.

See also Figure S1.

TOPBP1 was identified as the top hit (Figure 1B), which, given that it contains multiple BRCT domains that specifically recognize and bind phosphorylated proteins and was previously implicated in MDC1 binding (Blackford et al., 2015; Leung and Glover, 2011; Wang et al., 2011), we decided to pursue further.

Western blotting confirmed that TOPBP1 was specifically pulled down by the phosphorylated S196 peptide (Figure 1C). NBS1, on the other hand, previously shown to interact efficiently with doubly phosphorylated SDT phosphopeptides (Chapman and Jackson, 2008; Lloyd et al., 2009; Spycher et al., 2008), was not pulled down by either S196 peptide. These findings reveal a hitherto unnoticed TOPBP1 interaction site near the N terminus of MDC1 that is functionally distinct from the SDT repeats.

We next investigated whether S196 is required for MDC1-TOPBP1 interaction in the context of full-length MDC1. To do this, we mutated S196 to alanine in constructs expressing hemagglutinin (HA)-tagged wild-type (WT) MDC1 or a mutant MDC1 that cannot bind NBS1 due to alanine substitutions in all six SDT repeats (12A; Melander et al., 2008). Immunoprecipitation with HA affinity beads showed that the MDC1-S196A mutation abolished TOPBP1 binding whereas interaction with NBS1 remained intact (Figure 1D). In contrast, the MDC1-12A mutant was completely defective for NBS1 binding but could still interact with TOPBP1, whereas the double MDC1-S196A+12A mutant could bind to neither. We conclude from these results that MDC1 binding to TOPBP1 and NBS1 can be functionally separated, and that MDC1-TOPBP1 binding is mediated by the novel MDC1-S196 motif and not by the SDT repeats.

BRCT Domains 1 and 2 of TOPBP1 Interact with MDC1

TOPBP1 contains 9 BRCT domains, some of which occur in tandem, whereas others appear to function as single domains (Figure 2A). Of the BRCT domains in TOPBP1, only 1, 2, 5, and 7 contain the necessary lysine residues required for phosphopeptide binding (Leung and Glover, 2011). In line with this, BRCT0–2 interact with phosphorylated RAD9 or Treslin (Delacroix et al., 2007; Kumagai et al., 2010; Lee et al., 2007), BRCT4+5 bind to phosphorylated 53BP1 or BLM (Blackford et al., 2015; Cescutti et al., 2010), and BRCT7+8 interact with phosphorylated FANCJ (Gong et al., 2010).

Previously, we found that TOPBP1 interacts with MDC1 in a manner dependent on BRCT1 but not BRCT5 or BRCT7 (Blackford et al., 2015). However, the requirement for BRCT2 was not tested. Therefore, we mutated K250 in BRCT2 to alanine in GFP-tagged full-length TOPBP1 and compared it in co-immunoprecipitation experiments to the BRCT1 mutant we generated previously. As a control in these assays, we also included a mutant BRCT5 construct, because this domain interacts with BLM but not MDC1 (Blackford et al., 2015). As demonstrated previously, mutation of TOPBP1 BRCT1 abolished interaction with MDC1 and RAD9 but not BLM, whereas mutation of BRCT5 had no effect on MDC1 but BLM binding was lost (Figure 2B). Surprisingly, mutation of BRCT2 also abolished MDC1 binding, while leaving BLM and RAD9 unaffected. This suggested the existence of an additional TOPBP1 binding site in MDC1. Indeed, a conserved motif consisting of a serine residue (S168 in human MDC1) embedded in an acidic sequence environment is located just upstream of S196 (Figure 2C). Like S196, S168 was previously shown to be phosphorylated (Beausoleil et al., 2004). As this sequence motif strikingly resembles the TOPBP1-BRCT1 interaction site in RAD9 (Delacroix et al., 2007; Lee et al., 2007) (Figure S1A), we tested whether it functions as an additional TOPBP1 binding site in MDC1. Indeed, MDC1-S168A mutants were completely defective for TOPBP1 interaction and no further decrease in binding was observed for a S168A/S196A double mutant (Figure 2D). This suggested that the two phospho-binding sites in BRCT1 and BRCT2 of TOPBP1 directly contact phosphorylated S168 and S196 in MDC1. However, it remained unclear which BRCT domain of TOPBP1 contacts which phospho-epitope in MDC1. To fine-map the TOPBP1 interaction sites in MDC1, we assessed the binding affinity of S168 and S196 phosphopeptides for purified TOPBP1 BRCT0–2 by fluorescence polarization. This quantitative biophysical analysis revealed that purified TOPBP1 BRCT0–2 strongly binds to S168 and S196 phosphopeptides with an apparent dissociation constant (K_d) of less than 1 μM (Figure 2E). Interestingly, a K155E (BRCT1) mutant did not bind to the S168 phosphopeptide, indicating that BRCT1 mainly contacts the phosphorylated S168 motif. In contrast, both K155E and K250E (BRCT2) mutants were partially defective for binding to the phosphorylated S196 peptide, suggesting reduced specificity for TOPBP1 binding to this motif (Figure 2F). In contrast, MDC1-SDT phosphopeptides did not bind with significant affinity to either purified BRCT0–2 or BRCT4+5 (Figure S1B). These data support a model in which TOPBP1 BRCT1 interacts with phosphorylated S168 of MDC1, whereas BRCT2 most likely docks with phosphorylated S196.

MDC1 Phosphorylation and TOPBP1 Binding Are Mediated by CK2

Next, we wished to determine the kinase responsible for phosphorylation of MDC1 on S168 and S196. Both motifs are highly acidic and conform to the preferred Ser/Thr-x-x-Asp/Glu consensus sequence for CK2 (Meggio and Pinna, 2003). Therefore, we generated antibodies that specifically recognize phosphorylated MDC1-pS168 or MDC1-pS196 (Figure 3A) and then used them to examine the effect on MDC1 phosphorylation of treating cells with the selective CK2 inhibitor CX-4945 (Siddiqui-Jain et al., 2010). Results from this experiment indicated that S168 and S196 phosphorylation are CK2 dependent, because phosphorylation of both sites was abolished upon CX-4945 treatment (Figure 3B) but still occurred in cells treated with inhibitors of ATM, ATR, DNA-PK, or PLK1 (Figures S2A–S2D). Furthermore, CX-4945 did not affect DNA damage signaling events mediated by these kinases (Figure S2E). In addition, the interaction between MDC1 and TOPBP1 was undetectable in cells treated with the CK2 inhibitor (Figure 3C). To confirm the CK2-dependent interaction *in vitro*, we established that recombinant CK2 could directly phosphorylate a GST-tagged MDC1 fragment purified from *E. coli*, specifically on S168 and S196 (Figure 3D). Next, GST-pull-down experiments were carried out with this fragment and its mutant derivatives. The WT MDC1 fragment (containing the S168 and S196 motifs as well as two SDT repeats) pulled down significant quantities of TOPBP1 and NBS1 from HeLa nuclear extract only when pre-incubated with recombinant CK2 and ATP. In line with our data above, incorporation of the S196A mutation into the GST-MDC1 fragment abolished interaction with TOPBP1 (Figure 3E). Furthermore, mutation of the SDT repeats (SDT > ADA) abolished NBS1 but not TOPBP1 binding. Taken together, we conclude that CK2 phosphorylation of MDC1 on both S168 and S196 is required for direct binding to TOPBP1.

Direct Interaction with MDC1 Is Essential for TOPBP1 Recruitment to DSBs in Mitotic Cells

Next, we sought to investigate the biological role of the MDC1-TOPBP1 interaction. To this end, we knocked out endogenous MDC1 in the human osteosarcoma cell line U2OS using CRISPR-Cas9 (henceforth termed Δ*MDC1*; Figures 4A and 4B). Δ*MDC1* cells do not display a general ATM or ATR signaling defect in response to ionizing radiation (IR) but are, as expected, unable to recruit NBS1 and 53BP1 to IR-induced foci (Figures 4B and 4C). In addition, γH2AX foci intensity is reduced in Δ*MDC1* cells (Figure 4D), consistent with previous reports (Lou et al., 2006; Stewart et al., 2003; Stucki et al., 2005). We also found that western blot signals from our phospho-specific MDC1-pS168 and -pS196 antibodies were absent in Δ*MDC1* cells, as expected (Figure 4E).

Using WT U2OS, Δ*MDC1*, and Δ*53BP1* cells (Orthwein et al., 2015), we found that the TOPBP1 foci pattern induced by IR varies during the cell cycle (Figure 4F), with large foci that colocalize with γH2AX forming in G1 phase cells, and smaller foci that colocalize with the single-stranded DNA (ssDNA)-binding protein complex RPA forming in S/G2 phase cells. Moreover, the large TOPBP1 foci in G1 phase cells, but not the small S/G2 phase foci, were dependent on both MDC1 and 53BP1 (Figures 4F and S3A; Cescutti et al., 2010). Next, we addressed the question of whether MDC1 can recruit TOPBP1 in the absence of 53BP1 in G1 phase cells. While all IR-induced TOPBP1 foci were absent from MDC1-deficient G1 phase cells, TOPBP1 foci that colocalize with MDC1 could still be observed in Δ*53BP1* cells, although they are much smaller than those observed in WT cells (Figure 4G).

To assess the contribution of the MDC1-TOPBP1 interaction to TOPBP1 foci formation, we stably transfected Δ*MDC1* cells with GFP-tagged WT MDC1, three TOPBP1 interaction mutants (S168A, S196A, and a S168A/S196A double mutant), and the

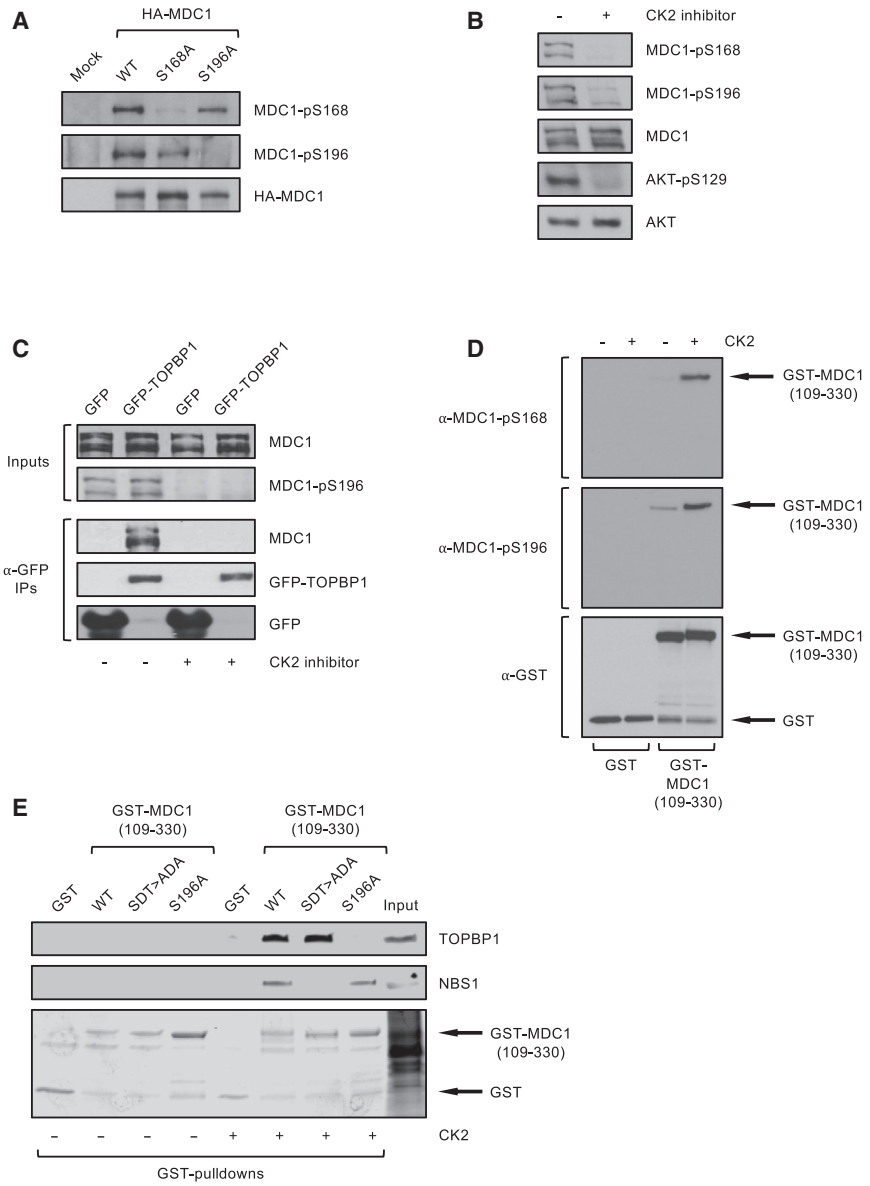


Figure 3. MDC1 Phosphorylation and TOPBP1 Binding Are Mediated by CK2

(A) HA-immunoprecipitations from 293FT cells transfected with the indicated constructs.

(B) 293FT cells treated with 10 μM CK2 inhibitor CX-4945 were harvested for western blotting with the indicated antibodies. AKT-pS129 is a positive control as AKT is a known CK2 substrate (Siddiqui-Jain et al., 2010).

(C) GFP pull-downs from 293FT cells transfected with the indicated GFP-tagged TOPBP1 constructs and treated with 10 μM CK2 inhibitor CX-4945 or DMSO vehicle control.

(D) Western blots of an *in vitro* kinase assay with recombinant CK2 as the kinase, and GST-tagged MDC1 fragment (encompassing residues 109–330) or GST alone as substrates.

(E) GST pull-downs from HeLa nuclear extracts with WT and mutant versions of GST-tagged MDC1 fragment (encompassing residues 109–330), pre-incubated or not with recombinant CK2 and ATP. See also Figure S2.

RNF8 interaction mutant TQXF > AQXF that cannot promote 53BP1 recruitment (Huen et al., 2007; Kolas et al., 2007; Mailand et al., 2007; Figure S3B). Visual inspection of these cells revealed IR-dependent accumulation of MDC1 and TOPBP1 in foci but did not show any apparent difference between cells expressing WT MDC1 and cells expressing the TOPBP1 binding mutants. In contrast, cells expressing the MDC1 TQXF > AQXF mutant were defective in forming TOPBP1 foci in G1 phase cells (Figure 5A). This indicated that TOPBP1 binding to MDC1 is not required for its efficient recruitment to sites of DSBs in interphase and suggested that TOPBP1 is predominately recruited by the RNF8-RNF168-53BP1 pathway. However, a more quantitative assessment of TOPBP1 foci formation by computer-aided segmentation and quantification of subcellular shapes (SQUASSH; Rizk et al., 2014) revealed a mild but significant reduction in MDC1-TOPBP1 colocalization in cells expressing the TOPBP1 binding mutants (Figures 5B, S3C, and S3D). These data suggest that while interaction with MDC1 mildly contributes to TOPBP1 accumulation in G1 phase cells, the RNF8-RNF168-53BP1 axis plays a much more dominant role. In addition, these data also suggest that despite its ability to form a stable complex with TOPBP1 *in vitro*, MDC1 is not able to substitute fully for 53BP1 as a TOPBP1 adaptor in G1 phase cells.

Next, because 53BP1 foci do not form during mitosis (Giunta et al., 2010; Nakamura et al., 2010; Nelson et al., 2009; van Vugt et al., 2010), we explored TOPBP1 recruitment to DSB sites in mitotic cells. After IR, TOPBP1 readily accumulates in foci that colocalize with γH2AX and MDC1 on mitotic chromosomes, both in U2OS cells and in the non-transformed human retinal pigmented epithelial cell line RPE-1 (Figures 5C and S3E). Significantly, while TOPBP1 foci formation on mitotic chromosomes was still dependent on γH2AX and MDC1, it occurred independently of 53BP1 (Figures 5D, 5E, and S3E). These observations suggested that, in contrast to G1 phase cells, MDC1 may recruit TOPBP1 to sites of DSBs directly in mitosis, through phosphorylation-dependent interactions with the S168 and S196 motifs. Indeed, foci containing the MDC1 mutant versions that are defective for TOPBP1 interaction were devoid of TOPBP1 on mitotic chromosomes, while expression of WT MDC1 in Δ*MDC1* cells fully restored TOPBP1 foci (Figures 5F and 5G). Consistent with our observation that TOPBP1 foci on mitotic chromosomes form independently of 53BP1, the TQXF > AQXF RNF8 binding mutant also rescued TOPBP1 foci in mitosis (Figures 5F and 5G). Finally, in line with our interpretation that MDC1 recruits TOPBP1 by direct

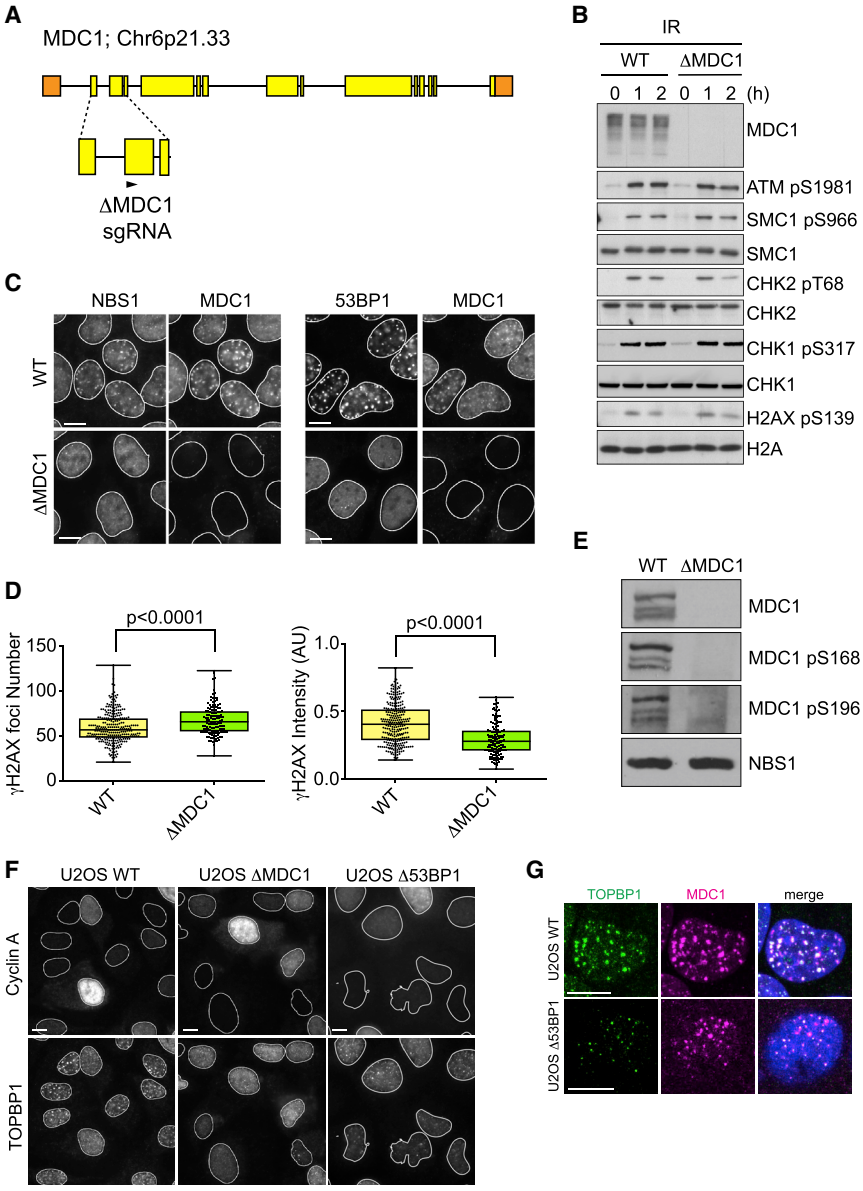


Figure 4. MDC1 Is Required for TOPBP1 Recruitment to Sites of DSBs in G1 Phase Cells

(A) Schematic representation of the human *MDC1* gene locus, illustrating the hybridization site of the gRNA selected for the generation of Δ*MDC1* cell line used in this study (gRNA sequence in the Key Resources Table).

(B) Western blots of total cell extract of irradiated WT U2OS cells and Δ*MDC1* cells showing that the ATM and ATR pathways are normally activated in the knock-out cell line in response to IR (3 Gy).

(C) Immunofluorescence experiment of irradiated WT U2OS cells and Δ*MDC1* cells stained with NBS1 and 53BP1 antibodies. Cells were co-stained with MDC1 antibodies to show lack of MDC1 expression in the knock-out cell line.

(D) Quantitative assessment of γH2AX foci in irradiated WT U2OS cells and Δ*MDC1* cells (box and whiskers represent minimum to maximum and individual data points are also shown; t test, $\alpha = 0.05$, at least 130 cells per condition).

(E) Western blot showing no signal with the MDC1-pS168 and -pS196 phospho-specific antibodies in Δ*MDC1* cells. NBS1 is a loading control.

(F) Immunofluorescence experiment of WT U2OS cells, Δ*MDC1* cells, and Δ*53BP1* cells stained with TOPBP1 antibodies 3 h after IR (3 Gy). Cells were co-stained with Cyclin A antibodies to distinguish G1 phase from S/G2 phase cells.

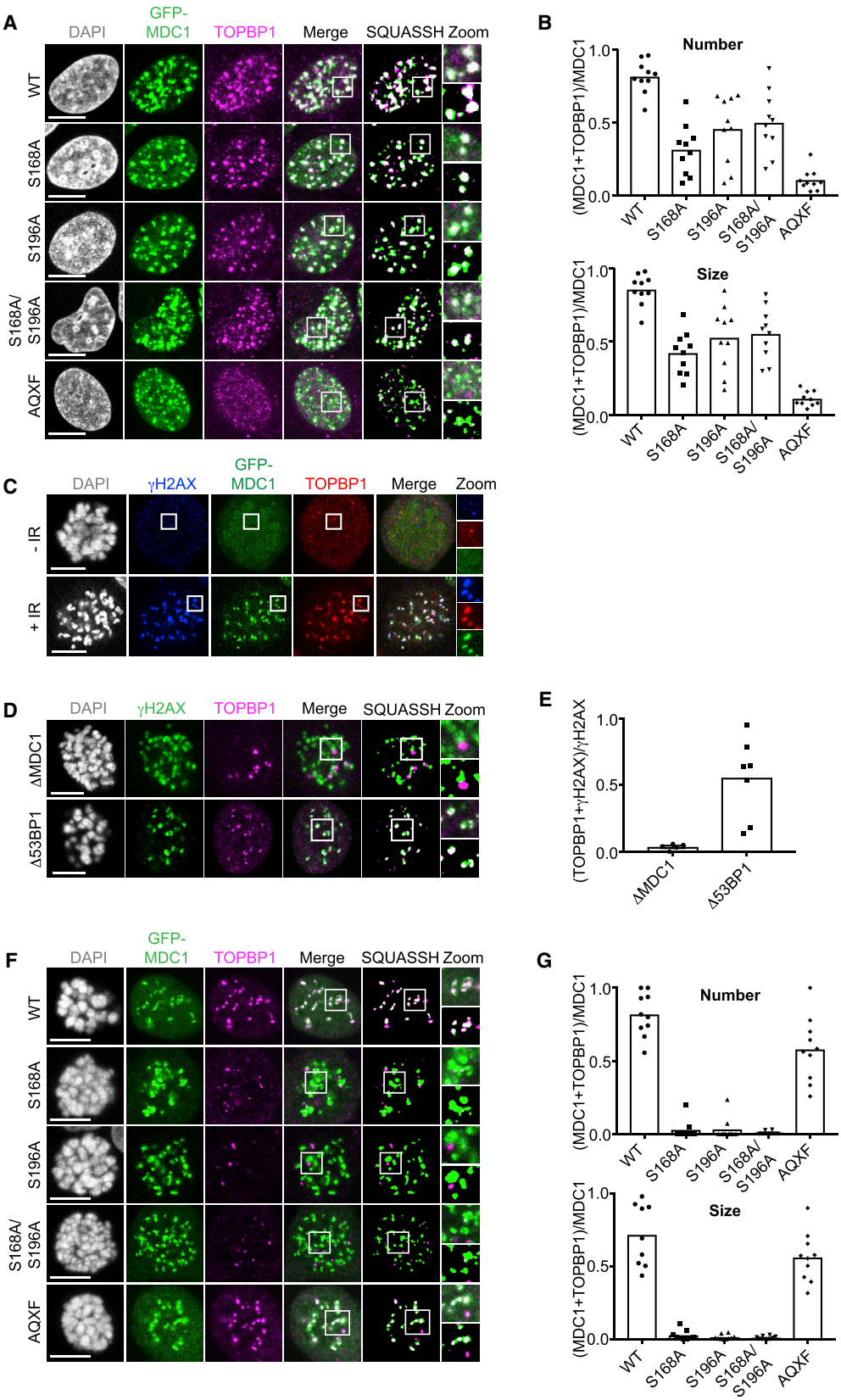
(G) Immunofluorescence experiment of irradiated WT U2OS cells and Δ*53BP1* G1 cells stained with MDC1 and TOPBP1 antibodies 3 h after IR (3 Gy).

All scale bars represent 10 μm. See also Figure S3.

interaction predominantly in mitosis, mutation of BRCT1 or BRCT2, but not BRCT5 or BRCT7, in GFP-tagged full-length TOPBP1 led to defective foci formation in mitosis (Figure S3F). In summary, these data reveal an unexpected complexity in the TOPBP1 recruitment mechanism to sites of DSBs: MDC1 controls TOPBP1 foci both in G1 phase and in mitosis, but only in mitosis does it primarily mediate TOPBP1 recruitment through direct interaction, as in G1 phase cells MDC1 predominantly controls TOPBP1 accumulation indirectly via the RNF8-RNF168-53BP1 axis.

MDC1-TOPBP1 Interaction Is Required for Maintenance of Chromosomal Stability during Mitosis

Next, we addressed the physiological consequences for cells lacking MDC1-TOPBP1 binding. First, given its specific role in the recruitment of TOPBP1 to sites of DSBs in mitosis, we sought to test whether protecting cells from the killing effects of DSBs acquired during mitosis may be a major physiological role of the MDC1-TOPBP1 interaction. Therefore, we synchronized parental U2OS, Δ*MDC1* cells, and Δ*MDC1* cells stably expressing GFP-tagged WT MDC1 or TOPBP1 binding mutants in mitosis and treated them with IR. Δ*MDC1* and MDC1-S168A/S196A-expressing mitotic cells, but not interphase cells, displayed marked radiosensitivity compared to WT MDC1-expressing cells (Figures 6A and S4A). To explore whether disruption of the MDC1-TOPBP1 interaction negatively affects DSB repair in the subsequent interphase after irradiation in mitosis, we synchronized parental U2OS, Δ*MDC1* cells, and Δ*MDC1* cells stably expressing WT MDC1 or TOPBP1 binding mutants in mitosis before irradiating them. As a marker for unrepaired DSBs, γH2AX foci were quantified 20 min, 6 h, and 24 h post-irradiation. There was a significant increase in residual γH2AX foci in cells expressing no MDC1 or the TOPBP1 binding mutants 24 h post-irradiation compared to cells expressing WT MDC1, indicating that in the absence of TOPBP1 recruitment in mitosis, a substantial number of DSBs are not repaired within this time period (Figure 6B).



(legend on next page)

During mitosis, untethered DNA breaks lead to the formation of acentric chromatin fragments that fail to segregate because they cannot interact with the mitotic spindle. These acentric fragments are randomly segregated to the cytoplasm of one of the daughter cells and are converted into micronuclei after nuclear envelope reformation (Bizard and Hickson, 2018). To test whether MDC1 and TOPBP1 act to prevent this, cells were first arrested in pro-metaphase and DSBs were induced by IR, followed by release from the mitotic block. Populations of parental U2OS and Δ*MDC1* cells expressing WT MDC1 contained few cells with micronuclei 4 h after release (Figure 6C). In contrast, Δ*MDC1* cells and Δ*MDC1* cells expressing TOPBP1 binding mutants showed significantly elevated numbers of cells with micronuclei, with more than 10% of cells in the population being accompanied by one or several micronuclei. This suggests that MDC1-mediated TOPBP1 recruitment to broken mitotic chromosomes may serve to stabilize clastogen-induced DSBs during mitosis for repair in the subsequent interphase. Next, given that a small number of γH2AX foci can be detected in almost every metaphase even in the absence of exogenous sources of genotoxic stress (Lukas et al., 2011), we addressed the consequences of loss of MDC1 and disruption of the MDC1-TOPBP1 interaction in unperturbed cells. Δ*MDC1* cell populations showed a significantly increased fraction of cells with micronuclei as compared to parental U2OS cells (Figure 6D). The majority of these micronuclei contained acentric chromatin fragments as they stained negative for the centromeric protein CENPA. Stable expression of WT MDC1, but not the TOPBP1-binding mutants, reduced the number of micronuclei to the level observed in parental U2OS cells, indicating that MDC1-TOPBP1 complex formation is important to stabilize spontaneous chromosome breaks during mitosis, possibly by bridging and tethering DNA ends (Figure 6D). The increase in spontaneous micronuclei formation upon disruption of the γH2AX-MDC1 cascade was even more pronounced in non-transformed RPE-1 cells, which generally feature much lower numbers of spontaneous micronuclei than cancer cells (Figure S4B). In line with this, we observed that a significantly increased number of chromosomes in metaphase spreads of Δ*MDC1* cells and S196A-expressing cells displayed abnormal structures, including, but not limited to, chromatid breaks, chromosome breaks, and dicentric chromosomes (Figures 7A, 7B, S5A, and S5B).

To obtain evidence for a DNA-end bridging role for MDC1 and TOPBP1, we performed Airyscan high-resolution confocal imaging of MDC1 and GFP-TOPBP1 foci in mitotic U2OS cells. Strikingly, using this technique, many IR-induced TOPBP1 structures in mitosis no longer resembled diffraction-limited spots but instead resembled filamentous assemblies (Figures 7C and 7D). In contrast, most MDC1 foci did not appear to form such structures but instead resembled cloud-like diffraction-limited spots. We also observed in all irradiated cells incidences of a TOPBP1 filament bridging two MDC1 foci, either within (intra) or occasionally between (inter) chromosomes (Figure 7C). In contrast, we observed few TOPBP1 foci colocalizing with or bridging MDC1 foci in unirradiated mitotic cells (Figure S6A). Nonetheless, we did occasionally observe TOPBP1 filaments in mitotic cells in the absence of exogenous DNA damage, but these did not colocalize with condensed chromatin (Figure S6A) and may represent structures that give rise to ultra-fine DNA bridges in anaphase, as observed previously (Broderick et al., 2015; Pedersen et al., 2015). TOPBP1 filaments were not observed in interphase cells, regardless of DNA damage.

In summary, our data show that the MDC1-TOPBP1 interaction is required for the maintenance of chromosomal stability and suggest that it does so by tethering broken chromosomes during mitosis.

DISCUSSION

The findings reported here reveal a previously unrecognized but highly conserved bi-partite protein interaction surface near the N terminus of MDC1 that is phosphorylated by CK2. An unbiased search for factors that interact with this region uncovered TOPBP1 as the primary phosphorylation-dependent interaction partner of this region of MDC1 and BRCT0–BRCT2 of TOPBP1 as the interacting domains.

Binding affinity of the phosphorylated S168 peptide to TOPBP1 BRCT0–BRCT2 is similar to the phosphorylated RAD9-S387 association with TOPBP1, with a K_d ranging between ~0.5 μM (MDC1; this study) and 2.0 μM (RAD9; Rappas et al., 2011), which is in the same order of magnitude of typical affinities between BRCT tandem domains and their phosphorylated interaction partners (see e.g., Lloyd et al., 2009; Stucki et al., 2005; Sun et al., 2017). A direct physical and phosphorylation-dependent association between the MDC1 SDT region and TOPBP1 BRCT4+5

Figure 5. Direct Interaction with MDC1 Is Essential for TOPBP1 Recruitment to DSBs in Mitotic Cells

(A) Confocal microscopy of U2OS cells expressing GFP-tagged MDC1 WT and mutants 3 h after treatment with 3 Gy IR.

(B) Quantitative analysis of GFP-MDC1 and TOPBP1 colocalization by SQUASSH. Upper graph: object number colocalization (fraction of objects in each channel that overlap ≥50%). Lower graph: object size colocalization (area of object overlap divided by total object area). Each data point represents one cell (n = 10); bars represent the mean. One-way ANOVA and Dunnett's multiple comparison test were performed to test for difference of WT versus mutants. All mutant cell lines are significantly different from WT ($p \leq 0.0006$).

(C) Confocal microscopy of U2OS cells arrested in mitosis by nocodazole (100 ng/mL) and treated with 0.5 Gy IR.

(D) Confocal microscopy of U2OS Δ*MDC1* and Δ*53BP1* cells arrested in mitosis by nocodazole (100 ng/mL) and treated with 0.5 Gy IR.

(E) Quantitative analysis of TOPBP1-γH2AX colocalization in Δ*MDC1* and Δ*53BP1* cells by SQUASSH. Each data point represents one cell (n = 8); bars represent the mean.

(F) Confocal microscopy of U2OS cells expressing GFP-tagged MDC1 WT and mutants, arrested in mitosis by nocodazole (100 ng/mL) 1 h after treatment with 0.5 Gy IR.

(G) Quantitative analysis of GFP-MDC1 and TOPBP1 colocalization by SQUASSH. Upper graph: object number colocalization. Lower graph: object size colocalization. Each data point represents one cell (n = 10); bars represent the mean.

All scale bars represent 10 μm. See also Figure S3.

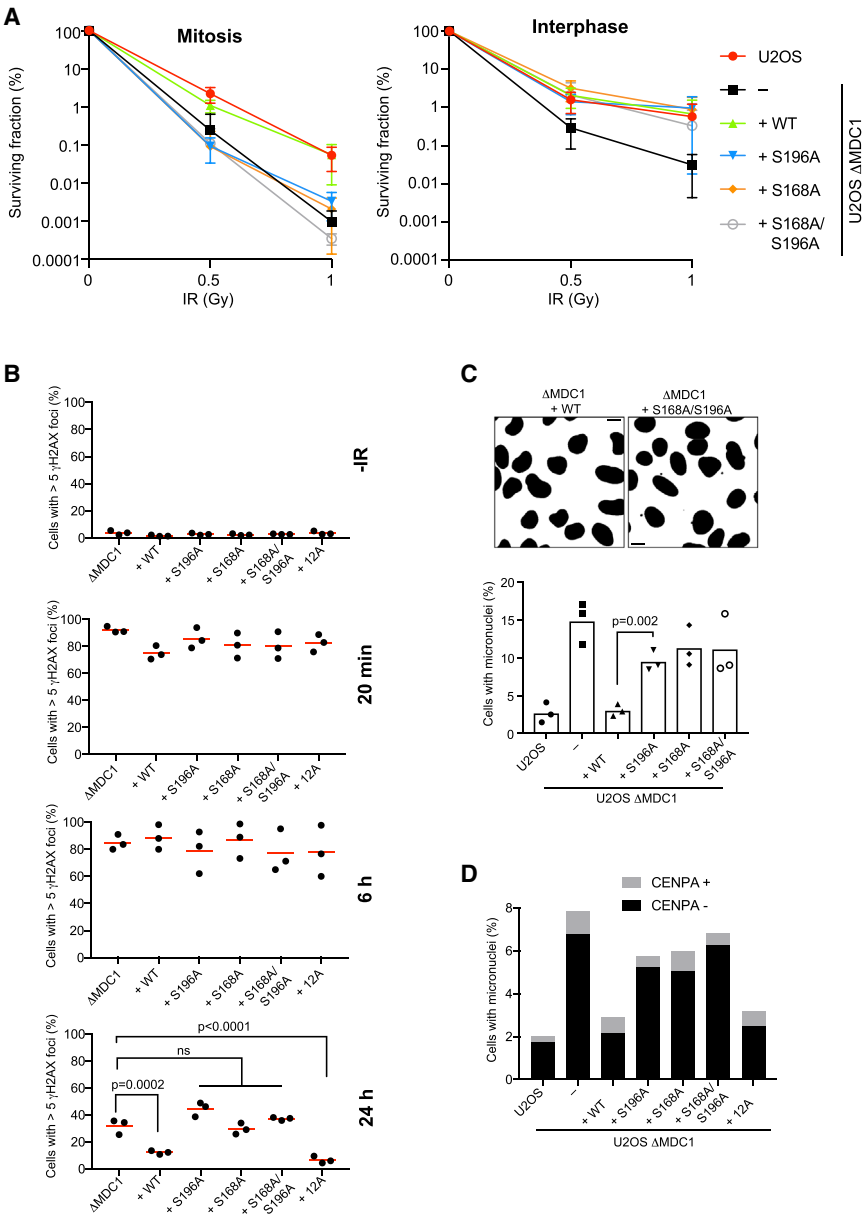


Figure 6. The MDC1-TOPBP1 Interaction Promotes Genome Stability during Mitosis

(A) Clonogenic survival assay of mitotic and interphase U2OS, Δ*MDC1*, and Δ*MDC1* cells stably transfected with MDC1 WT and TOPBP1 binding mutants (mean of 3 independent experiments; error bars: ±SD).

(B) Quantification of background γH2AX (−IR) and residual γH2AX foci 20 min, 6 h, and 24 h after irradiation of mitotic cells and subsequent release from mitotic arrest (red bars represent the mean; n = 3). One-way ANOVA and Dunnett's multiple comparison test were performed for the 24 h time point to test for difference of the mean of Δ*MDC1* versus complemented cell lines.

(C) Quantification of micronuclei formation 4 h after irradiation of mitotic cells and subsequent release from mitotic arrest (bars represent mean ± SD, n = 3, unpaired t test, $\alpha = 0.05$, at least 1,000 cells per condition).

(D) Quantification of micronuclei formation in untreated U2OS cell lines after staining of micronuclei for CENPA (stacked bars represent means of two independent experiments of at least 700 cells assessed per condition).

All scale bars represent 10 μm. See also Figure S4.

was previously reported and implicated in the replication checkpoint (Wang et al., 2011). However, in our hands and in another recent study (Choi and Yoo, 2016), mutation of TOPBP1-BRCT5 had no effect on MDC1-TOPBP1 binding, and we could not detect any measurable association between the phosphorylated MDC1-SDT region and TOPBP1 in our *in vitro* binding assays and co-immunoprecipitation experiments. Based on these findings, we propose that the NBS1 and TOPBP1 binding sites on MDC1 are structurally and functionally distinct and that the weak association between the MDC1 SDT region and TOPBP1 BRCT4+5 previously reported (K_d 94 ± 15 μM; Leung et al., 2013), is not physiologically relevant (as recently suggested by the authors of the original studies; Sun et al., 2017).

Our data demonstrate that the direct recruitment of TOPBP1 to DSBs via MDC1 is specifically important during mitosis. Interestingly, MDC1 phosphorylation on S168 and S196 did not detectably change in mitotic cells and was not induced by DNA damage (Figure S6B). These data are consistent with CK2 being constitutively active throughout the cell cycle. In line with this, the MDC1-TOPBP1 interaction also did not appear to change in response to these events (Figure S6C), and CK2 inhibition does not inhibit mitotic entry (Figure S6D). Thus, while MDC1-TOPBP1 complexes are present throughout the cell cycle, their physiologically relevant role seems to be confined to mitosis. Since MDC1 cannot substitute for 53BP1 in the accumulation of TOPBP1 in G1 foci, it is also unlikely that the mere exclusion of 53BP1 from mitotic chromatin could explain the switch in the TOPBP1 recruitment mechanism during mitosis. In this context, it is interesting to note that TOPBP1 also changes its foci pattern and recruitment mechanism at the G1/S transition. A possible mechanistic explanation for this switch in foci pattern came from recruitment studies carried out with full-length TOPBP1 and isolated TOPBP1 BRCT modules (Cescutti et al., 2010). Recruitment of TOPBP1 to foci in interphase cells consistently depends on two pairs of BRCT domains: G1 foci are dependent on BRCT0–2 and BRCT4+5 and colocalize with γH2AX, while S/G2 phase foci are dependent on BRCT0–2 and BRCT7+8 and colocalize with RPA-coated ssDNA. Here, we show that only one set of BRCT domains (BRCT0–2) is required for efficient accumulation of TOPBP1 on damaged mitotic chromosomes (Figure S3F). This suggests a switch in the recruitment

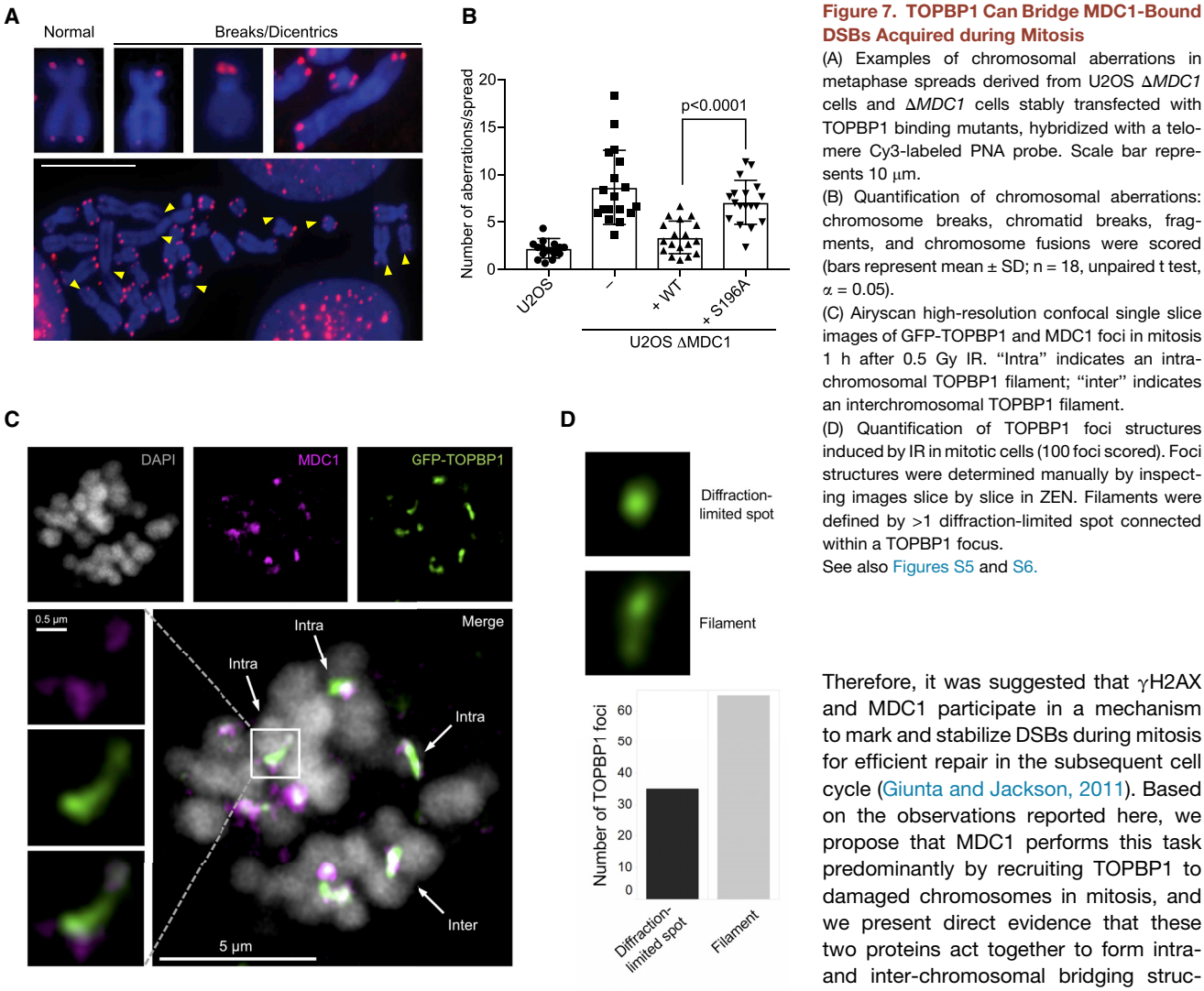


Figure 7. TOPBP1 Can Bridge MDC1-Bound DSBs Acquired during Mitosis

(A) Examples of chromosomal aberrations in metaphase spreads derived from U2OS Δ*MDC1* cells and Δ*MDC1* cells stably transfected with TOPBP1 binding mutants, hybridized with a telomere Cy3-labeled PNA probe. Scale bar represents 10 μm.

(B) Quantification of chromosomal aberrations: chromosome breaks, chromatid breaks, fragments, and chromosome fusions were scored (bars represent mean ± SD; n = 18, unpaired t test, $\alpha = 0.05$).

(C) Airyscan high-resolution confocal single slice images of GFP-TOPBP1 and MDC1 foci in mitosis 1 h after 0.5 Gy IR. "Intra" indicates an intra-chromosomal TOPBP1 filament; "inter" indicates an interchromosomal TOPBP1 filament.

(D) Quantification of TOPBP1 foci structures induced by IR in mitotic cells (100 foci scored). Foci structures were determined manually by inspecting images slice by slice in ZEN. Filaments were defined by >1 diffraction-limited spot connected within a TOPBP1 focus.

See also Figures S5 and S6.

mode of TOPBP1 in mitosis that may also be associated with a switch in its physiological function.

Thus, we propose that distinct modes of recruitment to sites of DSBs may have evolved to enable TOPBP1 to engage in diverse cellular response processes, depending on the cell-cycle stage. In G1, TOPBP1 has been implicated in G1/S checkpoint activation in response to irradiation via 53BP1 (Cescutti et al., 2010). In S/G2, TOPBP1 functions in multiple processes, including checkpoint signaling (via ATR), DNA replication initiation (via Treslin), and regulation of homologous recombination (via BLM). Mitotic cells are distinct from interphase cells in their ability to progress through M phase without activating a DNA damage checkpoint response (Rieder and Cole, 1998). Thus, it is unlikely that TOPBP1 fulfils the same roles at sites of mitotic DSBs as it does in G1 and in S/G2 phase. Furthermore, there is ample evidence that either DSBs are not repaired or that DSB repair pathways are extensively rewired in mitosis (reviewed in Heijink et al., 2013). Yet, DSBs do not go unnoticed in mitosis. ATM is activated, H2AX is efficiently phosphorylated, and MDC1 recognizes and is recruited to γH2AX-marked chromatin in mitosis. Therefore, it was suggested that γH2AX and MDC1 participate in a mechanism to mark and stabilize DSBs during mitosis for efficient repair in the subsequent cell cycle (Giunta and Jackson, 2011). Based on the observations reported here, we propose that MDC1 performs this task predominantly by recruiting TOPBP1 to damaged chromosomes in mitosis, and we present direct evidence that these two proteins act together to form intra- and inter-chromosomal bridging structures that may help to tether and stabilize broken chromosomes until they can be repaired in the following G1 phase. Consistent with this idea, we found that loss of MDC1 or disruption of the MDC1-TOPBP1 interaction leads to increased radiosensitivity specifically in mitotic cells and increased formation of micronuclei that contain mostly acentric chromatin fragments. This suggests that DSBs are not properly stabilized in the absence of MDC1 and TOPBP1 recruitment, and acentric fragments are not segregated into the daughter cells along with the chromosome from which they have broken off. Additionally, we observed that, upon disruption of the MDC1-TOPBP1 complex, a subset of DSBs remain unrepaired 24 h after DNA damage induction, suggesting that these persistent lesions would have benefited from marking and/or end-tethering during mitosis. A similar role was recently proposed for the MRN complex based on the observation that its NBS1 subunit forms small foci on mitotic chromosomes in response to IR (Giunta et al., 2010). We reproduced these data and found that, similar to the situation in interphase cells, the mitotic NBS1 foci are also dependent on the SDT repeat region in MDC1 (Figure S7A). However, mutation of the SDT repeats did

not render cells hypersensitive to IR in mitosis, nor did it negatively affect the stabilization of mitotic DSBs (Figures 6B, 6D, and S7B). Therefore, we conclude that the MDC1-MRN interaction is not critically implicated in the maintenance of genome stability during mitosis.

More than three decades ago, an IR-dependent, female-specific mutator (*mu2a*) on chromosome 3 in *Drosophila* was described that specifically increases the recovery of terminal deficiencies (i.e., chromosomes that have lost a natural telomere; Mason et al., 1984). *MU2* was later shown to be the *Drosophila* ortholog of vertebrate *MDC1* (Dronamraju and Mason, 2009). The prevalence of female-specific terminal deficiencies would be consistent with post-irradiation deposition of single-ended broken chromosomes into the zygote after the meiotic cell divisions, because acentric fragments would not move to the pole at anaphase in the absence of MU2 (MDC1) and would be lost. Based on this and on our own findings reported here, we propose that tethering of broken chromosomes during mitosis and meiosis may be an ancient and highly conserved function of MDC1. In this context, it is interesting to note that Mdb1, the *S. pombe* MDC1 ortholog, was recently shown to interact with the TOPBP1 ortholog Rad4 (Day et al., 2018). Whether the Mdb1-Rad4 complex performs functions similar to the MDC1-TOPBP1 complex in mammals remains to be determined.

Finally, it has been reported that TOPBP1 is capable of oligomerizing to form higher-order protein assemblies, although previous work has implicated multiple TOPBP1 motifs in this process (Bang et al., 2013; Liu et al., 2006). Nonetheless, an attractive hypothesis is that the TOPBP1 filamentous assemblies we observe bridging MDC1 foci represent large homo-oligomers containing many molecules of TOPBP1. Further work will be needed in the future to determine the mechanistic basis of higher-order TOPBP1 complex formation, what other proteins are present in such assemblies (if any), and what role they play in the response to mitotic DNA damage.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines and cell culture conditions
- METHOD DETAILS
  - Cloning and Mutagenesis
  - Peptide pull-downs
  - Mass spectrometry
  - SDS-PAGE and western blotting
  - Immunoprecipitations
  - Fluorescence polarization
  - *In vitro* kinase assays and GST-pull-downs
  - Generation of CRISPR/Cas9 MDC1 knock-out cell lines
  - Generation of stably transfected *ΔMDC1* cell lines
  - Immunofluorescence
  - Widefield microscopy
  - Confocal microscopy
  - Quantification of γH2AX foci
  - Quantification of micronuclei
  - Clonogenic survival assay
  - Metaphase analysis
  - Airyscan high-resolution confocal microscopy
  - Flow cytometry analysis of mitotic cells
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online at https://doi.org/10.1016/j.molcel.2019.02.014.

ACKNOWLEDGMENTS

We thank Daniel Durocher, Thanos Halazonetis, Steve Jackson, Jiri Lukas, Wojciech Niedzwiedz, and Feng Zhang for providing reagents. We are grateful to Steve Jackson and Madalena Tarsounas for critical reading of the manuscript and to Fena Ochs for help with Airyscan imaging. Imaging was performed with equipment maintained by the Center for Microscopy and Image Analysis, University of Zurich or the Wolfson Imaging Centre, Medical Research Council (MRC) Weatherall Institute of Molecular Medicine (WIMM), University of Oxford (supported by the MRC WIMM Strategic Alliance (G0902418), the MRC Molecular Haematology Unit (MC_UU_12009), the MRC Human Immunology Unit (MC_UU_12010), and the Wolfson Foundation (grant 18272). Cell sorting was carried out by the Flow Cytometry Core Facilities at the University of Zurich or MRC WIMM. The Blackford lab is supported by a Cancer Research UK Career Development Fellowship (C29215/A20772) to A.N.B. and an MRC WIMM DPhil Studentship (MR/N013468/1) to A.-M.K.S. and A.N.B. Work in the Pearl and Oliver lab is supported by Cancer Research UK Programme Grant C302/A24386. The Stucki lab is supported by the Forschungskredit Candoc program of the University of Zurich (FK-18-024) to P.-A.L., by a project grant from the Swiss National Foundation (31003A_163141), and by the Kanton of Zurich.

AUTHOR CONTRIBUTIONS

The project was conceived and supervised by A.N.B. and M.S. Biochemical experimental work was carried out by P.-A.L., S.E.J., and A.-M.K.S. Cell biological experimental work was carried out by P.-A.L., A.-M.K.S., and M.d.M.Z. Experimental tools and reagents were generated and characterized by P.-A.L., S.E.J., J.B., M.d.M.Z., and D.B. D.F. supervised D.B. Mass spectrometry data were contributed by S.B., R.F., and B.M.K. Fluorescence polarization data and the pSer168 phospho-specific antibody were contributed by M.D., A.W.O., and L.H.P. The paper was written by A.N.B. and M.S. with contributions from other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 27, 2018
Revised: December 30, 2018
Accepted: February 11, 2019
Published: February 18, 2019

REFERENCES

Bang, S.W., Kim, G.S., and Hwang, D.S. (2013). Oligomerization of TopBP1 is necessary for the localization of TopBP1 to mitotic centrosomes. Biochem. Biophys. Res. Commun. *436*, 31–34.

Beausoleil, S.A., Jedrychowski, M., Schwartz, D., Elias, J.E., Villén, J., Li, J., Cohn, M.A., Cantley, L.C., and Gygi, S.P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc. Natl. Acad. Sci. USA *101*, 12130–12135.

Bizard, A.H., and Hickson, I.D. (2018). Anaphase: a fortune-teller of genomic instability. Curr. Opin. Cell Biol. *52*, 112–119.

Blackford, A.N., and Jackson, S.P. (2017). ATM, ATR, and DNA-PK: the trinity at the heart of the DNA damage response. Mol. Cell *66*, 801–817.

Blackford, A.N., Nieminuszczy, J., Schwab, R.A., Galanty, Y., Jackson, S.P., and Niedzwiedz, W. (2015). TopBP1 interacts with BLM to maintain genome stability but is dispensable for preventing BLM degradation. Mol. Cell *57*, 1133–1141.

Broderick, R., Nieminuszczy, J., Blackford, A.N., Winczura, A., and Niedzwiedz, W. (2015). TOPBP1 recruits TOP2A to ultra-fine anaphase bridges to aid in their resolution. Nat. Commun. *6*, 6572.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. *7*, R100.

Cescutti, R., Negrini, S., Kohzaki, M., and Halazonetis, T.D. (2010). TopBP1 functions with 53BP1 in the G1 DNA damage checkpoint. EMBO J. *29*, 3723–3732.

Chapman, J.R., and Jackson, S.P. (2008). Phospho-dependent interactions between NBS1 and MDC1 mediate chromatin retention of the MRN complex at sites of DNA damage. EMBO Rep. *9*, 795–801.

Chiang, T.-W.W., le Sage, C., Larrieu, D., Demir, M., and Jackson, S.P. (2016). CRISPR-Cas9(D10A) nickase-based genotypic and phenotypic screening to enhance genome editing. Sci. Rep. *6*, 24356.

Choi, S.H., and Yoo, H.Y. (2016). Mdc1 modulates the interaction between TopBP1 and the MRN complex during DNA damage checkpoint responses. Biochem. Biophys. Res. Commun. *479*, 5–11.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Davis, S., Charles, P.D., He, L., Mowlds, P., Kessler, B.M., and Fischer, R. (2017). Expanding proteome coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) combined with broad specificity proteolysis. J. Proteome Res. *16*, 1288–1299.

Day, M., Rappas, M., Ptasinska, K., Boos, D., Oliver, A.W., and Pearl, L.H. (2018). BRCT domains of the DNA damage checkpoint proteins TOPBP1/Rad4 display distinct specificities for phosphopeptide ligands. eLife *7*, 94.

Delacroix, S., Wagner, J.M., Kobayashi, M., Yamamoto, K., and Karnitz, L.M. (2007). The Rad9-Hus1-Rad1 (9-1-1) clamp activates checkpoint signaling via TopBP1. Genes Dev. *21*, 1472–1477.

Dronamraju, R., and Mason, J.M. (2009). Recognition of double strand breaks by a mutator protein (MU2) in Drosophila melanogaster. PLoS Genet. *5*, e1000473.

Fischer, R., and Kessler, B.M. (2015). Gel-aided sample preparation (GASP)–a simplified method for gel-assisted proteomic sample generation from protein extracts and intact cells. Proteomics *15*, 1224–1229.

Giunta, S., and Jackson, S.P. (2011). Give me a break, but not in mitosis: the mitotic DNA damage response marks DNA double-strand breaks with early signaling events. Cell Cycle *10*, 1215–1221.

Giunta, S., Belotserkovskaya, R., and Jackson, S.P. (2010). DNA damage signaling in response to double-strand breaks during mitosis. J. Cell Biol. *190*, 197–207.

Gong, Z., Kim, J.-E., Leung, C.C.Y., Glover, J.N.M., and Chen, J. (2010). BACH1/FANCJ acts with TopBP1 and participates early in DNA replication checkpoint control. Mol. Cell *37*, 438–446.

Hari, F.J., Spycher, C., Jungmichel, S., Pavic, L., and Stucki, M. (2010). A divalent FHA/BRCT-binding mechanism couples the MRE11-RAD50-NBS1 complex to damaged chromatin. EMBO Rep. *11*, 387–392.

Heijink, A.M., Krajewska, M., and van Vugt, M.A.T.M. (2013). The DNA damage response during mitosis. Mutat. Res. *750*, 45–55.

Huen, M.S.Y., Grant, R., Manke, I., Minn, K., Yu, X., Yaffe, M.B., and Chen, J. (2007). RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. Cell *131*, 901–914.

Hustedt, N., and Durocher, D. (2016). The control of DNA repair by the cell cycle. Nat. Cell Biol. *19*, 1–9.

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071–1078.

Kolas, N.K., Chapman, J.R., Nakada, S., Ylanko, J., Chahwan, R., Sweeney, F.D., Panier, S., Mendez, M., Wildenhain, J., Thomson, T.M., et al. (2007). Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. Science *318*, 1637–1640.

Kumagai, A., Shevchenko, A., Shevchenko, A., and Dunphy, W.G. (2010). Treslin collaborates with TopBP1 in triggering the initiation of DNA replication. Cell *140*, 349–359.

Lee, J., Kumagai, A., and Dunphy, W.G. (2007). The Rad9-Hus1-Rad1 checkpoint clamp regulates interaction of TopBP1 with ATR. J. Biol. Chem. *282*, 28036–28044.

Lee, D.-H., Acharya, S.S., Kwon, M., Drane, P., Guan, Y., Adelmant, G., Kalev, P., Shah, J., Pellman, D., Marto, J.A., and Chowdhury, D. (2014). Dephosphorylation enables the recruitment of 53BP1 to double-strand DNA breaks. Mol. Cell *54*, 512–525.

Leung, C.C.Y., and Glover, J.N.M. (2011). BRCT domains: easy as one, two, three. Cell Cycle *10*, 2461–2470.

Leung, C.C.Y., Sun, L., Gong, Z., Burkat, M., Edwards, R., Assmus, M., Chen, J., and Glover, J.N.M. (2013). Structural insights into recognition of MDC1 by TopBP1 in DNA replication checkpoint control. Structure *21*, 1450–1459.

Liu, K., Paik, J.C., Wang, B., Lin, F.-T., and Lin, W.-C. (2006). Regulation of TopBP1 oligomerization by Akt/PKB for cell survival. EMBO J. *25*, 4795–4807.

Lloyd, J., Chapman, J.R., Clapperton, J.A., Haire, L.F., Hartsuiker, E., Li, J., Carr, A.M., Jackson, S.P., and Smerdon, S.J. (2009). A supramodular FHA/BRCT-repeat architecture mediates Nbs1 adaptor function in response to DNA damage. Cell *139*, 100–111.

Lou, Z., Minter-Dykhouse, K., Franco, S., Gostissa, M., Rivera, M.A., Celeste, A., Manis, J.P., van Deursen, J., Nussenzweig, A., Paull, T.T., et al. (2006). MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. Mol. Cell *21*, 187–200.

Lukas, C., Savic, V., Bekker-Jensen, S., Doil, C., Neumann, B., Pedersen, R.S., Grøfte, M., Chan, K.L., Hickson, I.D., Bartek, J., and Lukas, J. (2011). 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. Nat. Cell Biol. *13*, 243–253.

Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C., and Lukas, J. (2007). RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. Cell *131*, 887–900.

Mason, J.M., Strobel, E., and Green, M.M. (1984). mu-2: mutator gene in Drosophila that potentiates the induction of terminal deficiencies. Proc. Natl. Acad. Sci. USA *81*, 6090–6094.

Meggio, F., and Pinna, L.A. (2003). One-thousand-and-one substrates of protein kinase CK2? FASEB J. *17*, 349–368.

Melander, F., Bekker-Jensen, S., Falck, J., Bartek, J., Mailand, N., and Lukas, J. (2008). Phosphorylation of SDT repeats in the MDC1 N terminus triggers retention of NBS1 at the DNA damage-modified chromatin. J. Cell Biol. *181*, 213–226.

Nakamura, A.J., Rao, V.A., Pommier, Y., and Bonner, W.M. (2010). The complexity of phosphorylated H2AX foci formation and DNA repair assembly at DNA double-strand breaks. Cell Cycle *9*, 389–397.

Nelson, G., Buhmann, M., and von Zglinicki, T. (2009). DNA damage foci in mitosis are devoid of 53BP1. Cell Cycle *8*, 3379–3383.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. *302*, 205–217.

Orthwein, A., Fradet-Turcotte, A., Noordermeer, S.M., Canny, M.D., Brun, C.M., Strecker, J., Escribano-Diaz, C., and Durocher, D. (2014). Mitosis inhibits

DNA double-strand break repair to guard against telomere fusions. Science *344*, 189–193.

Orthwein, A., Noordermeer, S.M., Wilson, M.D., Landry, S., Enchev, R.I., Sherker, A., Munro, M., Pinder, J., Salsman, J., Dellaire, G., et al. (2015). A mechanism for the suppression of homologous recombination in G1 cells. Nature *528*, 422–426.

Pedersen, R.T., Kruse, T., Nilsson, J., Oestergaard, V.H., and Lisby, M. (2015). TopBP1 is required at mitosis to reduce transmission of DNA damage to G1 daughter cells. J. Cell Biol. *210*, 565–582.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Rappas, M., Oliver, A.W., and Pearl, L.H. (2011). Structure and function of the Rad9-binding region of the DNA-damage checkpoint adaptor TopBP1. Nucleic Acids Res. *39*, 313–324.

Rieder, C.L., and Cole, R.W. (1998). Entry into mitosis in vertebrate somatic cells is guarded by a chromosome damage checkpoint that reverses the cell cycle when triggered during early but not late prophase. J. Cell Biol. *142*, 1013–1022.

Rizk, A., Paul, G., Incardona, P., Bugarski, M., Mansouri, M., Niemann, A., Ziegler, U., Berger, P., and Sbalzarini, I.F. (2014). Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squassh. Nat. Protoc. *9*, 586–596.

Rogakou, E.P., Boon, C., Redon, C., and Bonner, W.M. (1999). Megabase chromatin domains involved in DNA double-strand breaks in vivo. J. Cell Biol. *146*, 905–916.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Siddiqui-Jain, A., Drygin, D., Streiner, N., Chua, P., Pierre, F., O'Brien, S.E., Bliesath, J., Omori, M., Huser, N., Ho, C., et al. (2010). CX-4945, an orally bioavailable selective inhibitor of protein kinase CK2, inhibits prosurvival and angiogenic signaling and exhibits antitumor efficacy. Cancer Res. *70*, 10288–10298.

Spycher, C., Miller, E.S., Townsend, K., Pavic, L., Morrice, N.A., Janscak, P., Stewart, G.S., and Stucki, M. (2008). Constitutive phosphorylation of MDC1 physically links the MRE11-RAD50-NBS1 complex to damaged chromatin. J. Cell Biol. *181*, 227–240.

Stewart, G.S., Wang, B., Bignell, C.R., Taylor, A.M.R., and Elledge, S.J. (2003). MDC1 is a mediator of the mammalian DNA damage checkpoint. Nature *421*, 961–966.

Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell *123*, 1213–1226.

Sun, L., Huang, Y., Edwards, R.A., Yang, S., Blackford, A.N., Niedzwiedz, W., and Glover, J.N.M. (2017). Structural Insight into BLM Recognition by TopBP1. Structure *25*, 1582–1588.e1583.

van Vugt, M.A.T.M., Gardino, A.K., Linding, R., Ostheimer, G.J., Reinhardt, H.C., Ong, S.-E., Tan, C.S., Miao, H., Keezer, S.M., Li, J., et al. (2010). A mitotic phosphorylation feedback network connects Cdk1, Plk1, 53BP1, and Chk2 to inactivate the G(2)/M DNA damage checkpoint. PLoS Biol. *8*, e1000287.

Vizcaíno, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., et al. (2016). 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. *44* (D1), D447–D456.

Wang, J., Gong, Z., and Chen, J. (2011). MDC1 collaborates with TopBP1 in DNA replication checkpoint control. J. Cell Biol. *193*, 267–273.

Wu, L., Luo, K., Lou, Z., and Chen, J. (2008). MDC1 regulates intra-S-phase checkpoint by targeting NBS1 to DNA double-strand breaks. Proc. Natl. Acad. Sci. USA *105*, 11200–11205.

Yi, T., Zhai, B., Yu, Y., Kiyotsugu, Y., Raschle, T., Etzkorn, M., Seo, H.-C., Nagiec, M., Luna, R.E., Reinherz, E.L., et al. (2014). Quantitative phosphoproteomic analysis reveals system-wide signaling pathways downstream of SDF-1/CXCR4 in breast cancer stem cells. Proc. Natl. Acad. Sci. USA *111*, E2182–E2190.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-53BP1 (IF)	Santa Cruz	sc-22760; RRID: AB_2256326
Rabbit polyclonal anti-AKT (WB)	Cell Signaling Technology	9272; RRID: AB_329827
Rabbit monoclonal anti-AKT-pS129 (WB)	Abcam	ab133458; RRID: AB_10895993
Rabbit monoclonal anti-ATM-pS1981 (WB)	Epitomics	2152-1; RRID: AB_991678
Rabbit polyclonal anti-Biotin (WB)	Abcam	ab53494; RRID: AB_867860
Goat polyclonal anti-BLM (WB)	Abcam	ab5446; RRID: AB_304894
Mouse monoclonal anti-Cyclin A (IF)	BD Biosciences	611269; RRID: AB_398797
Mouse monoclonal anti-CENPA (IF)	Abcam	ab13939; RRID: AB_300766
Mouse monoclonal anti-CHK1 (WB)	Santa Cruz	sc-8408; RRID: AB_627257
Rabbit polyclonal anti-CHK1-pS317 (WB)	Cell Signaling Technology	2344; RRID: AB_331488
Rabbit monoclonal anti-CHK1-pS345 (WB)	Cell Signaling Technology	2348; RRID: AB_331212
Rabbit polyclonal anti-CHK2 (WB)	Cell Signaling Technology	2662; RRID: AB_2080793
Rabbit polyclonal anti-CHK2-pT68 (WB)	Cell Signaling Technology	2661; RRID: AB_331479
Mouse monoclonal anti-CENPA (IF)	Abcam	ab13939; RRID: AB_300766
Mouse polyclonal anti-GFP (WB)	Roche	11814460001; RRID: AB_390913
Mouse monoclonal anti-GST (WB)	Santa Cruz	sc-138; RRID: AB_627677
Rabbit polyclonal anti-H2A (WB)	Merck	07-146; RRID: AB_11212920
Rabbit polyclonal anti-H2AX (WB)	Novus Biologicals	NB100-383; RRID: AB_10002060
Mouse monoclonal anti-γH2AX (IF/WB)	Merck	05-636; RRID: AB_309864
Rabbit polyclonal anti-H3 (WB)	Abcam	ab1791; RRID: AB_302613
Mouse monoclonal anti-H3-pS10 (FACS/WB)	Abcam	ab14955; RRID: AB_443110
Mouse monoclonal anti-HA (WB)	Santa Cruz	sc-57592; RRID: AB_629568
Rabbit polyclonal anti-KAP1 (WB)	Abcam	ab10483; RRID: AB_297222
Rabbit polyclonal anti-KAP1-pS824 (WB)	Bethyl Laboratories	IHC-00073; RRID: AB_577234
Mouse monoclonal anti-MDC1 (IF)	Abcam	ab50003; RRID: AB_881103
Rabbit polyclonal anti-MDC1 (WB)	Abcam	ab11171; RRID: AB_297810
Rabbit polyclonal anti-MDC1-pS168 (WB)	AMS Biotechnology	N/A
Rabbit polyclonal anti-MDC1-pS196 (WB)	21st Century Biochemicals	N/A
Rabbit polyclonal anti-NBS1 (IF)	Novus Biologicals	NB100-143; RRID: AB_10078050
Rabbit monoclonal anti-NBS1 (WB)	Abcam	ab32074; RRID: AB_777007
Mouse monoclonal anti-PLK1 (WB)	Merck	05-844; RRID AB_310836
Rabbit polyclonal anti-RAD9 (WB)	Abcam	ab70810; RRID: AB_1270159
Mouse monoclonal anti-RPA2 (WB)	Abcam	ab2175; RRID: AB_302873
Rabbit monoclonal anti-RPA2-pS4/pS8 (WB)	Bethyl Laboratories	A700-009; RRID: AB_2765278
Rabbit polyclonal anti-SMC1 (WB)	Abcam	ab9262; RRID: AB_307121
Rabbit polyclonal anti-SMC1-pS966 (WB)	Abcam	ab1276; RRID: AB_299409
Rabbit polyclonal anti-TOPBP1 (IF)	Bethyl Laboratories	A300-111A; RRID: AB_2272050
Rabbit polyclonal anti-TOPBP1 (WB)	Abcam	ab2402; RRID: AB_303044
Alexa Fluor 488 goat anti-rabbit IgG H&L (IF)	Abcam	ab150077; RRID: AB_2630356
Alexa Fluor 568 goat anti-rabbit IgG H&L (IF)	Abcam	ab175471; RRID: AB_2576207
Alexa Fluor 647 goat anti-mouse IgG H&L (IF)	Abcam	ab150115; RRID: AB_2687948
Alexa Fluor 647 goat anti-mouse IgG1 (FACS)	Thermo Fisher Scientific	A-21240; RRID: AB_2535809

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
HeLa nuclear extract	Ipracell	CC-01-20-25
Chemicals, Peptides, and Recombinant Proteins		
Alexa Fluor 488-labeled PNA CENPB probe	PNA Bio	F3004
Aphidicolin	Merck	178273
ATM inhibitor AZD0156	Selleck Chemicals	S8375
ATR inhibitor AZD6738	Cayman Chemical	21035
Camptothecin	Cayman Chemical	11694
CDK inhibitor RO-3306	Merck	217699
CK2 inhibitor CX-4945	Selleck Chemicals	S2248
cOmplete EDTA-free Protease Inhibitor Cocktail	Merck	5056489001
DAPI solution	BD Biosciences	564907
DNA-PK inhibitor NU7741	Cayman Chemical	14881
EZview Red Anti-HA Affinity Gel	Merck	E6779
GFP-Trap_MA	Chromotek	gtma-100
Glutathione Sepharose 4B	GE Healthcare	17075601
Karyomax Colcemid Solution In PBS	Thermo Fisher Scientific	15212012
Lipofectamine 2000	Thermo Fisher Scientific	11668019
MDC1-S196 peptide (Biotin-GGSGSRTTSSSVIVPESDEEGHSPV-NH_2)	Genosphere Biotechnologies	N/A
MDC1-pS196 peptide (Biotin-GGSGSRTTSSSVIVPE[pS] DEEGHSPV-NH_2)	Genosphere Biotechnologies	N/A
MDC1-pS168 peptide (Flu-GYGRLLLAED[pS]EEEVDFL)	Peptide Protein Research	N/A
MDC1-pS196 peptide (Flu-GYGGSVIVPE[pS]DEEGHSP)	Peptide Protein Research	N/A
MDC1 SDT motif 1 peptide (Flu-GYGFAFNLN[pS]D[pT]DVEEGQ)	Peptide Protein Research	N/A
MDC1 SDT motif 2 peptide (Flu-GYGQPPGED[pS]D[pT]DVDDDS)	Peptide Protein Research	N/A
MDC1 SDT motif 3 peptide (Flu-GYGPFGFID[pS]D[pT]DAEEER)	Peptide Protein Research	N/A
MDC1 SDT motif 4 peptide (Flu-GYGQESQAG[pS]D[pT]DVEEGK)	Peptide Protein Research	N/A
MDC1 SDT motif 5 peptide (Flu-GYGASMVIN[pS]D[pT]DDEEEV)	Peptide Protein Research	N/A
MDC1 SDT motif 6 peptide (Flu-GYGTTTERD[pS]D[pT]DVEEEE)	Peptide Protein Research	N/A
Nocodazole	Merck	M1404
PLK1 inhibitor BI 2536	Cayman Chemical	17385
SuperSignal West Femto Maximum Sensitivity Substrate	Thermo Fisher Scientific	34095
VECTASHIELD Antifade Mounting Media with DAPI	Vector Laboratories	H-1200
Critical Commercial Assays		
Bolt 4%–12% Bis-Tris Plus Gels	Thermo Fisher Scientific	NW04120BOX
Fixation/Permeabilization Solution Kit	BD Biosciences	554714
GenElute HP Endotoxin-Free Plasmid Maxiprep Kit	Merck	NA0410
Mini-PROTEAN 4%–15% TGX Stain-Free Protein Gels	Bio-Rad	4568086
Novex 10%–20% Tricine Protein Gels	Thermo Fisher Scientific	EC6625BOX
QuickLyse Miniprep Kit	QIAGEN	27405
QuikChange II Site-Directed Mutagenesis Kit	Agilent Technologies	200524
Telomere PNA FISH Kit/Cy3	Agilent Technologies	K5326
Deposited Data		
Mass spectrometry	This paper and PRIDE	PXD011850
Raw image files	This study, Mendeley Data	https://data.mendeley.com/datasets/26pch2jvzh.1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
293FT human embryonic kidney cell line	Thermo Fisher Scientific	R70007
U-2 OS (U2OS) human bone osteosarcoma cell line	ATCC	HTB-96
U2OS Δ53BP1	Daniel Durocher lab	Orthwein et al., 2015
U2OS ΔMDC1	This paper	N/A
U2OS ΔMDC1 +GFP-MDC1	This paper	N/A
U2OS ΔMDC1 +GFP-MDC1-S168A	This paper	N/A
U2OS ΔMDC1 +GFP-MDC1-S196A	This paper	N/A
U2OS ΔMDC1 +GFP-MDC1-S168A/S196A	This paper	N/A
U2OS ΔMDC1 +GFP-MDC1-AQXF	This paper	N/A
U2OS ΔMDC1 +GFP-MDC1-12A	This paper	N/A
RPE-1 human retinal pigmented epithelium cell line	Stephen Jackson lab	Chiang et al., 2016
RPE-1 H2AX$^{S139A/S139A}$	Stephen Jackson lab	Chiang et al., 2016
RPE-1 ΔMDC1	Stephen Jackson lab	Chiang et al., 2016
Oligonucleotides		
Mutagenesis primer MDC1-S168A forward: CTGTTGGCTGAGGACGCGGAGGAGGAAGTAG	Microsynth	N/A
Mutagenesis primer MDC1-S168A reverse: CTACTTCCTCCTCCGCGTCCTCAGCCAACAG	Microsynth	N/A
Mutagenesis primer MDC1-S196A forward: CTCTGTGATAGTTCCAGAGGCTGATGAAGAGGGGCATTCC	Microsynth	N/A
Mutagenesis primer MDC1-S196A reverse: GGAATGCCCCTCTTCATCAGCCTCTGGAACTATCACAGAG	Microsynth	N/A
Mutagenesis primer TOPBP1-BRCT2 (K250A) forward: GCAAGAACCAAAAGGTCAGGCGTATGAGTGTGCCAAGAGA	Eurofins Genomics	N/A
Mutagenesis primer TOPBP1-BRCT2 (K250A) reverse: TCTCTTGGCACACTCATACGCCTGACCTTTTGGTTCTTGC	Eurofins Genomics	N/A
sgRNA primer forward: CACCCACCTCGGGAAGAATGTGGT	Microsynth	N/A
sgRNA primer reverse: AAACACCACATTCTTCCCGAGGTG	Microsynth	N/A
Recombinant DNA		
pcDNA3-MDC1-HA-12A	Jiri Lukas lab	Melander et al., 2008
pcDNA3.1-GFP-MDC1-WT	Stephen Jackson lab	Kolas et al., 2007
pcDNA3.1-GFP-MDC1-12A	This paper	N/A
pcDNA3.1-GFP-MDC1-AQXF	Stephen Jackson lab	Kolas et al., 2007
pcDNA3.1-GFP-MDC1-S168A	This paper	N/A
pcDNA3.1-GFP-MDC1-S196A	This paper	N/A
pcDNA3.1-GFP-MDC1-S168A/S196A	This paper	N/A
pcDNA3.1-HA-MDC1-WT	Stephen Jackson lab	Kolas et al., 2007
pcDNA5/FRT/TO	Thermo Fisher Scientific	V6520-20
pcDNA5/FRT/TO-MDC1-HA-WT	This paper	N/A
pcDNA5/FRT/TO-MDC1-HA-12A	This paper	N/A
pcDNA5/FRT/TO-MDC1-HA-S196A-12A	This paper	N/A
pcDNA5/FRT/TO-MDC1-HA-S196A	This paper	N/A
pcDNA5/FRT/TO-MDC1-HA-S168A	This paper	N/A
pcDNA5/FRT/TO-MDC1-HA-S168/S196A	This paper	N/A
pGEX-4T-3-MDC1$^{109-330}$	Manuel Stucki lab	Spycher et al., 2008
pGEX-4T-3-MDC1$^{109-330}$ S196A	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pGEX-4T-3-MDC1$^{109-330}$ ADA	This paper	N/A
pGEX-4T-3-MDC1$^{109-330}$ ADA+S196A	This paper	N/A
pIRESneo2-TOPBP1	Thanos Halazonetis lab	Cescutti et al., 2010
pIRESneo2-TOPBP1-K154A/K155A (BRCT1)	Wojciech Niedzwiedz lab	Blackford et al., 2015
pIRESneo2-TOPBP1-K250A (BRCT2)	This paper	N/A
pIRESneo2-TOPBP1-K704A (BRCT5)	Wojciech Niedzwiedz lab	Blackford et al., 2015
pX330-U6-Chimeric_BB-CBh-hSpCas9	Feng Zhang lab	Addgene plasmid #42230; Cong et al., 2013
Software and Algorithms		
Boxshade	ExPASy Bioinformatics Resource Portal	https://embnet.vital-it.ch/software/BOX_form.html
CellProfiler	Carpenter et al., 2006	http://cellprofiler.org
Fiji	Schindelin et al., 2012	https://imagej.net/Fiji
GraphPad Prism v7.00	GraphPad Prism version 7.00 GraphPad Software, La Jolla California USA	http://graphpad.com
R	R Development Core Team	http://r-project.org
SQUASSH segmentation algorithm	Rizk et al., 2014	https://imagej.net/Squassh
T-Coffee multiple sequence alignment tool	Notredame et al., 2000	http://tcoffee.crg.cat
ZEN Black v2.1	ZEISS	https://www.zeiss.com

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents should be directed to the Lead Contact, Manuel Stucki (manuel.stucki@uzh.ch).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines and cell culture conditions

Source of all the cell lines used in this study is provided in the key resource table. 293FT, U2OS and RPE-1 cells were cultured in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal calf serum (FCS), 2 mM L-glutamine and penicillin-streptomycin antibiotics under standard cell culture conditions in a CO_2 incubator (37°C; 5% CO_2). Stably transfected U2OS cell lines were cultured in the presence of 400 μg/mL Geneticin. Cells were regularly tested for mycoplasma contamination.

METHOD DETAILS

Cloning and Mutagenesis

pcDNA5/FRT/TO-MDC1-HA-WT was generated by subcloning MDC1-HA-WT from pcDNA3.1-MDC1-HA-WT (gift from Steve Jackson; (Kolas et al., 2007)) into a modified pcDNA5/FRT/TO vector (Invitrogen). pcDNA5/FRT/TO-MDC1-HA-12A was generated by excision of a SmaI/AvrII restriction fragment containing the SDT region from pcDNA3-MDC1-HA-12A (gift from Jiri Lukas; (Melander et al., 2008)) and ligating it into pcDNA5/FRT/TO-MDC1-HA-WT that was previously digested with SmaI and AvrII. pcDNA5/FRT/TO-MDC1-HA-S168A, -S196A and -S168/S196A and pcDNA5/FRT/TO-MDC1-HA-S196A-12A were generated by site directed mutagenesis of pcDNA5/FRT/TO- MDC1-HA-WT and pcDNA5/FRT/TO-MDC1-HA-12A, respectively, using Quik-Change II Site-Directed Mutagenesis kit (Agilent Technologies). Sequences of the mutagenesis primers are in the key resource table. pGEX4T3-MDC1$^{109-330}$ fragment ("M2") was described elsewhere (Spycher et al., 2008). pGEX-4T-3-MDC1-M2-S196A, pGEX-4T-3-MDC1-M2-ADA-S196A and pGEX-4T-3-MDC1-M2-ADA were generated by PCR amplification of the M2 fragment using the pcDNA5/FRT/TO-MDC1-HA-12A and pcDNA5/FRT/TO-MDC1-HA-S196A-12A plasmids as templates, respectively. Fragments were then ligated into the pGEX-4T-3 vector. pcDNA3.1-GFP-MDC1-WT and pcDNA3.1-GFP-MDC1-AQXF plasmids were a gift from Steve Jackson and are described elsewhere (Kolas et al., 2007). S168A, S196A and S168A/S196A mutations were introduced in pcDNA3.1-GFP-MDC1-WT as described above for pcDNA5/FRT/TO-MDC1-HA plasmids. pcDNA3.1-GFP-MDC1-12A was generated by replacing an ApaI restriction fragment within the MDC1 coding sequence that contains all the six SDT repeats with a custom synthesized DNA fragment (BioCat GmbH), designed to have all six SDT motifs changed to ADA.

Peptide pull-downs

Biotinylated peptides (Genosphere Biotechnologies) were bound to streptavidin-coupled Dynabeads M-280 (Life Technologies). HeLa nuclear extracts (Ipracell) were diluted 1:1 with 2X dilution buffer (100 mM NaCl, 10 mM NaF, 0.2 mM EDTA, 0.4% Igepal CA-630, 20 mM HEPES-KOH, pH 7.4) supplemented with cOmplete protease inhibitor cocktail (Roche), and cleared by centrifugation. Dynabead-conjugated peptides were incubated with clarified extracts with end-to-end mixing at 4°C for 2 hr. Beads were washed with peptide pull-down buffer (50 mM NaCl, 50 mM KCl, 5 mM NaF, 0.1 mM EDTA, 0.2% Igepal CA-630, 10% glycerol, 20 mM HEPES-KOH, pH 7.4) supplemented with cOmplete protease inhibitor cocktail (Roche), before elution in 2X SDS sample buffer for mass spectrometry or SDS-PAGE. Peptide sequences were as follows: Biotin-GGSGSRTTSSSVIVPESDEEGHSPV-NH_2 (MDC1-S196) and Biotin-GGSGSRTTSSSVIVPE[pS]DEEGHSPV-NH_2 (MDC1-pS196).

Mass spectrometry

Precipitated material in SDS sample buffer was subjected to two rounds of chloroform-methanol precipitation followed by in-solution trypsin digestion as described previously (Fischer and Kessler, 2015). Digested and desalted samples were resuspended in 20 μL 2% acetonitrile and 0.1% formic acid and analyzed by nano-flow liquid chromatography tandem mass spectrometry (LC-MS/MS) as described previously (Davis et al., 2017). Samples were analyzed on an Orbitrap Fusion Lumos instrument, coupled to an Ultimate 3000 nUPLC with an EASY-Spray column (50 cm). MS1 scans were acquired at a resolution of 120000 in the Orbitrap and MS2 scan were acquired in the ion trap in rapid mode. Selected precursors were excluded for 60 s. MS data was processed with MaxQuant (Version 1.6.2.3) and searched against the UniProt/Swissprot database (human, retrieved 18/07/2017). We used the FastLFQ algorithm for quantitation and the match between runs feature for identification transfer. Proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository (PRIDE: PXD011850).

SDS-PAGE and western blotting

SDS-PAGE and western blotting were performed using 7% Tris-Bicine gels with the SE400 and TE42 systems from Hoefer, or Bolt 4%–12% Bis-Tris Plus gels from Thermo Fisher Scientific. Biotinylated peptides were resolved using Novex 10%–20% Tricine gels from Thermo Fisher Scientific. Additional SDS-PAGE and western blotting were performed using BioRad 4%–15% Mini-PROTEAN TGXTM Precast Gels. The following antibodies were used at the indicated dilutions: AKT (9272, Cell Signaling, 1/1000), AKT-pS129 (ab133458, Abcam, 1/3000), ATM-pS1981 (2152-1, Epitomics, 1/5000), Biotin (ab53494, Abcam, 1/100), BLM (ab5446, Abcam, 1/500), CHK1 (sc-8408, Santa Cruz Biotechnology, 1/1000), CHK1-pS317 (2344, Cell Signaling, 1/500), CHK1-pS345 (2348, Cell Signaling, 1/10,000), CHK2 (2662, Cell Signaling, 1/1000), CHK2-pT68 (2661, Cell Signaling, 1/1000), GFP (11814460001, Roche, 1/5000), GST (sc-138, Santa Cruz Biotechnology, 1/500), H2A (01-146, Merck, 1/2500), H2AX (NB100-383, Novus Biologicals, 1/5000), γH2AX (05-636, Merck, 1/1000), H3 (ab1791, Abcam, 1/30,000), H3-pS10 (ab14955, Abcam, 1/5000), HA (sc-57592, Santa Cruz Biotechnology, 1/1000), KAP1 (ab10483, Abcam, 1/5000), KAP1-pS824 (IHC-00073, Bethyl Laboratories, 1/500), MDC1 (ab11171, Abcam, 1/5000), MDC1-pS168 (1/500), MDC1-pS196 (1/2000), NBS1 (ab32074, Abcam, 1/2000), RAD9 (ab70810, Abcam, 1/3000), RPA2 (ab2175, Abcam, 1/10,000), RPA2-pS4/pS8 (A700-009, Bethyl Laboratories, 1/1000), SMC1 (ab9262, Abcam, 1/5000), SMC1-pS966 (ab1267, Abcam, 1/2500), TOPBP1 (ab2402, Abcam, 1/1500). To generate the MDC1-pS168 affinity purified rabbit polyclonal antibody, the synthetic peptide sequence RLLLAED[pS]EEEVDFL, was conjugated to keyhole limpet hemocyanin (KLH) via a non-native N-terminal cysteine residue and used for immunization (AMS Biotechnology). MDC1 pS196 affinity purified polyclonal antibody was custom made by 21st Century Biochemicals (Marlboro), using the synthetic peptide sequence SVIVPE[pS]DEEGHSP for immunization.

Immunoprecipitations

For preparation of lysates for immunoprecipitations (IPs), cells were washed twice in phosphate-buffered saline (PBS), and lysed in IP buffer (100 mM NaCl, 0.2% Igepal CA-630, 1 mM $MgCl_2$, 10% glycerol, 5 mM NaF, 50 mM Tris-HCl, pH 7.5), supplemented with cOmplete EDTA-free protease inhibitor cocktail and 25 U/mL Benzonase (Novagen). After nuclease digestion, NaCl and EDTA concentrations were adjusted to 200 mM and 2 mM, respectively, and lysates were cleared by centrifugation. Lysates were then incubated with 15 μL of either GFP-Trap magnetic agarose beads (ChromoTek) or EZview Red Anti-HA Affinity Gel (Sigma-Aldrich) for 2 h with end-to-end mixing at 4°C. Immunoglobulin- antigen complexes were washed five times with IP buffer before elution in 2X SDS sample buffer.

Fluorescence polarization

For source and sequence information of fluorescein-labeled (Flu) peptides see key resource table. Recombinant human TOPBP1 BRCT0-2 and BRCT4+5 were expressed and purified as follows: GST-tagged BRCT0-2 or His-SUMO-tagged BRCT4+5 were transformed into *Escherichia coli* expression strain Rosetta 2 (DE3)pLysS (Merck). Single transformed colonies were used to inoculate a 250 mL flask containing 100 mL of Luria–Bertani broth (LB) supplemented with carbenicillin (100 μg/mL) and chloramphenicol (34 μg/mL). Inoculated cultures were grown overnight at 37°C, at 220 rpm in an orbital shaking incubator. The following day, 20 mL of overnight culture was used to inoculate a 2-l flask containing 1 l of LB, supplemented with antibiotics as above. Cultures were grown at 37°C, 220 rpm, until the optical density at 600 nm reached 0.6–0.8. They were then removed from the incubator, and rapidly cooled on ice for 30 min. Recombinant protein expression was induced by the addition of 0.4 mM isopropyl β-D-1-thiogalactopyranoside

(IPTG), and the cultures incubated for a further 16 h at 220 rpm at 16°C. Cells were then harvested by centrifugation, and the resulting pellet was stored at −80°C until required. The cell pellet arising from 6 l of culture was resuspended in 100 mL of 25 mM HEPES (pH 7.5), 1 M NaCl, 5% glycerol, 0.1% Tween-20, supplemented with protease inhibitors (Roche), then lysed on ice by sonication delivered by a Jencons Ultrasonic Processor (18 × 5 s bursts at 50% amplitude). Cell debris and precipitated material were removed by high-speed centrifugation at 48,834 x g for 60 min. For GST-tagged BRCT0-2, the supernatant was applied to a 5 mL HiTrap GST column (GE Healthcare), then washed with buffer containing 50 mM HEPES pH 7.5, 1 M NaCl, 250 μM TCEP. Retained protein was eluted by application of the lysis buffer supplemented with 20 mM glutathione. For His- SUMO-tagged BRCT4+5, the supernatant was applied to a 5 mL HiTrap TALON crude column (GE Healthcare), washed first with buffer containing 50 mM HEPES pH 7.5, 1 M NaCl, 250 μM TCEP, and then again with buffer containing 50 mM HEPES pH 7.5, 200 mM NaCl, 250 μM TCEP and 10 mM imidazole. Retained protein was eluted by application of the same buffer supplemented with 250 mM imidazole. Rhinovirus 3C PreScission protease (GE Healthcare) was added to the eluate and incubated overnight at 4°C to cleave the affinity tag. Both eluted TOPBP1 proteins were concentrated to a final volume of 5 mL (Vivaspin 20, 10 kDa MWCO, Sartorius Stedim Biotech) then applied to a HiLoad Superdex 200 16/60 size exclusion chromatography column (GE Healthcare) pre-equilibrated in 25 mM HEPES pH 7.5, 200 mM NaCl, 1 mM EDTA, 250 μM TCEP, 0.02% Tween-20. Flu-labeled peptides at a concentration of 100 nM were incubated at room temperature with increasing concentrations of TOPBP1 BRCT0-2 or BRCT4+5 in 25 mM HEPES (pH 7.5), 200 mM NaCl, 1 mM EDTA, 0.5 mM TCEP, 0.02% Tween-20. Fluorescence polarization was measured using a CLARIOstar multimode microplate reader (BMG Labtech). Data were analyzed using GraphPad Prism by non-linear fitting with a one-site total binding model. Any non-specific binding component was subtracted from the data for presentation purposes only. All data represent the mean of four separate experiments, and error bars represent one standard deviation.

In vitro kinase assays and GST-pull-downs

GST fusion proteins were expressed in the competent *Escherichia coli* expression strain BL21(DE3)pLysS (Promega) followed by protein purification with glutathione–Sepharose beads. GST fusion proteins were mock treated or treated with 120 U recombinant CK2 (New England Biolabs) with 10 mM ATP in NEB PK buffer (New England Biolabs) and incubated at 30°C for 45 min. For GST-pull-downs, 5 μg of the GST-MDC1 fusion proteins were then mixed with 250 μg HeLa nuclear extract and incubated for 1 h at 4°C. Glutathione–Sepharose beads (GE Healthcare) were added and the suspension was incubated at 4°C for 2 h on a rotating wheel. The beads were washed three times with wash buffer (50 mM Tris, pH 7.5, 120 mM NaCl, 1 mM DTT, and 0.3% NP-40) and bound proteins were eluted by addition of 2X SDS sample buffer.

Generation of CRISPR/Cas9 MDC1 knock-out cell lines

The U2OS Δ*MDC1* cell line was generated as follows, according to a protocol published previously (Ran et al., 2013). Sources and sequences of the reagents used are provided in the key resource table. The sgRNA target sequence was cloned into the pX330-U6-Chimeric_BB-CBh-hSpCas9 vector (a gift from Feng Zhang; Addgene plasmid #42230; http://n2t.net/addgene/42230; RRID:Addgene 42230) and verified by sequencing. Transfection of U2OS cells was done by electroporation of 10^6 cells with 5 μg of plasmid DNA on a NEON transfection system (Invitrogen), using the following electroporation parameters: pulse voltage 1230 V; pulse width 10 ms; pulse number 4. Clonal cell lines were isolated by dilution in 96-well plates. Total extract of single clones was prepared and analyzed by western blotting, using rabbit polyclonal anti-MDC1 antibody (Abcam). Total deletion of the MDC1 gene was verified in one of the clonal cell lines by Sanger sequencing of a PCR-amplified genomic fragment that was cloned in the pCR II Blunt TOPO plasmid (sequences available upon request).

Generation of stably transfected *ΔMDC1* cell lines

For generation of the GFP-MDC1 expressing stable cell lines, U2OS Δ*MDC1* cells were transfected with the indicated plasmids using Lipofectamine 2000, following the standard protocol provided by the supplier. The media was changed 10 h post-transfection. Two days post-transfection, cell colonies were grown in the presence of 400 μg/mL Geneticin G418 for 10 days with a media change every 3 days. Single colonies were picked using cloning rings and expanded. Clones were analyzed for GFP-MDC1 expression by western blotting and immunofluorescence, respectively. To improve homogeneous GFP-MDC1 expression, positive clones were sorted using a BD FACSAria III 4L cell sorter.

Immunofluorescence

Cells were grown on glass coverslips and fixed with either ice-cold methanol for 10 min (for staining with TOPBP1 antibody), or with 4% buffered formaldehyde (for staining with all other antibodies) for 15 min at room temperature, and subsequently permeabilized for 5 min in PBS containing 0.2% Triton X-100. Following 1 h of blocking in blocking buffer (10% FBS, 3% BSA in PBS), primary antibody incubations were performed at room temperature for 2 h. Coverslips were washed three times with PBS and secondary antibody incubations were performed for 1 h at room temperature in the dark. After washing with PBS three times, coverslips were mounted on glass microscopy slides with VECTASHIELD mounting medium containing 0.5 μg/mL 4',6-diamidino-2-phenylindole dihydrochloride (DAPI). The following antibodies were used at the indicated dilutions: 53BP1 (sc-22760, Santa Cruz, 1/500), CENPA (ab13939, Abcam, 1/200), Cyclin A (611269, BD Biosciences, 1/100), γH2AX (05-636, Millipore, 1/500), MDC1 (ab50003, Abcam, 1/300), NBS1 (NB100-143, Novus, 1/200), RPA2 (ab2175, Abcam, 1:250), TOPBP1 (A300-111A, Bethyl, 1/500).

Widefield microscopy

Widefield image acquisition was done on a ZEISS AxioObserver Z1 widefield microscope, equipped with a Lumencor SpectraX illumination system and a Hamamatsu Orca Flash 4.0 V2, sCMOS, cooled fluorescence camera (16-bit, 2048 × 2048 pixel (4 MP), pixel size 6.5 μm). A 63x, 1.4-NA, i-plan apochromat oil-immersion objective was used. For three-emission wavelength detection we combined DAPI with EGFP or Alexa Fluor 488 and Alexa Fluor 568. For optimal representation in figures, images were adjusted for brightness and exported as RGB TIF files using Fiji (Schindelin et al., 2012). Images in Figure S2G were deconvoluted with Huygens software (Scientific Volume Imaging) using a computed theoretical PSF based on known microscopic parameters.

Confocal microscopy

Confocal images were acquired with a Leica SP8 inverse confocal laser scanning microscope with a 63x, 1.4-NA Plan-Apochromat oil-immersion objective. For four-emission wavelength detection we combined DAPI with EGFP, Alexa Fluor 568 and Alexa Fluor 647. For three- emission wavelength detection we combined DAPI with EGFP and Alexa Fluor 568. The sequential scanning mode was used and the number of overexposed pixels was kept at a minimum. At least 10 single cells were recorded per condition with 10-15 z sections and a resolution of 512 × 512 pixels (voxel size 100 × 100 × 298 nm), 8-bit depth. For optimal representation in figures, maximum intensity projections were calculated and images were adjusted for brightness and exported as RGB TIF files using Fiji. For quantitative assessment of protein colocalization, the SQUASSH plugin (part of the MosaicSuite) for ImageJ and Fiji was used (Rizk et al., 2014). SQUASSH combines the image-segmentation task with biological reality through prior knowledge about the imaged objects (in this case DNA damage foci), the image-formation process and the noise present in the image. Maximum intensity projections of two channel confocal micrographs of single cells (EGFP and Alexa Fluor 568) were used for the SQUASSH analysis. Segmentation parameters were as follows: regularization ch1: 0.05, regularization ch2: 0.05, minimum object intensity ch1: 0.55, minimum object intensity ch2: 0.55, include subpixel segmentation: yes, exclude z edge: yes. Noise model: Poisson, standard deviation xy: 0.8 pixels, standard deviation z: 0.8 pixels. A script for the free open-source statistical software R is automatically generated by the SQUASSH plugin, which calculates object number colocalization (number of objects that overlap $\geq$ 50% divided by total number of objects) and object size colocalization (area of object colocalization divided by total object area). Due to slight heterogeneity of the GFP signal in the stably transfected cells we excluded signal co-localization from the analysis.

Quantification of γH2AX foci

For quantification of γH2AX foci in Figure 5B, cells were grown on coverslips, synchronized using RO-3306 and released for 40 min to allow cells to enter mitosis. Mitotic cells were irradiated at 0.5 Gy and subsequently released in fresh medium and left to grow for 24 h in the cell culture incubator. Cells were then fixed and stained with mouse monoclonal anti-γH2AX antibody as described above. Quantification was done manually. For quantification of γH2AX foci in Figure S3D, unsynchronized cells were irradiated with 3 Gy, followed by incubation at 37°C for 1 h. Cells were then fixed and stained with mouse monoclonal anti-γH2AX antibody as described above. Image acquisition was done using widefield microscopy. Quantification was done using CellProfiler (Carpenter et al., 2006) and R (CellProfiler analysis pipeline and R script available upon request).

Quantification of micronuclei

For quantification of micronuclei formation after irradiation of mitotic cells, cells where either untreated or irradiated with 0.5 Gy in prometaphase after a nocodazole arrest and subsequently released from the cell cycle arrest by washing twice with PBS and adding fresh medium for 4 h to allow cells to enter G1 phase. Cell were then fixed and stained with DAPI as described above. For quantification of spontaneous micronuclei formation, asynchronous and untreated cell populations were grown on coverslips, fixed and stained with DAPI as described above. Images were captured by widefield microscopy. Micronuclei were counted manually and nuclei were counted using the CellProfiler software. To assess the presence of centromeres in the micronuclei, fixed cells were stained with CENPA antibodies (see above). To probe for the presence of telomeric repeats in the micronuclei, fluorescence *in situ* hybridization (FISH) was done using Telomere PNA FISH Kit/Cy3 (Agilent Technologies) according to the protocol provided by the supplier.

Clonogenic survival assay

For clonogenic survival assay of interphase cells, 5×10^5 cells were counted and plated, followed by irradiation with the indicated doses and replating in 6-well plates. For mitotic cells, 2.5×10^6 cells were counted, plated and arrested in prometaphase using 100 ng/mL nocodazole for 16 h, followed by irradiation with the indicated doses. Following a mitotic shake off, cells were counted and re-plated in 6-well plates. For both interphase and mitotic assays, cells were grown with medium change every 3 days for 12 days. Plates were washed twice with PBS and incubated for 1 h in methylene blue fixation/staining solution (80% EtOH, 1% methylene blue powder) at room temperature. After rigorous washing with warm tap water, plates where left to dry overnight and colonies were counted manually using an eCountTM colony counter pen.

Metaphase analysis

Cells were grown to 90% confluency on a 10-cm plate and were treated with 0.1 μg/mL KaryoMax Colcemid (Thermo Fisher Scientific) for 2 h. Cells were trypsinized and transferred to a 15 mL Falcon tube, centrifuged at 1000 rpm for 5 min and carefully

resuspended in 5 mL hypotonic buffer (15% FBS, 75 mM KCl) and incubated for 15 min at 37°C. Cells were again pelleted at 1000 rpm for 5 min, the supernatant was discarded and the cell pellet was resuspended in 200 μL of hypotonic buffer. Cells were fixed by adding drop-wise 7 mL of EtOH:acetic acid 3:1 while slowly vortexing followed by incubation for 20 min. After centrifugation at 1000 rpm for 5 min, supernatant was discarded and cells were re-suspended in remaining 200 μL of the fixation buffer. 20-25 μL of the cell suspension was then dropped in a 45° angle onto a wet glass slide and air-dried for 10 min. Metaphases were stained with DAPI (VECTASHIELD with DAPI), covered with a glass coverslip and sealed with nail varnish. Telomere FISH was conducted using the Telomere PNA FISH Kit/Cy3 (Agilent Technologies) according to the protocol provided by the supplier. For telomere/centromere co-FISH, the same protocol as above was applied with the following adjustments: fixation of metaphase spreads was done with 3:1 MeOH:acetic acid for 20 min on ice. The CENPB probe (Alexa 488-labeled PNA CENPB probe from PNA Bio) was dissolved in formamide and combined with the telomere PNA probe to a final concentration of 5 μM. The hybridization time and temperature were adjusted to 8 min at 70°C.

Airyscan high-resolution confocal microscopy

Cells were grown in 35-mm #1.5H glass bottom dishes (IB-81158, Thistle Scientific). Cells were washed once in PBS before fixing for 10 min in 4% paraformaldehyde (PFA; Santa Cruz Biotechnology sc-281692) at room temperature. PBS was added simultaneously during PFA removal to prevent samples from drying due to PFA evaporation. Cells were washed once in PBS, before permeabilization in 0.2% Triton X-100 (Thermo Fisher Scientific) in PBS for 5 min. Samples were washed once in PBS before incubation with primary antibodies diluted in antibody buffer (DMEM supplemented with 10% FCS and 0.05% sodium azide; filtered) for 1 h. Cells were washed three times in 0.2% Tween 20 (Sigma-Aldrich) in PBS, before incubation with secondary antibodies diluted in antibody buffer containing 0.5 μg/mL DAPI in the dark for 30 min. Samples were washed three times in 0.2% Tween 20 in PBS and once in PBS before imaging in glass bottom dishes. Airyscan confocal imaging was carried out using an LSM 880 Airyscan inverted microscope (ZEISS) equipped with a 63x/1.4 numerical aperture (NA) Plan Apochromat objective and an Airyscan 32-pinhole detector unit. DAPI was detected using a 405-nm diode laser and 420-480 nm plus 495-550 nm band pass emission filters, Alexa 488 and GFP were detected with the 488-nm line of an argon laser and 420-480 nm plus 495-550 nm band pass emission filters, and Alexa 568 was detected using a 561-nm diode-pumped solid-state laser and 495-550 nm band pass plus 570 nm long pass emission filters. Z stacks of entire cells were acquired using 0.16 μm step size. Raw data were processed using Airyscan processing with Wiener Filter strength 6 using ZEN Black software version 2.1, yielding 16-bit images with approximately 180 nm lateral resolution. Display of images was adjusted for intensity for optimal display of structures of interest. 3D data were visualized and analyzed using Imaris 9.1.2 (Bitplane, Oxford Instruments) in-built animation tool. Intensities were adjusted for optimal display purposes.

Flow cytometry analysis of mitotic cells

Cells were treated with 10 μM CX-4945 or DMSO vehicle control for 16 h before harvesting by trypsinisation. Cells were washed with 1 mg/mL BSA in PBS, before fixation and permeabilization using Fixation/Permeabilization Solution Kit (BD Biosciences) according to the manufacturer's instructions. Cells were stained with H3-pS10 primary antibody (ab14955, Abcam, 1/100) for 1 h at RT, followed by Alexa Fluor 647 goat anti-mouse IgG1 (A-21240, Thermo Fisher Scientific, 1/200) for 1 h at RT. Cells were resuspended in 250 μg/mL RNase A, 2 μg/mL DAPI and 0.02% sodium azide in PBS, incubated at 37°C in the dark for 30 min before being analyzed using an Attune NxT flow cytometer (Thermo Fisher Scientific).

QUANTIFICATION AND STATISTICAL ANALYSIS

Each experiment was repeated at least twice, and representative experiments are depicted. Statistical calculations were done using GraphPad Prism 7 (GraphPad Software Inc.). Unpaired t tests were used to compare two populations. One-way ANOVA was used to compare more than two populations. Linear regression was used to compare clonogenic survival data. Sample sizes and α levels are indicated in the Figure Legends. In all cases: ns indicates $p > 0.05$.

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-13981-x OPEN

Treacle controls the nucleolar response to rDNA breaks via TOPBP1 recruitment and ATR activation

Clémence Mooser[1,5], Ioanna-Eleni Symeonidou[1,5], Pia-Amata Leimbacher[1], Alison Ribeiro[1], Ann-Marie K. Shorrocks[2,3], Stephanie Jungmichel[4], Sara C. Larsen[4], Katja Knechtle[1], Arti Jasrotia[1], Diana Zurbriggen[1], Alain Jeanrenaud[1], Colin Leikauf[1], Daniel Fink[1], Michael L. Nielsen[4], Andrew N. Blackford[2,3] & Manuel Stucki[1*]

Induction of DNA double-strand breaks (DSBs) in ribosomal DNA (rDNA) repeats is associated with ATM-dependent repression of ribosomal RNA synthesis and large-scale reorganization of nucleolar architecture, but the signaling events that regulate these responses are largely elusive. Here we show that the nucleolar response to rDNA breaks is dependent on both ATM and ATR activity. We further demonstrate that ATM- and NBS1-dependent recruitment of TOPBP1 in the nucleoli is required for inhibition of ribosomal RNA synthesis and nucleolar segregation in response to rDNA breaks. Mechanistically, TOPBP1 recruitment is mediated by phosphorylation-dependent interactions between three of its BRCT domains and conserved phosphorylated Ser/Thr residues at the C-terminus of the nucleolar phosphoprotein Treacle. Our data thus reveal an important cooperation between TOPBP1 and Treacle in the signaling cascade that triggers transcriptional inhibition and nucleolar segregation in response to rDNA breaks.

[1] Department of Gynecology, University of Zurich, Wagistrasse 14, CH-8952 Schlieren, Switzerland. [2] Department of Oncology, Medical Research Council Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK. [3] Cancer Research UK/Medical Research Council Oxford Institute for Radiation Oncology, University of Oxford, Oxford OX3 7DQ, UK. [4] The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Faculty of Health and Medical Sciences, Bledgamsvej 3B DK-2200, Copenhagen, Denmark. [5] These authors contributed equally: Clémence Mooser, Ioanna-Eleni Symeonidou. *email: manuel.stucki@uzh.ch

Double-strand breaks (DSBs) are highly toxic lesions that need to be repaired accurately in order to avoid chromosomal instability, a hallmark of various human disorders, including cancer. Mammalian cells respond to DSBs by activating a complex network of signaling pathways generally referred to as the DNA damage response (DDR)[1], which regulates various cellular processes such as the induction of cell-cycle checkpoints, apoptosis or senescence and the coordination of DNA repair pathway choice.

Central players in the cellular response to DSBs are the phosphoinositide 3-kinase (PI3K)-related protein kinases (PIKKs), including ataxia-telangiectasia mutated (ATM) and ATM and RAD3-related (ATR). ATM functions mainly in response to DSBs while ATR is activated by a much wider range of genotoxic insults. Both ATM and ATR require specific cofactors for their recruitment and optimal activation. The MRE11/RAD50/NBS1 (MRN) complex is required for ATM and ATR activation in response to DSBs[2–5]. ATR is recruited to sites of DNA damage via its interacting protein ATRIP that binds to ssDNA coated with the ssDNA-binding protein RPA (RPA-ssDNA)[6]. Such RPA-ssDNA is generated by nucleolytic processing of DNA lesions or by the uncoupling of DNA polymerases and helicases during DNA replication[7]. Several additional cofactors are required for the optimal activation of ATR at sites of DNA damage. Among these, the best studied is TOPBP1, which contains an ATR activation domain (AAD) that stimulates ATR kinase activity[8].

DSBs also induce transcriptional repression[9–14]. This effect has first been described in the nucleoli where the ribosomal gene (rDNA) arrays that make up for the highest transcribed sites in the human genome are located[9]. In human cells, ~300 rDNA repeats are distributed among nucleolar organizer regions (NORs). The NORs are located on the short arms of the five acrocentric chromosomes and encode the precursor ribosomal RNA (pre-rRNA) that is synthesized by the RNA polymerase I (Pol I) transcription machinery. Nucleoli are composed of three distinct subdomains that reflect the stages of ribosome biogenesis. Fibrillar centers (FCs) contain pools of transcription factors and non-transcribed rDNA sequences. Pol I transcription takes place at the interface between FCs and the surrounding dense fibrillar component (DFC) where early processing of the synthesized pre-rRNA occurs. FCs and DFC are embedded in the granular component (GC) where late processing of pre-rRNAs occurs, yielding mature rRNAs that together with the ribosomal proteins assemble into ribosome subunits[15].

The induction of breaks in the rDNA repeats triggers rapid ATM-dependent shutdown of Pol I transcription in the nucleoli[9]. Transcriptional repression is accompanied by a large-scale structural alteration of the nucleoli whereby the rDNA repeats along with associated proteins translocate from inside the nucleoli into the nucleolar periphery, where they accumulate in focal structures that are called nucleolar caps[16–19]. These structural changes—often referred to as nucleolar segregation—are also induced by inhibition of Pol I transcription by small-molecule inhibitors, thus suggesting that transcriptional repression is the underlying cause of nucleolar segregation[20]. While nucleolar segregation has been proposed to facilitate the repair of rDNA breaks[16,17,19], the mechanism by which ATM signaling suppresses rRNA transcription and the functional implication of transcriptional repression and nucleolar segregation have not yet been studied in detail.

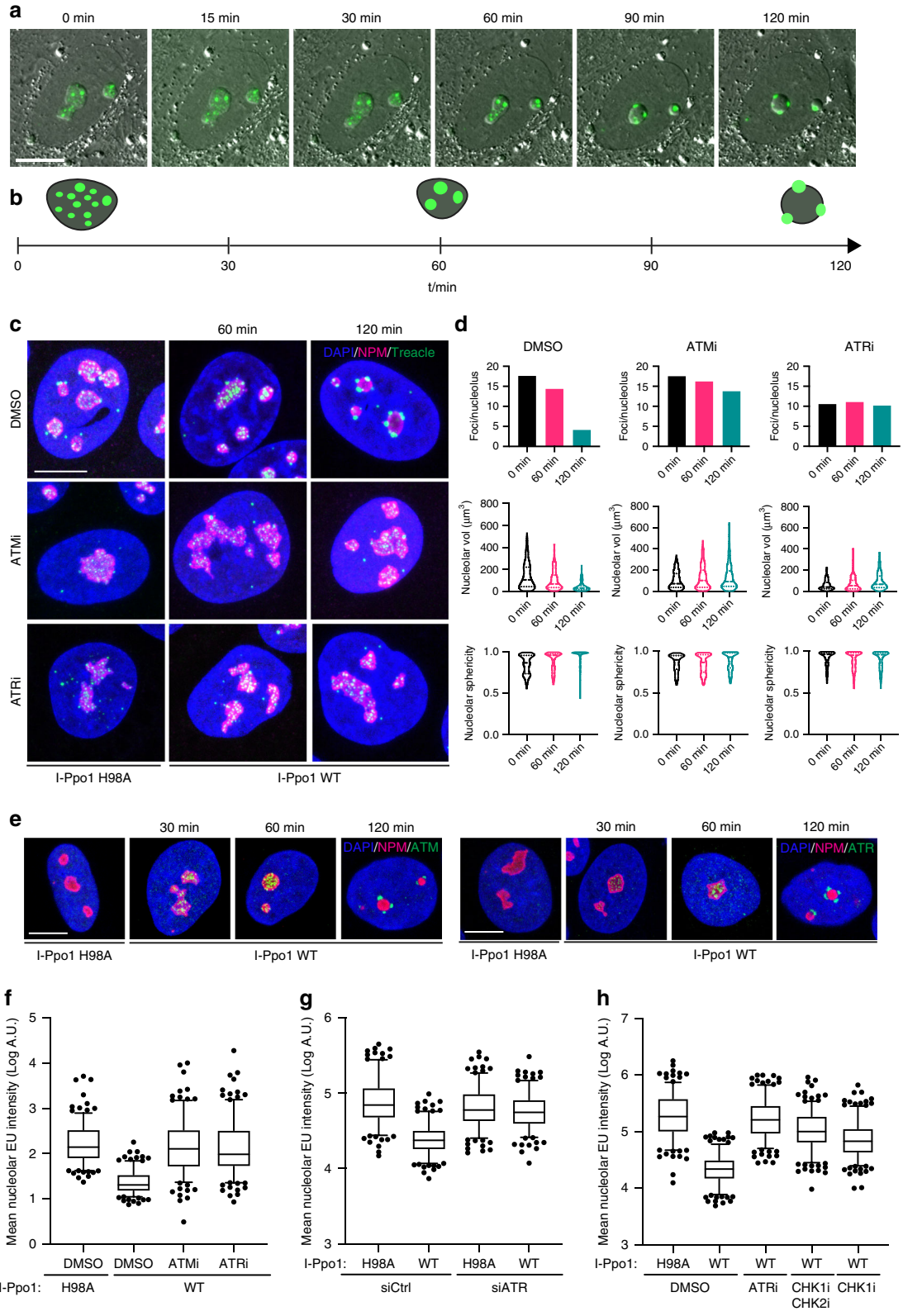
We and others recently discovered that NBS1, a subunit of the MRN complex, is rapidly and transiently recruited in the nucleoli in response DSBs[21,22]. This recruitment event is dependent on a direct phosphorylation-dependent interaction between the N-terminal FHA/BRCT region of NBS1 and Treacle (also called *TCOF1*), a low complexity nucleolar phosphoprotein whose gene is mutated in Treacher-Collins syndrome, a congenital disorder of craniofacial development[23]. Even though nucleolar localization of NBS1 has been associated with transcriptional repression of the rDNA repeats, the exact mechanism by which NBS1 contributes to the shutdown of Pol I transcription remains elusive.

Here we show that transcriptional repression and nucleolar segregation in response to targeted induction of DSBs in the rDNA repeats depend on both ATM and ATR activity. We furthermore show that Treacle recruits the ATR activator TOPBP1 in the nucleoli in an ATM- and NBS1-dependent manner. Mechanistically, TOPBP1 is recruited through direct phosphorylation-dependent interactions with the C-terminal region of Treacle. We propose that Treacle orchestrates the nucleolar response to rDNA breaks primarily through recruiting the two key adaptors NBS1 and TOPBP1 in the nucleoli, both of which are implicated in ATR activation.

Results

ATR-dependent Pol I inhibition and nucleolar segregation. In order to follow the dynamics of the cellular response to rDNA breaks in real time, we generated U2OS cell lines stably expressing low levels of green fluorescence protein (GFP)-tagged Treacle (Supplementary Fig. 1a). Characterization of these cell lines showed that GFP-Treacle exhibited the same subcellular localization as the endogenous Treacle, concentrating at the FC and DFC within the nucleoli (Fig. 1a). To induce targeted breaks in the rDNA repeats, we adopted the I-Ppo1 mRNA transfection approach previously described[16]. I-Ppo1 recognizes a sequence within the 28S portion of the rDNA repeat as well as a few sites elsewhere in the genome[24]. Live-cell microscopy was used to monitor changes in GFP-Treacle localization and changes in nucleolar morphology over time. Upon induction of breaks in the rDNA repeats, GFP-Treacle foci rapidly fused to form a few bright centers within the nucleoli before they migrated in the nucleolar periphery. At the same time, the nucleoli changed from irregularly shaped to round. By 120 min post I-Ppo1 mRNA transfection, the majority of nucleoli in the transfected cell population had completely segregated (Fig. 1a and Supplementary Movie 1).

We next asked if the DDR is implicated in the translocation of GFP-Treacle from inside the nucleoli to the nucleolar periphery in response to rDNA breakage. Surprisingly, we found that nucleolar segregation is not only blocked by ATM inhibitors, but also by ATR inhibitors (Supplementary Fig. 1b). We next assessed the localization of endogenous Treacle in response to I-Ppo1 mRNA transfection by quantitative 3D confocal microscopy (Supplementary Fig. 2a). These studies confirmed that migration of endogenous Treacle to the nucleolar periphery as well as changes in the nucleolar volume and sphericity in response to rDNA breaks are dependent on both ATM and ATR activity (Fig. 1c, d, Supplementary Fig. 2c). We observed that nucleolar segregation is blocked by ATR inhibitors in the majority of cells, irrespective of the cell-cycle stage (Supplementary Fig. 2d). The differences between control cells and inhibitor treated cells were not caused by variability in the I-Ppo1 expression levels nor did inhibitor treatment affect I-Ppo1 translation (Supplementary Fig. 3a–c). We next investigated ATM- and ATR recruitment to sites of I-Ppo1-induced DSBs in the nucleoli. Both ATM and ATR first accumulated inside of the nucleoli and at later timepoints, localized to nucleolar caps (Fig. 1e, Supplementary Fig. 2b). We also assessed ATM and ATR downstream signaling in response to I-Ppo1 transfection. CHK2 was efficiently phosphorylated on



Thr68 in an ATM-dependent manner, and CHK1 was phosphorylated on both Ser317 and Ser345 in an ATR-dependent manner (Supplementary Fig. 1c).

Next, we measured transcriptional activity within the nucleoli in response to rDNA breaks by 5-ethynyl uridine (EU) RNA labeling followed by Click-iT chemistry and by quantitative real-time PCR (qRT-PCR). Both assays showed strong ATM and ATR dependency of Pol I transcriptional inhibition in response to targeted induction of rDNA breaks by I-Ppo1 and in response to ionizing radiation (Fig. 1f, g and Supplementary Fig. 1d, e). Notably, CHK1 and CHK2 inhibition by two small-molecule inhibitors also significantly compromised transcriptional inhibition, albeit somewhat less efficiently than ATR inhibition (Fig. 1h).

Fig. 1 ATR-dependent Pol I inhibition and nucleolar segregation. a Time-lapse microscopy of U2OS cells expressing low levels of GFP-tagged Treacle, after transfection with I-Ppo1. **b** Timeline of the morphological changes of the nucleoli after I-Ppo1 transfection. **c** Morphological changes of the nucleoli after I-Ppo1 transfection in ATMi (KU-55933) and ATRi (VE-821) treated cells. Displayed are maximum intensity projections of confocal z-stacks. NPM stands for nucleophosmin **d** Quantification of Treacle foci number per nuceolus, nucleolar volume and nucleolar sphericity by segmentation of 3D reconstituted z-stacks. DMSO/0 min $n = 262$, DMSO/60 min $n = 239$, DMSO/120 min $n = 264$, ATMi/0 min $n = 174$, ATMi/60 min $n = 166$, ATMi/120 min $n = 179$, ATRi/0 min $n = 208$, ATRi/60 min $n = 218$, ATRi/120 min $n = 201$ independent cells. Bars represent means, dotted lines within the violin plots represent median and quartiles. **e** ATM and ATR recruitment to sites of nucleolar DSBs in response to I-Ppo1 transfection. Displayed are maximum intensity projections of confocal z-stacks. NPM stands for nucleophosmin. **f** Quantification of nucleolar EU incorporation after I-Ppo1 transfection in I-Ppo1 H98A/DMSO ($n = 215$), DMSO ($n = 205$), ATMi (KU-55933, $n = 206$) and ATRi (VE-821, $n = 219$) treated cells. **g** Quantification of nucleolar EU incorporation in siCtrl I-Ppo1 H98A and WT ($n = 203$) and siATR I-Ppo1 H98A and WT ($n = 210$) treated cells. **h** Quantification of nucleolar EU incorporation after I-Ppo1 H98A ($n = 229$) and I-Ppo1 WT transfection in control (DMSO, $n = 231$), ATRi (VE-821, $n = 237$), CHK1i/CHK2i (AZD7762, $n = 246$) and CHK1i (GDC-0575, $n = 255$) treated cells. **f**–**h** Boxes represent the 25–75 percentile range with median and whiskers represent the 5–95 percentile range. Data points outside of this range are shown individually. All scalebars = 10 µm. Source data are provided as a Source Data file.

In summary, these data reveal that transcriptional inhibition and nucleolar segregation in response to rDNA breakage require both ATM and ATR activity throughout the cell cycle.

NBS1 regulates the nucleolar response to rDNA breaks. As we and others previously showed that NBS1 interacts with Treacle and is rapidly recruited in the nucleoli in response to DSBs in an ATM- and Treacle-dependent manner[21,22], we next wished to determine if NBS1 is also implicated in the nucleolar response to I-Ppo1 expression. To this end, we targeted exon 2 of NBS1 in U2OS cells using CRISPR-Cas9 and isolated several clones with no detectable expression of full-length NBS1 (Supplementary Fig. 4a–c). One clone was subsequently analyzed in detail. Similar to cell lines derived from Nijmegen breakage syndrome patients, NBS1 knock-out U2OS cells showed increased radio-sensitivity, displayed defects in G2/M checkpoint activation, DSB resection and ATR activation in response to ionizing radiation (IR), and had MRE11 predominantly localized in the cytoplasm (Supplementary Fig. 5a–f). As it was recently shown that the C-terminal MRE11-interacting region in NBS1 is sufficient for viability and ATM activation[25], we suspected that the NBS1 knock-out U2OS cells may express a hypomorphic C-terminal fragment of NBS1. Indeed, the NBS1 mRNA is overexpressed in these cells and they express low levels of an ~40 kDa protein that is responsive to NBS1 mRNA knockdown by siRNA (Supplementary Fig. 5g, h). Based on these findings, we concluded that these cells most likely express low levels of a truncated NBS1 protein lacking the N-terminus. Therefore, we henceforth refer to these cells as NBS1ΔN.

We next addressed the role of NBS1 in transcriptional repression and nucleolar segregation in response to I-Ppo1 transfection. No nucleolar segregation was evident in NBS1ΔN cells within the first 2 h after transfection with I-Ppo1. (Fig. 2a). In addition, Pol I inhibition upon I-Ppo1 expression was compromised in NBS1ΔN cells (Fig. 2b). We next constructed NBS1ΔN cell lines stably expressing full-length NBS1, tagged at the C-terminus with the bright monomeric GFP variant mNeonGreen (mNG). Time-lapse microscopy of these cells revealed normal nucleolar segregation in response to I-Ppo1 expression. In addition, we consistently observed that NBS1-mNG is recruited into the nucleoli prior to nucleolar segregation (Fig. 2c, d, Supplementary Movie 2). NBS1-mNG co-localized with Treacle in the nucleoli and the introduction of point mutations in the N-terminal FHA and BRCT domains of NBS1 that have previously been shown to abrogate their ability to interact with phosphorylated proteins (R28A, K160M)[26], compromised nucleolar recruitment of NBS1 (Fig. 2e). Since NBS1 contacts Treacle via its FHA/BRCT domains[21], these data suggest that direct interaction with Treacle may also mediate its recruitment in the nucleoli. Indeed, depletion of endogenous Treacle by siRNA strongly reduced NBS1 nucleolar recruitment in response to I-Ppo1 expression (Fig. 2f). Treacle-mediated recruitment of NBS1 was prevented by ATM and ATR inhibition (Supplementary Fig. 6a). Since we also observed MRE11 recruitment in the nucleoli after I-Ppo1 transfection (Supplementary Fig. 6b), we propose that NBS1 acts in the nucleolar response to rDNA breaks in the context of the entire MRN complex.

In summary, these results confirm that the MRN complex is an essential upstream component of the signaling pathways that regulate transcriptional repression and nucleolar segregation in response to rDNA breaks[9,21].

rDNA resection occurs downstream of ATR activation. The MRN complex has been proposed to act upstream of ATR activation in response to DSBs through its essential role in DNA end resection, which produces ssDNA, the main signal recognized by ATR[4,5,27]. We therefore considered the possibility that ATM- and MRN-dependent resection of rDNA breaks may be required for ATR activation in the nucleoli. I-Ppo1 expression triggered the accumulation of the ssDNA binding protein RPA within nucleolar caps (Fig. 3a). In addition, RPA2, phosphorylated on Ser4 and Ser8 (RPA2 pS4/8), an established marker for resection, was also enriched within nucleolar caps (Supplementary Fig. 7a). At early time points after I-Ppo1 transfection, we observed RPA foci also inside the nucleoli, suggesting that DNA end resection occurs already before the broken rDNA repeats have moved out of the nucleoli (Fig. 3a). It was proposed that breaks in the rDNA repeats are resected throughout the cell cycle[16]. However, we found strong cell cycle dependency of RPA foci formation within the first 2 h after I-Ppo1 transfection, with only Cyclin A-positive S/G2-phase cells showing significant accumulation of RPA in the nucleolar periphery (Fig. 3b, c and Supplementary Fig. 7b, c). In NBS1ΔN cells, no RPA foci formed within the first 2 h after I-Ppo1 transfection (Fig. 3d, e). RPA pS4/8 foci were also strongly reduced in NBS1ΔN cells and by the MRE11 inhibitor Mirin, indicating that resection of broken rDNA is dependent on the nuclease activity of MRN complex (Supplementary Fig. 7a).

Surprisingly, we observed that treatment of cells with ATM and ATR inhibitors prior to I-Ppo1 transfection strongly abrogated RPA foci formation, suggesting that DNA end resection is not required for ATR activation in the nucleoli (Fig. 3f, g). In support of this, depletion of the DNA end resection factor CtIP failed to compromise nucleolar segregation, even though it strongly inhibited RPA foci formation in response to I-Ppo1 expression (Supplementary Fig. 7d).

These findings thus reveal a difference in the regulation of DNA end resection between DSBs induced in the rDNA repeats as compared to DSBs induced elsewhere in the genome that do not require ATR activity for resection.

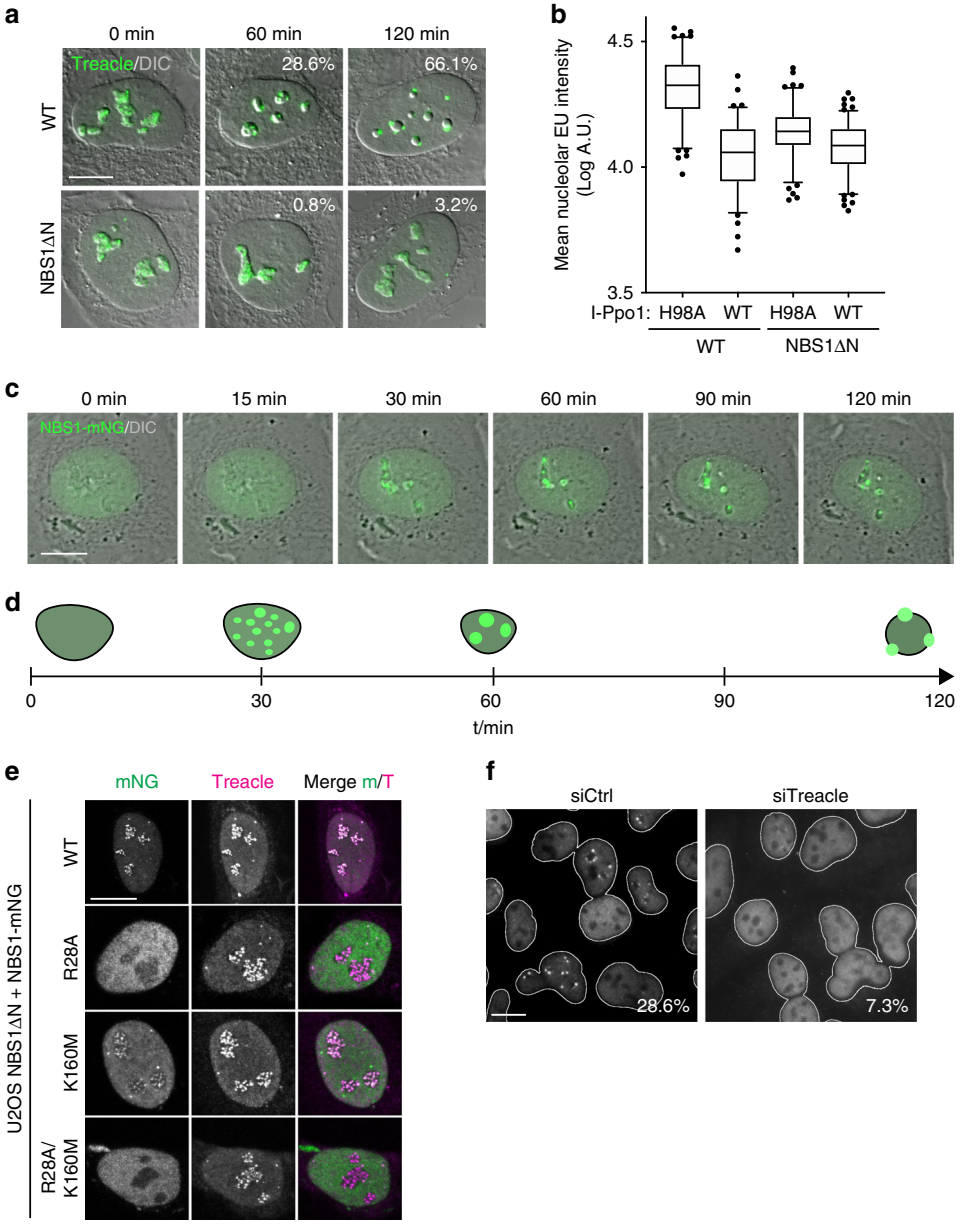


Fig. 2 NBS1 regulates the nucleolar response to rDNA breaks. a Timecourse of endogenous Treacle localization after I-Ppo1 transfection in normal U2OS cells and NBS1ΔN cells. **b** Quantification of nucleolar EU incorporation after I-Ppo1 H98A ($n = 111$) and I-Ppo1 WT ($n = 87$) transfection in control U2OS and after I-Ppo1 H98A ($n = 116$) and I-Ppo1 WT ($n = 124$) transfection in NBS1ΔN cells (boxes represent the 25–75 percentile range with median and whiskers represent the 5–95 percentile range. Data points outside of this range are shown individually). **c** Time-lapse microscopy of U2OS cells expressing mNG-tagged full-length NBS1 (wild type), after transfection with I-Ppo1. **d** Timeline of NBS1-mNG recruitment in relation to morphological changes in the nucleoli after I-Ppo1 transfection. **e** Confocal microscopy of NBS1-mNG wild type and point mutants in the NBS1ΔN background. **f** Localization of endogenous NBS1 2 h after I-Ppo1 transfection in siCtrl and siTreacle-treated cells (percentages of cells with >2 NBS1 caps are indicated; one of two experiments is shown). All scalebars = 10 μm. Source data are provided as a Source Data file.

TOPBP1 recruitment in response to rDNA breaks. Having established that DNA end resection is not required for nucleolar segregation in response to I-Ppo1 expression and occurs downstream of ATR activation, we next wished to investigate the mechanism of ATR activation in the nucleoli. The ATR activator TOPBP1 was previously implicated in nucleolar stress responses and its overexpression was shown to trigger ATR-dependent Pol I inhibition and nucleolar segregation even in the absence of rDNA damage[28]. Upon induction of rDNA breaks by I-Ppo1, TOPBP1 first accumulated inside of the nucleoli where it co-localized with Treacle. At later timepoints, after nucleolar segregation had taken place, TOPBP1 accumulated in nucleolar caps (Fig. 4a).

We next asked if TOPBP1 recruitment in the nucleoli is dependent on NBS1, ATM and ATR activity. Only few NBS1ΔN cells contained nucleolar TOPBP1 foci, and expression of wild-type NBS1 but not FHA/BRCT mutated NBS1 rescued this defect (Fig. 4b). In addition, TOPBP1 accumulation in nucleolar caps was also blocked by pre-treatment of cells with either ATM or ATR inhibitors (Fig. 4c, d). Interestingly though, TOPBP1 foci did form within the nucleoli of a substantial percentage of ATR-inhibited U2OS cells, suggesting that not the initial recruitment of TOPBP1 but rather its subsequent accumulation in the nucleolar periphery is dependent on ATR activity (Fig. 4c).

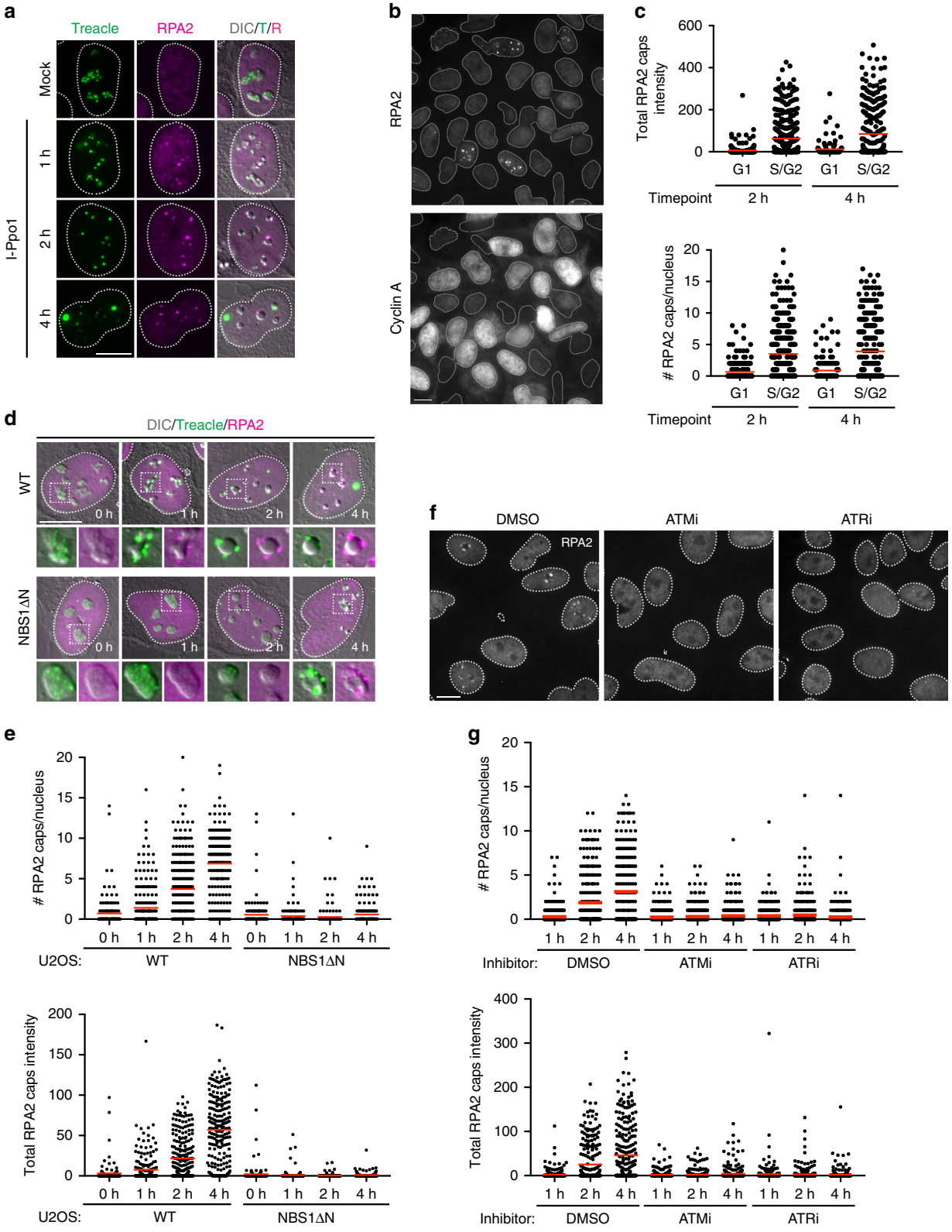


Fig. 3 rDNA resection occurs downstream of ATR activation. a Timecourse of RPA2 foci formation in U2OS cells after I-Ppo1 transfection. **b** Timecourse of RPA2 foci formation in cells counterstained for Cyclin A. **c** Quantification of the experiment in **b** (all data points shown, red bars represent mean). **d** Timecourse of RPA2 foci formation in U2OS WT and NBS1ΔN cells after I-Ppo1 transfection. **e** Quantification of the experiment in **d** (all data points shown, red bars represent mean). **f** Timecourse of RPA2 foci formation in U2OS cells treated with ATMi (KU-55933) and ATRi (VE-821) after I-Ppo1 transfection. **g** Quantification of the experiment in **f** (all data points shown, red bars represent mean). All scalebars = 10 μm. Source data are provided as a Source Data file.

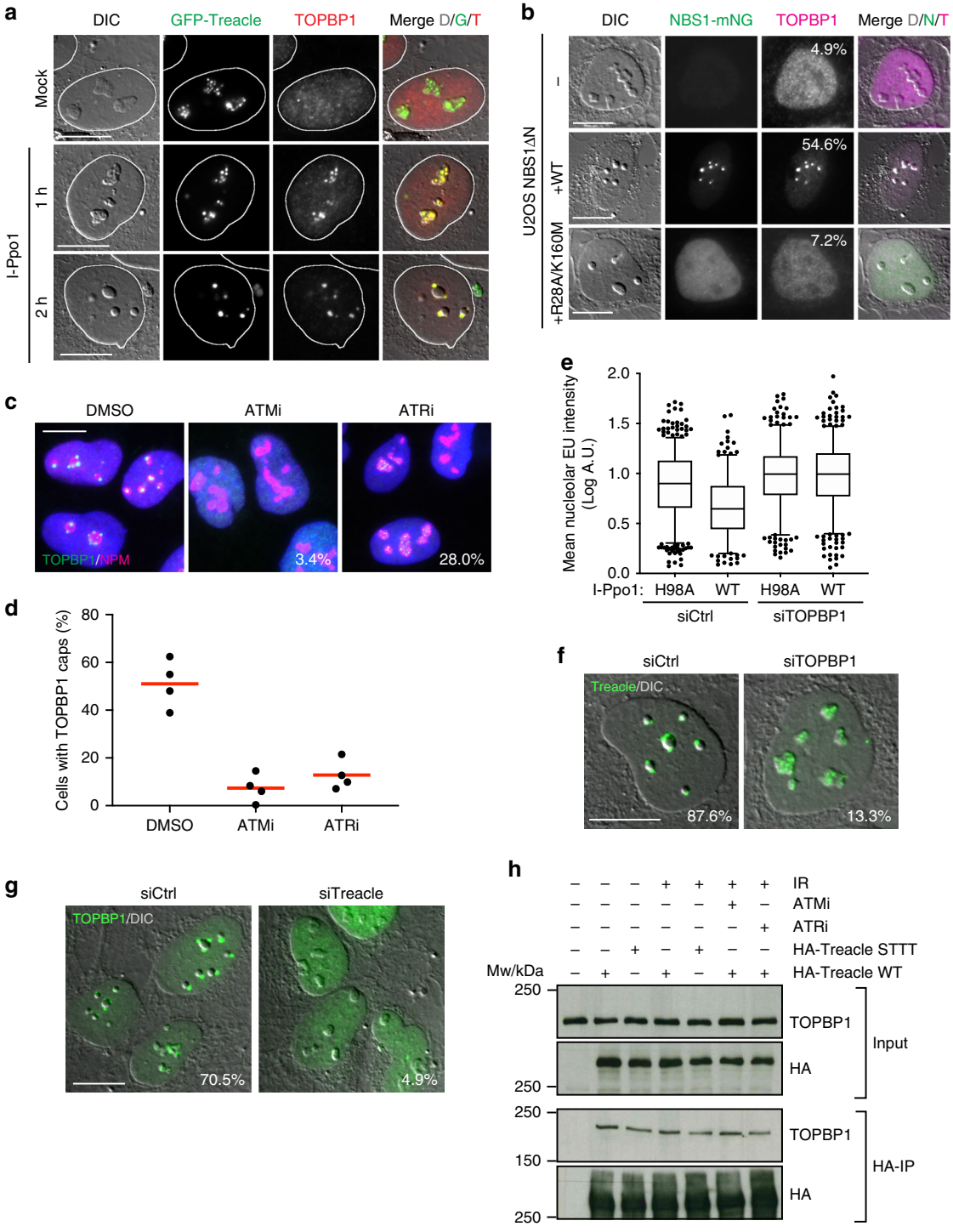


Fig. 4 TOPBP1 recruitment in response to rDNA breaks. a Timecourse of TOPBP1 localization in GFP-Treacle expressing U2OS cells after I-Ppo1 transfection. **b** TOPBP1 localization 2 h after I-Ppo1 transfection in NBS1ΔN cells and NBS1ΔN cells complemented with wild-type and mutant NBS1-mNG (percentage of cells with > 2 TOPBP1 caps are indicated; one of two experiments is shown). **c** TOPBP1 localization 2 h after I-Ppo1 transfection in ATMi- and ATRi-treated cells versus vehicle (DMSO)-treated cells (percentage of cells with TOPBP1 foci inside of the nucleoli are indicated; one of two experiments is shown). **d** Quantification of cells with TOPBP1 caps after ATMi and ATRi treatment versus vehicle (DMSO) treatment (red bars represent mean). **e** Quantification of nucleolar EU incorporation after I-Ppo1 H98A ($n = 692$) and I-Ppo1 WT ($n = 649$) transfection in control U2OS and after I-Ppo1 H98A ($n = 450$) and I-Ppo1 WT ($n = 521$) transfection in siTOPBP1-treated cells (boxes represent the median with 25–75 percentile range and whiskers represent the 5–95 percentile range. Data points outside of this range are shown individually). **f** Endogenous Treacle localization 2 h after I-Ppo1 transfection in control siRNA and TOPBP1 siRNA transfected cells (percentage of cells with > 2 Treacle caps are indicated; one of two experiments is shown). **g** TOPBP1 localization in Treacle-depleted U2OS cells and control cells 2 h after I-Ppo1 transfection (percentage of cells with > 2 TOPBP1 caps are indicated; one of two experiments is shown). **h** HA-immunoprecipitations from 293FT cells transfected with the indicated HA-tagged Treacle variants, in the presence or absence of ATMi and ATRi and with and without IR treatment as indicated. STTT mutated Treacle: S171A, T173A, T203A, T210A. All scalebars = 10 µm. Source data are provided as a Source Data file.

Downregulation of TOPBP1 expression in U2OS cells by siRNA significantly compromised transcriptional inhibition in response to I-Ppo1 expression (Fig. 4e and Supplementary Fig. 8a). Moreover, downregulation of TOPBP1 also blocked nucleolar segregation in the majority of cells (Fig. 4f). Based on the observation that TOPBP1 interacts with components of the MRN complex[29,30], we considered the possibility that the MRN complex may recruit TOPBP1 in the nucleoli through direct interaction. Consistent with this idea, downregulation of Treacle, the nucleolar adaptor for NBS1, also severely compromised TOPBP1 nucleolar recruitment and cap formation in response to rDNA damage (Fig. 4g). However, co-immunoprecipitation experiments revealed that HA-tagged Treacle mutants that lack the NBS1 interaction sites[21] still efficiently co-immunoprecipitated TOPBP1. Moreover, TOPBP1–Treacle complex formation was constitutive and did not depend on DNA damage or ATM and ATR activity (Fig. 4h). These data suggested that TOPBP1 may interact with Treacle directly and not via the MRN complex. In support of this idea, and consistent with previous findings[31], we identified Treacle as a candidate TOPBP1-interacting protein by mass spectrometric analysis of TOPBP1-associated proteins both in the presence and absence of DNA damage (Supplementary Fig. 9a, b and Supplementary Data 1).

In summary, these data reveal that TOPBP1 is an important regulator of the nucleolar response to rDNA damage and suggest that the MRN complex and Treacle mediate the recruitment of TOPBP1 in the nucleoli in response to rDNA breaks for subsequent ATR activation.

Characterization of the TOPBP1–Treacle interaction. We next sought to characterize the TOPBP1–Treacle interaction in more detail. Treacle consists of three structurally distinct regions: an N-terminal region that contains the phosphorylated Ser and Thr residues required for NBS1 interaction[21], a central region consisting of 10 consecutive acidic and serine-rich sequence stretches and a C-terminal region containing nuclear and nucleolar localization signals (Fig. 5a). In order to map the region of Treacle required for its interaction with TOPBP1, we generated a series of deletion mutants of HA-tagged Treacle and assessed their impact on the interaction with TOPBP1 by co-immunoprecipitation experiments. These data revealed that the main TOPBP1 interaction site is located within the C-terminal part of Treacle and that the NBS1 and TOPBP1 interaction sites on Treacle are distinct and do not overlap (Fig. 5b).

TOPBP1 is a modular protein consisting of nine BRCT domains (numbered from 0 to 8), two of which occur in tandems, one in a unique triple-BRCT configuration, and two appear to function as single domains[32,33]. BRCT 1, 2, 5 and 7 contain the necessary residues for phosphopeptide binding[34]. The AAD is located between BRCT domains 6 and 7 (Fig. 5c). BRCT tandems usually contain a conserved lysine residue that directly contacts the phosphate group of a modified protein ligand[35]. Mutating such residues significantly reduces ligand-binding affinity. We therefore mutated Lys residues in BRCT domains 1, 2, 5 and 7 to Ala in GFP-tagged full-length TOPBP1 and used these constructs in co-immunoprecipitation experiments. The analysis showed that, in accordance with previous work, mutation of TOPBP1 BRCT1 abolished TOPBP1 interaction with RAD9 but not with BLM, whereas mutation of BRCT5 disrupted TOPBP1 interaction with BLM but had little effect on RAD9 binding to TOPBP1 (Fig. 5d)[36,37]. Surprisingly, mutation of three BRCT domains (1, 2 and 5) completely abolished Treacle binding to TOPBP1, thus suggesting the presence of three phosphorylated TOPBP1 binding sites in Treacle (Fig. 5d). Of the three purified BRCT tandems fused to GST, only BRCT0-2 pulled down significant amounts of Treacle from Hela nuclear extracts (Supplementary Fig. 8c)

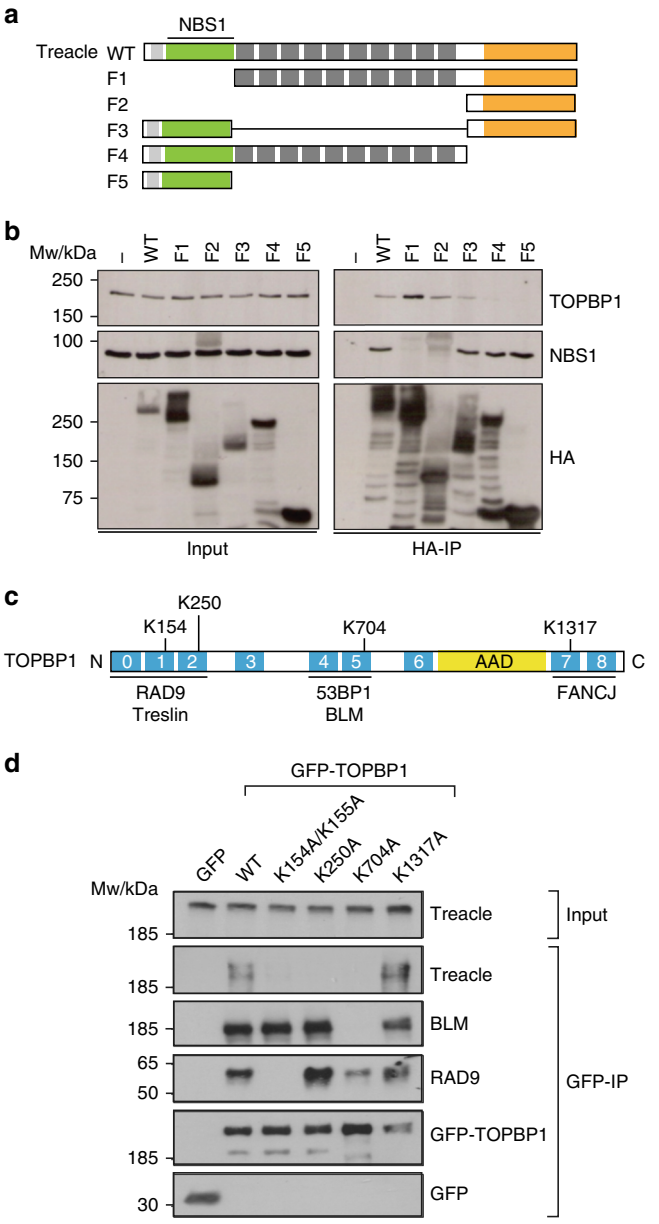


Fig. 5 Characterization of the TOPBP1–Treacle interaction. a Schematic showing the layout of conserved domains and motifs in Treacle as well as HA-tagged deletion constructs of Treacle, lacking either the N-terminal region (F1), both the N-terminal and the central repeat region (F2), the central repeat region (F3), the C-terminal region (F4) or both the central repeat and the C-terminal region (F5). HA-immunoprecipitation of full-length and five deletion constructs depicted in **a**. **c** Schematic showing the layout of conserved domains and motifs in TOPBP1. Names of known TOPBP1-binding partners are shown below the domains they interact with and phospho-interacting Lys residues in the three BRCT tandems are highlighted. AAD: ATR-activating domain. **d** GFP-pulldowns from 293FT cells transfected with the indicated GFP-tagged TOPBP1 constructs. Source data are provided as a Source Data file.

suggesting that interaction of Treacle with the isolated TOPBP1 BRCT4 + 5 may be too weak for detection in a GST-pulldown experiment.

A conserved TOPBP1 phospho-interaction motif in Treacle. Since complex formation with Treacle requires the phosphopeptide-binding function of three TOPBP1 BRCT domains, we next

mapped phosphorylation sites within the C-terminal region of Treacle by mass spectrometry. Several Ser/Thr phospho-residues were identified with high (>90%) positional probability (Supplementary Data 2). Of these, only a few are conserved. Three phosphorylation sites (Thr1223, Ser1227 and Ser1228) within a highly conserved acidic sequence stretch stand out because along with their surrounding amino acids, they bear significant similarities to known TOPBP1 BRCT0-2 interaction motifs in RAD9 and Treslin (Supplementary Fig. 8d)[31,36]. Interestingly, a highly conserved Ser/Gln (SQ) motif (the ATM/ATR target motif) previously suggested to interact with NBS1[22] is located immediately downstream of the putative BRCT0-2 interaction site. In addition, amino acids N-terminal of this SQ motif closely resemble the TOPBP1 BRCT4-5 interaction site in BLM (Supplementary Fig. 9d)[37,38]. Mutating the central two Ser residues within the putative TOPBP1 BRCT0-2 recognition site severely compromised TOPBP1–Treacle interaction (Fig. 6a). Mutation of the conserved SQ motif also slightly reduced the binding of HA-tagged Treacle to endogenous TOPBP1.

We next sought to investigate if mutation of these phospho-residues in Treacle would affect TOPBP1 recruitment in the nucleoli and transcriptional repression in response to rDNA breaks. To this end, we generated cell lines stably expressing siRNA-resistant GFP-tagged full-length Treacle (wild type and point mutants, respectively) and assessed in each cell line how well ectopic expression of Treacle could rescue the effect of endogenous Treacle downregulation. The analysis showed that in mock-transfected cells, downregulation of endogenous Treacle almost completely abrogated TOPBP1 nucleolar accumulation (Fig. 6b, c). This defect was rescued to control levels by stable ectopic expression of GFP-tagged wild-type Treacle. However, none of the mutants was able to fully complement for the loss of endogenous Treacle expression. Notably, the triple mutant was almost completely defective for TOPBP1 accumulation. In addition, the mutant Treacle-expressing cells showed significantly higher levels of nucleolar EU incorporation after I-Ppo1 transfection when endogenous Treacle was depleted, suggesting a defect in I-Ppo1-induced transcriptional repression in the absence of Treacle–TOPBP1 interaction (Fig. 6d).

In summary, these data reveal that phosphorylation-dependent TOPBP1 binding to a conserved sequence motif within the Treacle C-terminal domain is essential for its recruitment in the nucleoli and is necessary for optimal transcriptional repression in response to rDNA breaks.

Treacle- and TOPBP1-mediated nucleolar recruitment of ATR. The mechanism of ATR activation by TOPBP1 is not well understood. We thus considered the possibility that TOPBP1 may mediate ATR recruitment in the nucleoli in response to rDNA breaks. ATR localization in the nucleoli was strongly reduced in Treacle- and TOPBP1-depleted cells (Fig. 7a, b), suggesting that Treacle- and TOPBP1-mediated accumulation of ATR in the nucleoli may be required to ensure transcriptional repression and nucleolar segregation in response to rDNA breaks. In support of this idea, we observed that TOPBP1 overexpression not only triggered nucleolar segregation as previously reported[28], but it also mediated robust ATR accumulation in nucleolar caps (Fig. 7d). ATR activation by TOPBP1 relies on a conserved tryptophan residue that lies within the AAD (W1145). Both nucleolar segregation and ATR accumulation were completely abrogated upon overexpression of a W1145R mutant version of TOPBP1 even though this mutant was still enriched in the nucleoli where it co-localized with Treacle (Fig. 7d, Supplementary Fig. 8d). In Treacle-depleted cells, overexpression of wild-type TOPBP1 did not trigger nucleolar segregation (Fig. 7e). In summary, these findings support the idea that the nucleolar response to rDNA breaks depends on Treacle- and TOPBP1-mediated accumulation of ATR in the nucleoli.

Loss of Treacle and TOPBP1 impairs rDNA repair. We next aimed to establish the physiological significance of TOPBP1/ATR-mediated transcriptional inhibition and nucleolar segregation. Previous studies suggested that these events are required for efficient repair of rDNA breaks[17–19]. We therefore studied recruitment of the DSB repair machinery in response to rDNA breaks in Treacle- and TOPBP1-deficient cells. Depletion of either Treacle or TOPBP1 by siRNA prior to rDNA damage resulted in significantly reduced H2AX phosphorylation and 53BP1 accumulation in the nucleolar periphery. Moreover, Treacle- and TOPBP1-depleted cells were also unable to efficiently recruit RPA and the homologous recombination (HR) factors BRCA1 and RAD51 to nucleolar caps after I-Ppo1 treatment (Fig. 8a–d).

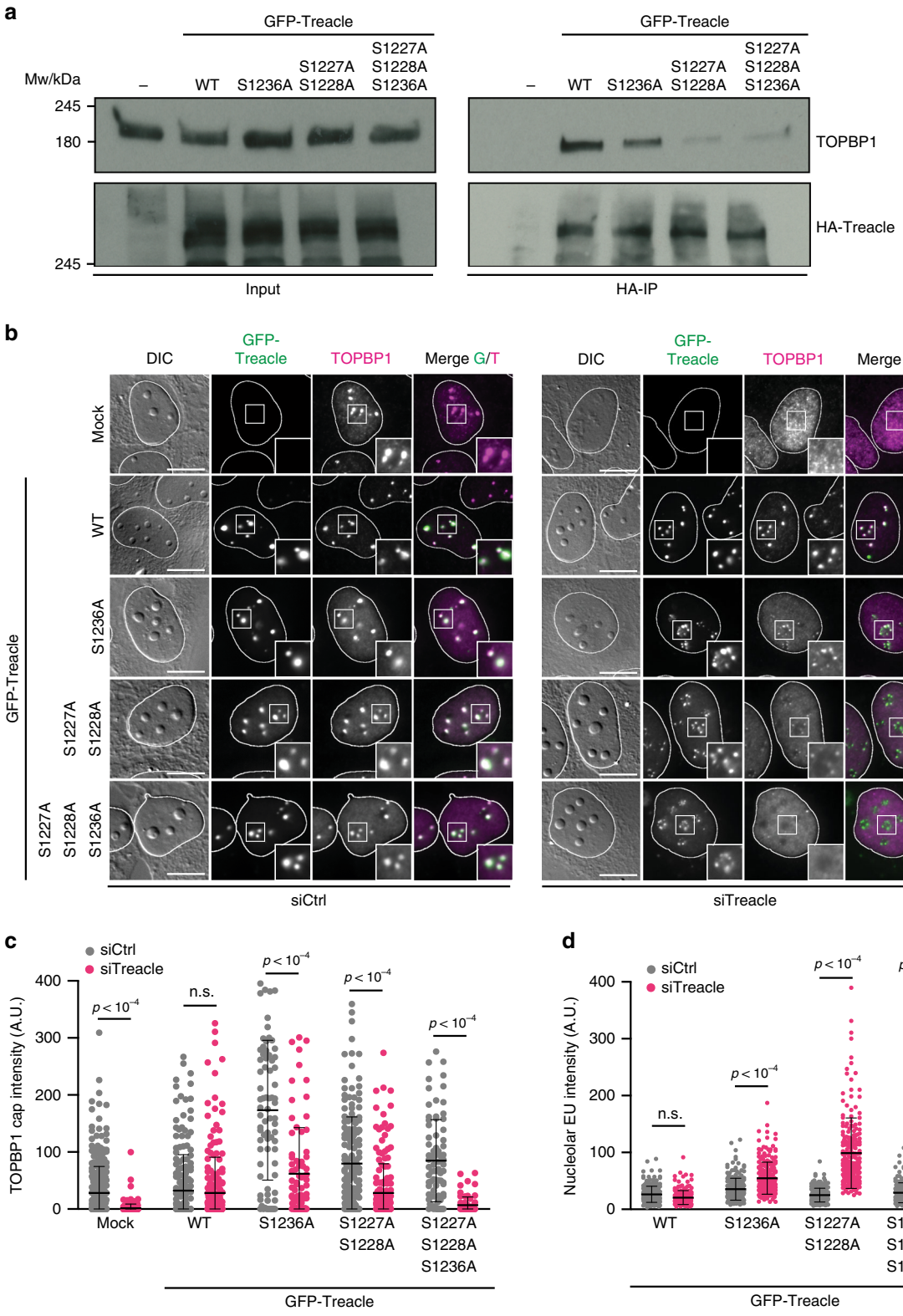
We next depleted Treacle and TOPBP1 by siRNA and measured cell viability in cells transfected with I-Ppo1 mRNA. Depletion of both Treacle and TOPBP1 resulted in a reduced ability of cells to survive DNA damage induced by I-Ppo1 transfection (Fig. 8e). In addition, NBS1ΔN cells were also hypersensitive to I-Ppo1 mRNA transfection when compared to wild-type U2OS cells (Fig. 8f). While the reduced survival may be caused by defective rDNA repair, we cannot rule out a contribution of defective DSB repair at I-Ppo1 sites that are located outside of the rDNA repeats.

Taken together, these data highlight the importance of Treacle as a nucleolar-specific adaptor for NBS1 and TOPBP1 and support a model in which these DDR factors accumulate in the nucleoli upon rDNA breakage in order to recruit and activate the ATR kinase, whose activity is required to repress Pol I transcription and mediate migration of the broken rDNA repeats out of the nucleoli into nucleolar caps, where they can be efficiently accessed by the DSB repair machinery (Fig. 8g).

Discussion

The findings reported here reveal a role of ATR kinase activity in the suppression of Pol I transcription and nucleolar segregation in response to rDNA breaks. ATR's role in DSB-induced cell-cycle checkpoint activation is well established[39]. However, its role in the cellular response to DSBs was mostly assigned to S/G2 phases of the cell cycle when sister chromatids are present. This is because extensive regions of RPA-coated ssDNA (RPA-ssDNA), a well-established signal for ATR activation, are not generated at sites of DSBs in G1-phase cells as DNA end resection that generates 3′ ssDNA overhangs is suppressed in G1 (ref. [40]). Despite this well-established model, evidence for ATR activation in G1-phase cells has been recently presented[41,42]. We report here that repression of Pol I transcription and nucleolar segregation are dependent on ATR activity throughout the cell cycle, even in G1 phase, when no sister chromatids are present, and resection of broken rDNA does not take place. Our data thus challenge the current view that RPA-ssDNA is an essential signal for ATR activation in response to DSBs and suggest that ATR may be efficiently activated without RPA-ssDNA in the nucleoli. Accordingly, there is evidence that extensive regions of RPA-ssDNA may not be the signal for ATR activation in response to DSBs also outside of the nucleoli[43–45].

An important element in the process of transcriptional repression and nucleolar segregation in response to rDNA breaks is the accumulation of the MRN complex and TOPBP1 in the nucleoli. Depletion of Treacle not only abrogates recruitment of both MRN and TOPBP1 in the nucleoli, but also results in



defective recruitment of ATR in the nucleoli and renders cells incapable to undergo nucleolar segregation. However, Treacle-mediated recruitment of the MRN complex and TOPBP1 are unlikely to represent the initial signal for ATM and ATR activation in the nucleoli. We therefore consider it more likely that Treacle lies in the center of an auto-amplification mechanism, which ultimately leads to the accumulation of ATR in the nucleoli. Several lines of evidence support this notion: first, efficient Treacle-mediated accumulation of NBS1 and TOPBP1 requires ATM and ATR activity, indicating that the initial activation of these kinases at sites of rDNA breaks occurs independently of Treacle. Second, TOPBP1 recruitment in the nucleoli occurs independently of ATR in a subset of cells, and third, overexpression of TOPBP1 leads to ATR-dependent transcriptional repression in the nucleoli followed by nucleolar segregation even in the absence of rDNA breaks[28].

Fig. 6 A conserved TOPBP1 phospho-interaction motif in Treacle. a HA-immunoprecipitations from 293FT cells transfected with the indicated HA-tagged Treacle variants. **b** TOPBP1 localization 2 h after I-Ppo1 transfection in GFP-Treacle-expressing cell lines after treatment with control siRNA (left panel) and Treacle siRNA (right panel). **c** Quantification of the experiment in **b**, mock: siCtrl ($n = 301$) versus siTreacle ($n = 290$), WT: siCtrl ($n = 226$) versus siTreacle ($n = 242$), S1236A: siCtrl ($n = 75$) versus siTreacle ($n = 79$), S1227A/S1228A: siCtrl ($n = 169$) versus siTreacle ($n = 177$) and S1227A/S1228A/S1236A: siCtrl ($n = 77$) versus siTreacle ($n = 69$) cells. **d** Quantification of nucleolar EU incorporation after I-Ppo1 transfection in GFP-Treacle-expressing cell lines after treatment with control siRNA and Treacle siRNA. WT: siCtrl ($n = 216$) versus siTreacle ($n = 200$), S1236A: siCtrl ($n = 224$) versus siTreacle ($n = 224$), S1227A/S1228A: siCtrl ($n = 222$) versus siTreacle ($n = 227$) and S1227A/S1228A/S1236A: siCtrl ($n = 229$) versus siTreacle ($n = 232$) cells. **c**, **d** Statistical significance was calculated using a one-way analysis of variance and Sidak's multiple comparison test. Bars represent mean ± S.D. All scale bars = 10 μm. Source data are provided as a Source Data file.

Our data show that the interaction between Treacle and TOPBP1 is constitutive and DNA damage independent. The TOPBP1 interaction site in Treacle closely resembles known interaction motifs of the TOPBP1 BRCT0-2 module in RAD9 and Treslin[31,36,46]. In addition, our data are consistent with a constitutive and direct interaction between Treacle and TOPBP1 BRCT0-2. While BRCT4 + 5 appear to be also required for stable association between TOPBP1 and Treacle in cells, they do not efficiently interact with Treacle on their own. The reason for this is currently not clear but the sequence similarity of the motif around Ser1236 in Treacle with the TOPBP1 interaction site in BLM makes it a promising candidate for a TOPBP1 BRCT4 + 5 interaction site[37]. Further biochemical work is needed to identify the TOPBP1 BRCT4 + 5 interaction site in Treacle.

Our analysis revealed that DNA end resection of broken rDNA repeats takes place in S/G2 cells. This observation differs from the data presented by a previous work[16] showing cell-cycle independent end resection of the rDNA breaks. However, the discrepancy in the results between the two studies could be attributed to the different timing of the resection assay. We observed resection activity at early time points, and in particular within the first 2 h after transfection of the I-Ppo1 mRNA. At later time points (4 h), we also started to detect resection activity in G1-phase cells, especially in untransformed RPE-1 cells (Supplementary Fig. 7a, b), consistent with the data published by van Sluis and McStay, who assessed resection exclusively at late time points (6 h) after I-Ppo1 mRNA transfection[16]. We therefore propose that prolonged exposure of cells to I-Ppo1 generates persistent DSBs, which may be resected also in G1-phase cells.

Using live-cell imaging, we observed rapid fusion of FC and DFC in the nucleoli upon I-Ppo1 expression throughout the cell cycle, followed by "floating" of the fused FC and DFC to the nucleolar surface where they remain anchored in the nucleolar periphery. Moreover, we found that fusion of FC/DFC and nucleolar segregation occurs downstream of ATR activation, but independently of resection. This is remarkably different from what was observed for DSBs induced outside of the nucleoli, which are resected in an ATM-dependent, but ATR-independent manner[4,5].

Nucleolar material exhibits liquid-demixing behavior[47]. Thereby, FC, DFC and GC within the nucleoli may represent coexisting immiscible liquid phases determined by differences in the biophysical properties of their constituents[48]. Ongoing transcription is an essential element of the maintenance of nucleolar structure because it may prevent coalescence of DFC-surrounded GCs to form larger centers[20]. Thus, when transcription is inhibited, GC and DFC start to fuse, which may induce a liquid-phase de-mixing process that results in the "floating" of the exposed GC to the nucleolar periphery. Our live-cell imaging data clearly support such a liquid-like behavior of the nucleoli and suggests liquid de-mixing processes as the driving force behind nucleolar segregation. In summary, we propose that nucleolar reorganization induced by rDNA breaks may be a consequence of inhibition of transcription resulting in enhanced ability of FC and DFC to coalesce.

The mechanism of Pol I inhibition in response to rDNA breaks is not yet clear. ATM-dependent inhibition of RNA polymerase II transcription in response to DSBs has been shown to involve local histone H2A ubiquitylation and the recruitment of a specific chromatin remodeler as well as the cohesion complex[10,11,14]. Posttranslational histone modifications also seem to be key events in Pol I inhibition because ATM-dependent H2B S14 phosphorylation by MST2 kinase is required for transcriptional repression in the nucleoli[19]. Direct modification of the Pol I transcription machinery by ATM and ATR may also be involved as several key Pol I transcription factors appear to be targeted by PIKKs[49]. However, the fact that transcriptional inhibition is dependent on both PIKK activity as well as downstream signaling events involving CHK1 and CHK2 kinases, it is likely that more than one effector molecule mediates DSB-induced transcriptional inhibition in the nucleoli.

Methods

Cell culture and drug treatments. All cell lines were grown in a sterile cell culture environment and routinely tested for mycoplasma contamination. 293FT (transformed human embryonal kidney cell line), U2OS (human osteosarcoma cell line), HeLa (cervical carcinoma cell line) and RPE-1 (immortalized human retinal pigmented cell line) cells were cultured in Dulbecco's modified Eagle medium (DMEM), supplemented with 10% fetal calf serum (FCS), 2 mM L-glutamine and penicillin–streptomycin antibiotics under standard cell culture conditions in a CO_2 incubator (37 °C; 5% CO_2). U2OS cell lines stably transfected with GFP-Treacle and tetracyclin repressor expression constructs were cultured in the presence of 200 μg/ml Zeocin and 10 μg/ml Blasticidin S (Invitrogen). U2OS cell lines stably transfected with NBS1-mNG expression constructs and with GFP-ATR expression constructs (gift from J. Lukas) were cultured in the presence of 1 μg/ml Puromycin. For pulsed EU incorporation, cells were incubated for 30 min in medium containing 1 mM EU. The Click-iT EU Alexa Fluor 594 Imaging Kit (Thermo Fisher Scientific) was used for EU detection. Unless stated otherwise, the following compounds were used at the indicated final concentrations: ATM inhibitor KU-55933 (5 μM; Selleckchem), ATR inhibitor VE-821 (5 μM; Selleckchem), ATR inhibitor HY-19323 (5 μM; MedChemExpress), CHK1/CHK2 inhibitor AZD7762 (300 nM; TargetMol), CHK1 inhibitor GDC-0575 (300 nM; MedChemExpress). X-ray irradiation of cells was performed with a YXLON.SMART 160E-1.5 device (150 kV, 6 mA; YXLON International SA) delivering 11.8 mGy per second. Soft X-rays were largely filtered out with a 3-mm aluminum filter.

Cloning and mutagenesis. The pcDNA4-TO-strep-HA-GFP-Treacle construct for GFP-Treacle expression, the deletion mutants F1-F5 and the point mutant S171A/T173A/T203A/T210A (STTT) has been described[21]. Mutant derivatives S1236A, S1227A/S1228A and S1227A/S1228A/S1236A of this construct were generated by site-directed mutagenesis using the QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies). Sequences of the mutagenesis primers were:

S1236A forward: GGAGAGACTGGGCTGGCGCCACATC
S1236A reverse: GATGTGGTGGCGCCAGCCCAGTCTCTCC
S1227A/S1228A forward:
GAGACCGCAGCAGCAGAGGCCGCCGAGGATGATGTGGTG
S1227A/S1228A reverse:
CACCACATCATCCTCGGCGGCCTCTGCTGCGGTCTC

pcDNA5/FRT/TO-NBS1-mNG were generated by PCR amplification of mNG from pmNeonGreen-C1 (Allele Biotechnology) and ligation into pcDNA5/FRT/TO-NBS1-WT, -R28A, -K160M and R28A/K160M, respectively[50]. Plasmids pIRES V5 I-Ppo1 and pIRES I V5 -Ppo1(H98A) for I-Ppo1 mRNA purification (gift from Brian McStay) were described[16]. pIRESneo2-TOPBP1-WT, -K154A/K155A (BRCT1), -K704A (BRCT5), -K1317A (BRCT7) were described[37,51,52]. The mutant derivative W1145R of this construct was generated by site-directed mutagenesis using the QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies). Sequences of the mutagenesis primers were:

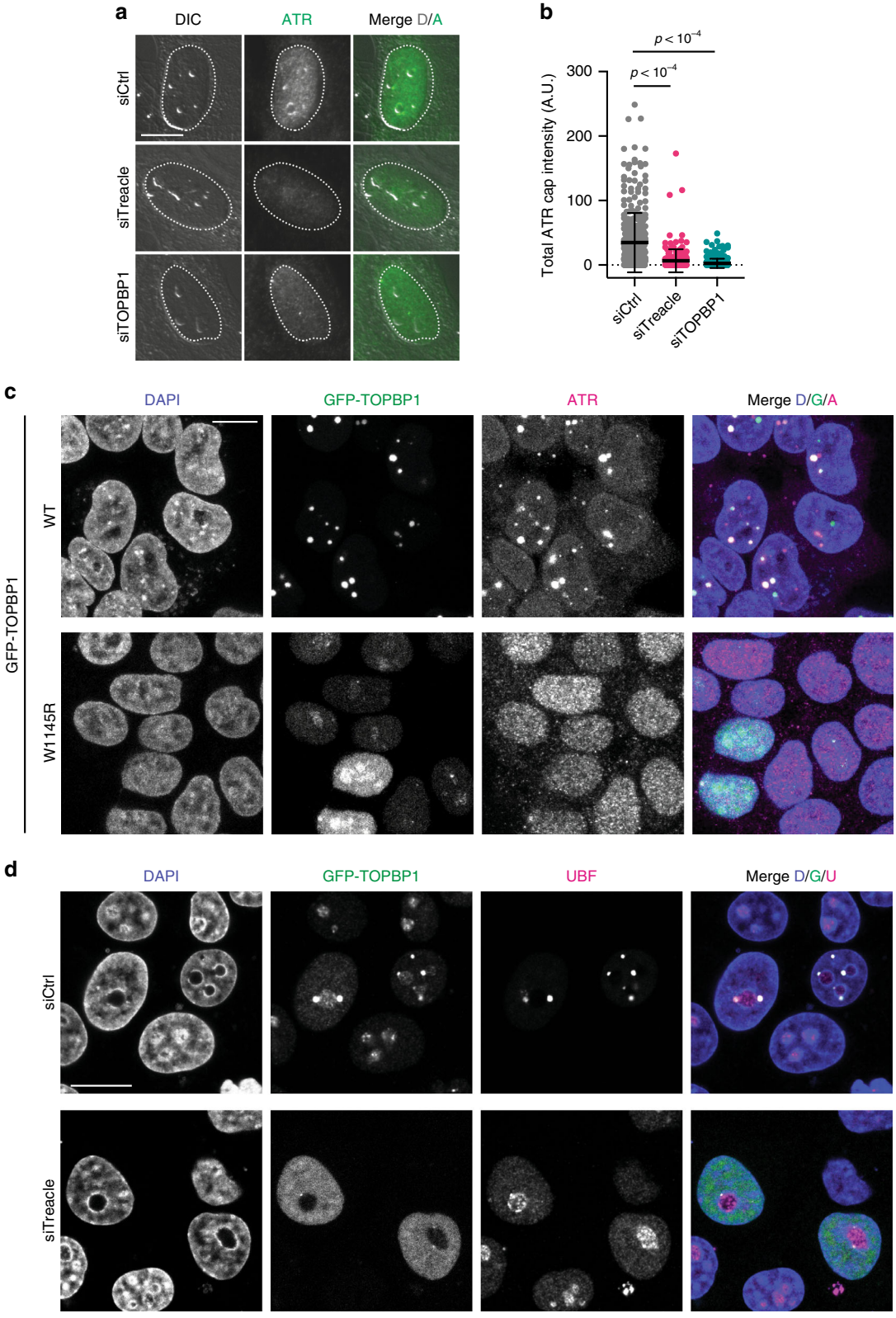


Fig. 7 Treacle- and TOPBP1-mediated nucleolar recruitment of ATR. a ATR localization in Treacle- and TOPBP1-depleted U2OS cells and control cells 2 h after I-Ppo1 transfection. **b** Quantification of the experiment in **a**, siCtrl ($n = 364$), siTreacle ($n = 214$) and siTOPBP1 ($n = 245$) treated cells. Statistical significance was calculated using a one-way analysis of variance and Sidak's multiple comparison test. Bars represent mean ± S.D. **c** ATR localization in 293-T cells overexpressing GFP-TOPBP1 wild type and W1145R AAD point mutant. **d** UBF localization in Hela cells overexpressing GFP-TOPBP1 and transfected with control siRNA and Treacle siRNA, respectively. All scale bars = 10 μm. Source data are provided as a Source Data file.

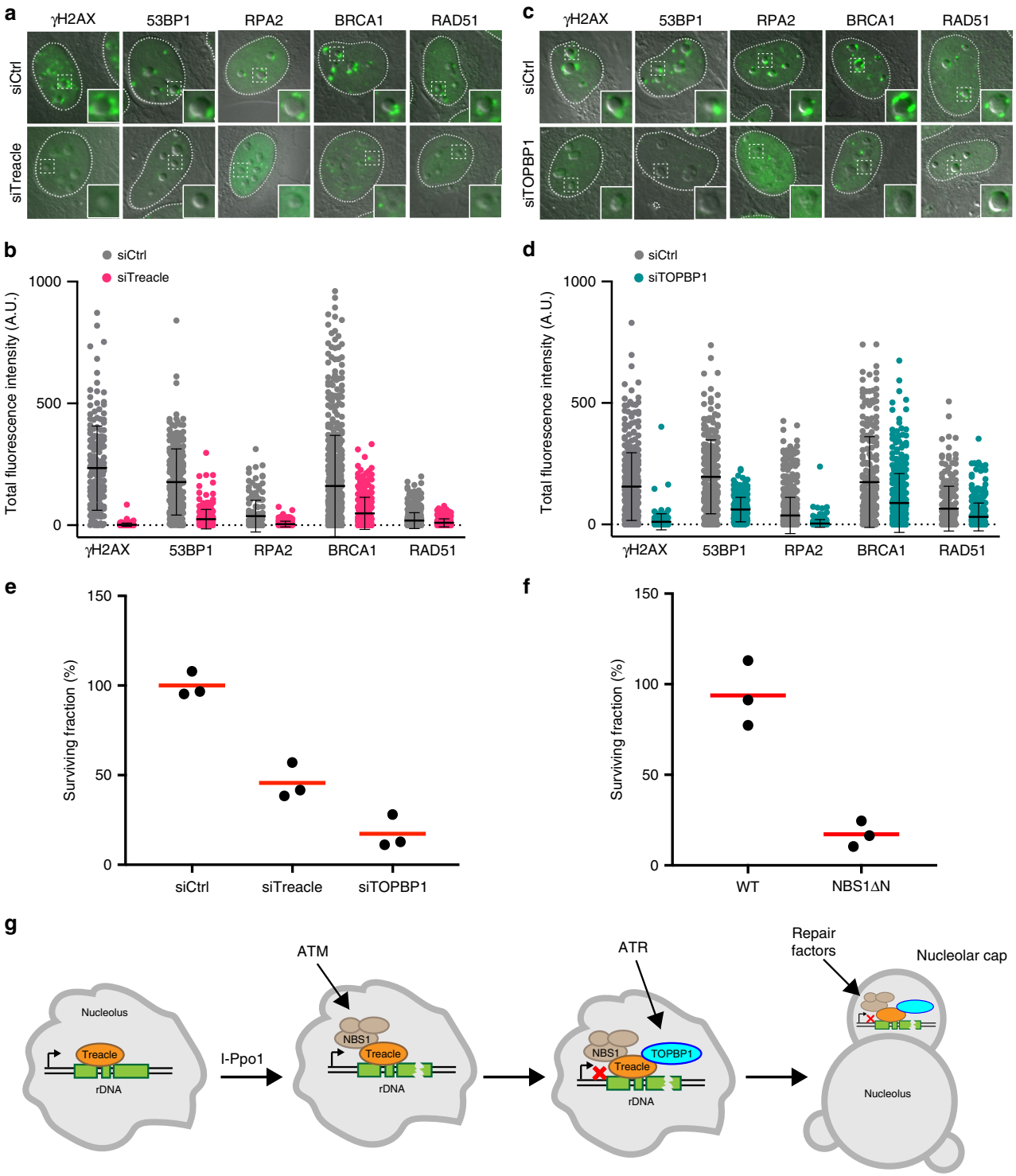


Fig. 8 Loss of Treacle and TOPBP1 impairs rDNA repair. a γH2AX, 53BP1, RPA2, RAD51 and BRCA1 localization in control depleted (siCtrl) and Treacle-depleted (siTreacle) U2OS cells 2 h after I-Ppo1 transfection. **b** Quantification of the experiment in **a**, γH2AX: siCtrl ($n = 177$) versus siTreacle ($n = 200$), 53BP1: siCtrl ($n = 261$) versus siTreacle ($n = 239$), RPA2: siCtrl ($n = 152$) versus siTreacle ($n = 168$) BRCA1: siCtrl ($n = 525$) versus siTreacle ($n = 246$) and RAD51: siCtrl ($n = 525$) versus siTreacle ($n = 246$) cells (all datapoints are shown, bar represents mean, error bars represent S.D.) **c** γH2AX, 53BP1, RPA2, RAD51 and BRCA1 localization in control depleted (siCtrl) and TOPBP1-depleted (siTOPBP1) U2OS cells 2 h after I-Ppo1 transfection. **d** Quantification of the experiment in **c**, γH2AX: siCtrl ($n = 403$) versus siTOPBP1 ($n = 205$), 53BP1: siCtrl ($n = 239$) versus siTOPBP1 ($n = 256$), RPA2: siCtrl ($n = 649$) versus siTOPBP1 ($n = 359$) BRCA1: siCtrl ($n = 228$) versus siTOPBP1 ($n = 348$) and RAD51: siCtrl ($n = 228$) versus siTOPBP1 ($n = 348$) cells (all datapoints are shown, bar represents mean, error bars represent S.D.). **e** Clonogenic surivival analysis of I-Ppo1-transfected control (siCtrl), Treacle-depleted (siTreacle) and TOPBP1-depleted (siTOPBP) U2OS cells (red bars represent the mean). **f** Clonogenic surivival analysis of I-Ppo1-transfected WT and NBS1DN U2OS cells (red bars represent the mean). **g** Model (see text for details). Source data are provided as a Source Data file.

W1145R forward:
GCCTTCCCAAAATGAACAGATCATTCGGGATGACCCTAC
W1145R reverse:
GTAGGGTCATCCCGAATGATCTGTTCATTTTGGGAAGGC.

Generation of stable cell lines. To isolate cell lines with low GFP-Treacle expression, U2OS cells were co-transfected with pcDNA6/TR (Invitrogen) and pcDNA4-TO-strep-HA-GFP-Treacle[21] constructs and stably transfected cells were selected in the presence of 400 µg/ml Zeocin and 5 µg/ml Blasticidin S (Invitrogen). Subsequently, clones with low levels of Doxycycline-independent "leaky" expression of GFP-Treacle were isolated. NBS1ΔN cells stably expressing NBS1-mNG were generated by co-transfection of pcDNA5/FRT/TO-NBS1-mNG with pBabe-Puro, followed by selection with 1 µg/ml Puromycin. To isolate cell lines with homogenous NBS1-mNG expressing, pooled colonies were sorted using a BD FACSAria II 4L cell sorter.

Generation and characterization of the NBS1ΔN cell line. For the generation of NBS1 knock-out cells, the sgRNA target sequence (50-GCGTTGAGTACGTTGT TGGA-30) was cloned into the PX330-U6-hSpCas9 vector and verified by sequencing. The targeting vector was then transfected into U2OS cells, and clonal cell lines were isolated by dilution in 96-well plates. Single clones were analyzed for NBS1 expression by immunofluorescence[50]. Several clones without detectable NBS1 expression were isolated and one of them was further characterized. Frame-shift-inducing insertion/deletion (indel) mutations in all of the three NBS1 alleles present in U2OS cells were verified by Sanger sequencing of a PCR-amplified genomic fragment that was cloned in the pCR II Blunt TOPO plasmid.

RNA transfection. The control siRNA (siCtrl) and siRNA against Treacle (siTreacle), TOPBP1 (siTOPBP1), ATR (siATR) and CtIP (siCtIP) were obtained from Microsynth AG. The sequences of the siRNAs were: siCtrl: UGGUUUACA UGUCGACUAA-dTdT, siTreacle: CCACCAUGGGUUGGAACUAAAUU-dTdT, siTOPBP1: ACAAAUACAUGGCUGGUUA-dTdT, siATR: CCUCCGUGAUGUU GCUUA-dTdT and siCtIP: GCTAAAACAGGAACGAATC-dTdT. Cells were grown in six-well plate 24 h prior the transfection. The siRNA transfection was done with 10 pM siRNA using Lipofectamine RNAiMAX (Invitrogen). Cells were split on coverslips 24 h after the transfection and harvested 72 h after the transfection.

For the generation of the in vitro transcribed I-Ppo1, we used the vectors pIRES I-Ppo1 wild type and H98A (a gift from Brian McStay[16]). Plasmids were linearized with NotI and transcribed in vitro using the MEGAscript T7 kit (Ambion) according to the manufacturer's instructions. The I-Ppo1 mRNA was then polyadenylated using the Poly(A) tailing kit (Ambin) according to the manufacturer's instructions. For I-Ppo1, mRNA transfection cells were grown on glass coverslips in six- or 24-well plate 24 h prior the I-Ppo1 transfection. The mRNA transfection was done using the Lipofectamine MessengerMax Reagent (Invitrogen) according to the manufacturer's instruction.

SDS-PAGE and western blotting. SDS-PAGE and western blotting were performed using BioRad 4–15% Mini-PROTEAN TGXTM Precast Gels. Additional SDS-PAGE and western blotting was done using Bolt 4–12% Bis-Tris Plus gels from Thermo Fisher Scientific. The following antibodies were used at the indicated dilutions: BLM (goat, Abcam, ab5446, 1/500), GFP (mouse, Roche, 11814460001, 1/5000), HA (mouse, Santa Cruz Biotechnology, sc-57592, 1/1000), NBS1 (rabbit, Abcam, ab32074, 1/2000), NBS1 (mouse, GeneTex, GTX70224, 1/1000), TOPBP1 (A300-111A, Bethyl Laboratories, 1/5000), TOPBP1 (rabbit, Abcam, Ab2402, 1/1500), ATM pS1981 (rabbit, Epitomics, 2152-1, 1/5000), ATM (rabbit, Calbiochem, PC-116, 2/250), Treacle (rabbit, Sigma, HPA038237, 1/500), RAD9 (rabbit, Abcam, ab70810, 1/3000), RAD50 (mouse, GeneTex, 13B3, GTX70288, 1/1000), MRE11 (mouse, Abcam, 12D7, ab214, 1/500), Tubulin (mouse, Sigma, DM1A; T6199, 1/2000), RPA-pS4/S8 (rabbit, Bethyl laboratories, A700-009, 1/1000), RPA2 (mouse, Abcam, ab2175, 1/10000), CHK1 pS317 (rabbit, Cell Signaling, 2344, 1/500), CHK1 pS345 (rabbit, Cell Signaling, 2348, 1/500) CHK1 (mouse, Santa Cruz Biotechnology, sc-8408, 1/1000), CHK2 pT68 (rabbit, Cell Signaling, 2661, 1/1000), CHK2 (rabbit, Cell Signaling, 2662, 1/1000).

Immunoprecipitations. For GFP-Trap immunoprecipitations, cells were washed twice in phosphate-buffered saline (PBS), and lysed in IP buffer (100 mM NaCl, 0.2% Igepal CA-630, 1 mM $MgCl_2$, 10% glycerol, 5 mM NaF, 50 mM Tris-HCl, pH 7.5), supplemented with cOmplete EDTA-free protease inhibitor cocktail and 25 U/ml Benzonase (Novagen). After nuclease digestion, NaCl and EDTA concentrations were adjusted to 200 mM and 2 mM, respectively, and lysates were cleared by centrifugation. Lysates were then incubated with 15 µl of GFP-Trap magnetic agarose beads (ChromoTek) with end-to-end mixing at 4 °C. Immunoglobulin–antigen complexes were washed five times with IP buffer before elution in 2× SDS sample buffer. For HA immunoprecipitations, cells were washed twice in PBS, and lysed in IP buffer (100 mM NaCl, 1 mM $MgCl_2$, 2 mM EDTA, 10% glycerol, 5 mM NaF, 0.2% Nonidet 40 (IP40), 50 mM Tris-HCl, pH 7.6), supplemented with cOmplete EDTA-free protease inhibitor cocktail and 25 U/ml Benzonase (Novagen). After nuclease digestion, lysates were cleared by centrifugation. Lysates were then incubated with 40 µl of anti HA-Agarose beads (Sigma) with end-to-end mixing at 4 °C. Immunoglobulin–antigen complexes were washed three times with IP buffer before elution in 2× SDS sample buffer.

GST pulldowns. GST fusion proteins were expressed in the competent *Escherichia coli* expression strain BL21(DE3)pLysS (Promega) followed by protein purification with glutathione–Sepharose beads. A total of 5 µg of purified GST-TOPBP1 BRCT fusion proteins were mixed with 33 µl of HeLa nuclear extract (7 mg/ml; Ipracell) and incubated for 1.5 h at 4 °C. Glutathione–Sepharose beads (GE Healthcare) were added and the suspension was incubated at 4 °C for 1.5 h on a rotating wheel. The beads were washed three times with wash buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM DTT, and 0.2% NP-40) and bound proteins were eluted by addition of 2× SDS sample buffer.

Immunofluorescence. Cells were grown on glass coverslips and fixed with either ice-cold methanol for 10 min or with 4% buffered formaldehyde for 15 min at room temperature, and subsequently permeabilized for 5 min in PBS containing 0.3% Triton X-100. Following 1 h of blocking in blocking buffer (10% FBS, 3% BSA in PBS), primary antibody incubations were performed at 4 °C overnight. Coverslips were washed three times with PBS and secondary antibody incubations were performed for 1 h at room temperature in the dark. After washing with PBS for three times, coverslips were mounted on glass microscopy slides with Vectrashield mounting medium containing 0.5 µg/ml 4′,6-diamidino-2-phenylindole dihydrochloride (DAPI). The following antibodies were used at the indicated dilutions: Treacle (rabbit, Sigma Life Science, HPA038237, 1/100), TOPBP1 (rabbit, Bethyl, A300-111A, 1/500), Nucleophosmin (mouse, Thermofisher, MA5-17141, 1/300), BRCA1 (mouse, sc-6454, Santa Cruz, 1/100), RAD51 (rabbit, Santa Cruz, sc-8349, 1/100), γH2AX (mouse, Millpore, 05-636, 1/500), MDC1 (mouse, Abcam, Ab50003, 1/300), Cyclin A (mouse, BD biosciences, 611269, 1/100), RPA2 (mouse, Abcam, Ab2175, 1/250), RPA2 pS4/S8 (rabbit, Bethyl A300-245, 1/400), 53BP1 (mouse, gift from T. Halazonetis, 1/20) NBS1 (rabbit, Novus, NB100-143, 1/200), MRE11 (mouse, GeneTex, GTX70212, 1/1000), UBF (mouse, Santa Cruz, sc-13121, 1/100), ATM (rabbit, Abcam, ab32420, 1/250), ATR (rabbit, Cell Signaling E1S3S, 1/100), V5 (mouse, Abcam, ab27671, 1/500).

Widefield and confocal microscopy. Widefieled image acquisition was done on a Zeiss AxioObserver Z1 widefield microscope, equipped with a Lumencor SpectraX illumination system and a Hamamatsu Orca Flash 4.0 V2, sCMOS, cooled fluorescence camera (16 bit, 2048 × 2048 pixel (4 MP), pixel size 6.5 µm). A 63×, 1.4-NA, i-plan apochromat oil-immersion objective was used. For optimal representation in figures, images were adjusted for brightness and exported as RGB TIF files using Fiji[53].

Confocal images were acquired with a Leica SP8 inverse confocal laser scanning microscope with a 63×, 1.4-NA Plan-Apochromat oil-immersion objective. The sequential scanning mode was used and the number of overexposed pixels was kept at a minimum. Fields were recorded at a resolution of 512 × 512 pixels, 8 bit depth or 1024 × 1024 pixels, 8 bit. For optimal representation in figures, maximum intensity projections were calculated and images were adjusted for brightness and exported as RGB TIF files using Fiji[53].

Image quantification. The xyz confocal datasets (z-stacks) were analyzed using IMARIS 9.2 (Bitplane). Caps/foci were segmented and counted by the integrated intensity-based spot detection tool and nucleoli were surface rendered to calculate nucleolar volume and sphericity.

Nucleolar caps and DNA damage foci were quantified using CellProfiler 3.0[51]. First, nuclei segmentation was performed by the intensity-based primary object detection module using the DAPI signal. For nucleolar caps and foci segmentation, the primary object detection module was used on the respective channels after applying a feature enhancement filter. Downstream data manipulation and graphical representation of the data were done using R 3.4.2 (R Development Core Team). CellProfiler pipelines and R scripts are available upon request.

Quantification of nucleolar EU incorporation was done with Fiji[53]. Nucleoli were segmented manually by using the DIC channel and a Wacom graphic tablet. Nucleoli were marked as regions of interest (ROI) and added to the ROI manager. Fluorescence intensities were then measured within the ROIs in the EU channel.

Time-lapse microscopy. GFP-Treacle- or NBS1-mNG-expressing cells were plated 24 h prior the imaging in the µ-Slide 8-well chamber coverslip (Ibidi). I-Ppo1 transfection was done using Lipofectamine MessengerMax Reagent (Invitrogen) according to the manufacturer's instruction. Cells were followed over time with a Zeiss AxioObserver Z1 widefield microscope equipped with a Definite Focus hardware autofocus control unit. Temperature and CO_2 levels were kept at 37 °C and 5% CO_2 for the duration of the experiment. A 63×, 1.4-NA, i-plan apochromat oil-immersion objective was used. Phase-contrast images and wavelength detection EGFP were acquired at 3 min intervals over a period of 2–3 h. Movies of individual cells were generated by applying the StackReg plugin in Fiji (Rigid body transformation) to compensate for cell movement. Time-lapse recordings were then animated and exported as GIF files or rendered to mp4 video files with Adobe Photoshop CC 2018.

Real-time PCR. Quantification of rRNA transcription by real-time PCR was essentially done as described previously[21]. Unless otherwise indicated, RNA

isolation and cDNA synthesis were carried out 6 h after I-Ppo1 transfection. Real-time PCR was performed using 10 ng of cDNA with SYBR Green I Master (Roche, Cat. no. 04707516001). Samples were analyzed on a LightCycler480 and quantified with LightCycler480 quantification software.

Clonogenic survival assay after I-Ppo1 transfection. Cells were treated either with control siRNA (siCtrl), siTreacle or siTOPBP1 and 48 h post transfection, cells were transfected with I-Ppo1 mRNA. To ensure the same conditions, cells were counted and plated in six-well plates 4 h after I-Ppo1 transfection. For each condition, 1000 cells per well were plated, in triplicates. Colonies were grown for 10 days, with medium change every 4 days. Plates were washed twice in PBS and incubated for 1 h in Methylene blue fixation/staining solution (80% EtOH, 1% Methylene blue powder) at room temperature. After rigorous washing with warm tap water, plates were left to dry overnight and colonies were counted manually using an eCount™ colony counter pen.

Mass spectrometry. Mass spectrometric experiments were performed on a nanoscale UHPLC system (EASY-nLC1000 from Proxeon Biosystems) connected to an Orbitrap Q-Exactive HF equipped with a nanoelectrospray source (Thermo Fisher Scientific). Peptides were auto-sampled and separated on a 15-cm analytical column (75 µm ID) in-house packed with 1.9 µm C18 beads using a 77-min gradient ranging from 5 to 40% acetonitrile in 0.1% formic acid at a flow rate of 250 nl per min. The mass spectrometer was operated in data-dependent acquisition mode and all samples were analyzed using a previously described "sensitive" acquisition method[52]. All raw data analysis was performed with MaxQuant software suite[54] version 1.5.3.30 supported by the Andromeda search engine[55]. For generation of the theoretical spectral library, a HUMAN.fasta database was extracted from UniProt. Cysteine carbamidomethylation was included as a fixed modification, whereas N-terminal acetylation, methionine oxidation and phosphorylation (S, T and Y), were included as variable modifications. Mass tolerance for precursors was set to 20 ppm in the first MS/MS search and 4.5 ppm in the main MS/MS search. Data were automatically filtered by posterior error probability to achieve a false discovery rate of < 1% (default), at the peptide-spectrum match and the site-specific levels. For enrichment and mass spectrometric analysis of Treacle phosphopeptides, 293T cells were transfected with pcDNA4-TO-strep-HA-GFP-Treacle and collected in high-salt RIPA buffer (50 mM Tris pH7.5, 400 mM NaCl, 1 mM EDTA, 1% Nonidet P-40, 0.1% Na-deoxycholate, 2 mM Na-orthovanadate, 5 mM NaF, 5 mM glycero-2-phosphate, protease inhibitors (Roche)). Proteins from cleared lysate were precipitated in acetone, dissolved in urea, reduced with dithiothreitol, alkylated with chloroacetamide30, and digested using Lys-C and modified sequencing-grade trypsin (Sigma). Protease digestion was terminated by addition of trifluoroacetic acid and peptides were purified using reversed-phase Sep-Pak C18 cartridges (Waters). Phosphopeptides were enriched on a TiO_2 column prior to mass spectrometry.

Statistical analysis. Graphs were generated with Prism 8 for macOS (GraphPad Software Inc., Version 8.1.2). Unless otherwise indicated, horizontal bars represent the mean values and error bars the standard deviation. In box plot graphs, boxes represent the 25–75 percentile range with median, and whiskers represent the 9–95 percentile range. Data points outside this range are shown individually. Statistical tests were applied as described in the Figure legends and were calculated with Prism 8. In all quantitative immunofluorescence studies, at least 100 random cells per group were scored.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[56] partner repository with the dataset identifier PXD015138. The source data underlying Figs. 1d, f–h, 2b, 3c, e, g, 4d, e, h, 5b, d, 6a, c, d, 7b and 8b–d and Supplementary Figs. 1c–e, 3b–d, 7c and 8a are provided as a Source Data file. All other relevant data, CellProfiler pipelines and R scripts are available from the authors upon request.

Received: 14 March 2019; Accepted: 10 December 2019;
Published online: 08 January 2020

References

1. Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. *Mol. Cell* **40**, 179–204 (2010).
2. Carson, C. T. et al. The Mre11 complex is required for ATM activation and the G2/M checkpoint. *EMBO J.* **22**, 6610–6620 (2003).
3. Uziel, T. et al. Requirement of the MRN complex for ATM activation by DNA damage. *EMBO J.* **22**, 5612–5621 (2003).
4. Jazayeri, A. et al. ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. *Nat. Cell Biol.* **8**, 37–45 (2006).
5. Myers, J. S. & Cortez, D. Rapid activation of ATR by ionizing radiation requires ATM and Mre11. *J. Biol. Chem.* **281**, 9346–9350 (2006).
6. Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542–1548 (2003).
7. Blackford, A. N. & Jackson, S. P. ATM, ATR, and DNA-PK: the trinity at the heart of the DNA damage response. *Mol. Cell* **66**, 801–817 (2017).
8. Kumagai, A., Lee, J., Yoo, H. Y. & Dunphy, W. G. TopBP1 activates the ATR-ATRIP complex. *Cell* **124**, 943–955 (2006).
9. Kruhlak, M. J. et al. The ATM repair pathway inhibits RNA polymerase I transcription in response to chromosome breaks. *Nature* **447**, 730–734 (2007).
10. Shanbhag, N. M., Rafalska-Metcalf, I. U., Balane-Bolivar, C., Janicki, S. M. & Greenberg, R. A. ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell* **141**, 970–981 (2010).
11. Kakarougkas, A. et al. Requirement for PBAF in transcriptional repression and repair at DNA breaks in actively transcribed regions of chromatin. *Mol. Cell* **55**, 723–732 (2014).
12. Iannelli, F. et al. A damaged genome's transcriptional landscape through multilayered expression profiling around in situ-mapped DNA double-strand breaks. *Nat. Commun.* **8**, 15656 (2017).
13. Ui, A., Nagaura, Y. & Yasui, A. Transcriptional elongation factor ENL phosphorylated by ATM recruits polycomb and switches off transcription for DSB repair. *Mol. Cell* **58**, 468–482 (2015).
14. Meisenberg, C. et al. Repression of transcription at DNA breaks requires cohesin throughout interphase and prevents genome instability. *Mol. Cell* **73**, 1–20 (2018).
15. Mangan, H., Gailín, M. Ó. & McStay, B. Integrating the genomic architecture of human nucleolar organizer regions with the biophysical properties of nucleoli. *FEBS J.* **30**, 1598 (2017).
16. van Sluis, M. & McStay, B. A localized nucleolar DNA damage response facilitates recruitment of the homology-directed repair machinery independent of cell cycle stage. *Genes Dev.* **29**, 1151–1163 (2015).
17. Harding, S. M., Boiarsky, J. A. & Greenberg, R. A. ATM dependent silencing links nucleolar chromatin reorganization to DNA damage recognition. *Cell Rep.* **13**, 251–259 (2015).
18. Warmerdam, D. O., van den Berg, J. & Medema, R. H. Breaks in the 45S rDNA lead to recombination- mediated loss of repeats. *Cell Rep.* **14**, 2519–2527 (2016).
19. Pefani, D. E., Tognoli, M. L., Pirincci Ercan, D., Gorgoulis, V. & O'Neill, E. MST2 kinase suppresses rDNA transcription in response to DNA damage by phosphorylating nucleolar histone H2B. *EMBO J.* **37**, e98760–16 (2018).
20. van Sluis, M. & McStay, B. Nucleolar reorganization in response to rDNA damage. *Curr. Opin. Cell Biol.* **46**, 81–86 (2017).
21. Larsen, D. H. et al. The NBS1-Treacle complex controls ribosomal RNA transcription in response to DNA damage. *Nat. Cell Biol.* **16**, 792–803 (2014).
22. Ciccia, A. et al. Treacher Collins syndrome TCOF1 protein cooperates with NBS1 in the DNA damage response. *Proc. Natl Acad. Sci. USA* **111**, 18631–18636 (2014).
23. Trainor, P. A., Dixon, J. & Dixon, M. J. Treacher Collins syndrome: etiology, pathogenesis and prevention. *Eur. J. Hum. Genet.* **17**, 275–283 (2009).
24. Berkovich, E., Monnat, R. J. & Kastan, M. B. Roles of ATM and NBS1 in chromatin structure modulation and DNA double-strand break repair. *Nat. Cell Biol.* **9**, 683–690 (2007).
25. Kim, J. H. et al. The Mre11-Nbs1 interface is essential for viability and tumor suppression. *Cell Rep.* **18**, 496–507 (2017).
26. Hari, F. J., Spycher, C., Jungmichel, S., Pavic, L. & Stucki, M. A divalent FHA/BRCT-binding mechanism couples the MRE11-RAD50-NBS1 complex to damaged chromatin. *EMBO Rep.* **11**, 387–392 (2010).
27. Adams, K. E., Medhurst, A. L., Dart, D. A. & Lakin, N. D. Recruitment of ATR to sites of ionising radiation-induced DNA damage requires ATM and components of the MRN protein complex. *Oncogene* **25**, 3894–3904 (2006).
28. Sokka, M., Rilla, K., Miinalainen, I., Pospiech, H. & Syvaoja, J. E. High levels of TopBP1 induce ATR-dependent shut-down of rRNA transcription and nucleolar segregation. *Nucleic Acids Res.* **43**, 4975–4989 (2015).
29. Morishima, K.-I. et al. TopBP1 associates with NBS1 and is involved in homologous recombination repair. *Biochem. Biophys. Res. Commun.* **362**, 872–879 (2007).
30. Duursma, A. M., Driscoll, R., Elias, J. E. & Cimprich, K. A. A role for the MRN complex in ATR activation via TOPBP1 recruitment. *Mol. Cell* **50**, 116–122 (2013).
31. Day, M. et al. BRCT domains of the DNA damage checkpoint proteins TOPBP1/Rad4 display distinct specificities for phosphopeptide ligands. *Elife* **7**, 94 (2018).
32. Huo, Y.-G., Bai, L., Xu, M. & Jiang, T. Crystal structure of the N-terminal region of human Topoisomerase III^2 binding protein 1. *Biochem. Biophys. Res. Commun.* **401**, 401–405 (2010).
33. Rappas, M., Oliver, A. W. & Pearl, L. H. Structure and function of the Rad9-binding region of the DNA-damage checkpoint adaptor TopBP1. *Nucleic Acids Res.* **39**, 313–324 (2011).

34. Wardlaw, C. P., Carr, A. M. & Oliver, A. W. TopBP1: A BRCT-scaffold protein functioning in multiple cellular pathways. *DNA Repair* **22**, 165–174 (2014).
35. Leung, C. C. Y. & Glover, J. N. M. BRCT domains: easy as one, two, three. *Cell Cycle* **10**, 2461–2470 (2011).
36. Delacroix, S., Wagner, J. M., Kobayashi, M., Yamamoto, K. I. & Karnitz, L. M. The Rad9-Hus1-Rad1 (9-1-1) clamp activates checkpoint signaling via TopBP1. *Genes Dev.* **21**, 1472–1477 (2007).
37. Blackford, A. N. et al. TopBP1 interacts with BLM to maintain genome stability but is dispensable for preventing BLM degradation. *Mol. Cell* **57**, 1133–1141 (2015).
38. Sun, L. et al. Structural insight into BLM recognition by TopBP1. *Structure* **25**, 1582–1588.e3 (2017).
39. Cimprich, K. A. & Cortez, D. ATR: an essential regulator of genome integrity. *Nat. Rev. Mol. Cell Biol.* **9**, 616–627 (2008).
40. Hustedt, N. & Durocher, D. The control of DNA repair by the cell cycle. *Nat. Cell Biol.* **19**, 1–9 (2017).
41. Gamper, A. M. et al. ATR kinase activation in G1 phase facilitates the repair of ionizing radiation-induced DNA damage. *Nucleic Acids Res.* **41**, 10334–10344 (2013).
42. Bøe, C. A., Håland, T. W., Boye, E., Syljuåsen, R. G. & Grallert, B. A novel role for ATR/Rad3 in G1 phase. *Sci. Rep.* **8**, 6880 (2018).
43. Ball, H. L., Myers, J. S. & Cortez, D. ATRIP binding to replication protein A-single-stranded DNA promotes ATR-ATRIP localization but is dispensable for Chk1 phosphorylation. *Mol. Biol. Cell* **16**, 2372–2381 (2005).
44. Kim, S.-M., Kumagai, A., Lee, J. & Dunphy, W. G. Phosphorylation of Chk1 by ATM- and Rad3-related (ATR) in Xenopus egg extracts requires binding of ATRIP to ATR but not the stable DNA-binding or coiled-coil domains of ATRIP. *J. Biol. Chem.* **280**, 38355–38364 (2005).
45. Kousholt, A. N. et al. CtIP-dependent DNA resection is required for DNA damage checkpoint maintenance but not initiation. *J. Cell Biol.* **197**, 869–876 (2012).
46. Boos, D. et al. Regulation of DNA replication through Sld3-Dpb11 interaction is conserved from yeast to humans. *Curr. Biol.* **21**, 1152–1157 (2011).
47. Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc. Natl Acad. Sci. USA* **108**, 4334–4339 (2011).
48. Feric, M. et al. Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686–1697 (2016).
49. Matsuoka, S. et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160–1166 (2007).
50. Anand, R. et al. NBS1 promotes the endonuclease activity of the MRE11-RAD50 complex by sensing CtIP phosphorylation. *EMBO J.* **38**, e101005–e101016 (2019).
51. Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
52. Kelstrup, C. D., Young, C., Lavallee, R., Nielsen, M. L. & Olsen, J. V. Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. *J. Proteome Res.* **11**, 3487–3497 (2012).
53. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
54. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
55. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
56. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2018).

Acknowledgements

We thank Brian McStay, Jiri Lukas and Thanos Halazonetis for providing valuable reagents, Dafni Pefani for providing technical advice and Dorthe Larsen for sharing unpublished observations. Imaging was performed with equipment maintained by the Center for Microscopy and Image Analysis, University of Zurich. Cell sorting was carried out by the Flow Cytometry Core Facility, University of Zurich. The Blackford lab is supported by a Cancer Research UK Career Development Fellowship (C29215/A20772) to A.N.B. and an MRC WIMM DPhil Studentship (MR/N013468/1) to A.-M.K.S. and A.N.B. The Nielson lab is supported by the Novo Nordisk Foundation Center for Protein Research, the Novo Nordisk Foundation (grant agreement numbers NNF14CC0001 and NNF13OC0006477), and the Danish Council of Independent Research (grant agreement numbers 4002-00051, 4183-00322A and 8020-00220B). The Stucki lab is supported by the Forschungskredit Candoc program of the University of Zurich (FK-18-024) to P.-A.L., a project grant from the Swiss National Foundation (31003A_163141), the Promedica foundation and a grant from the Helmut Horten Foundation.

Author contributions

The project was conceived and supervised by M.S. Biochemical and cell biological experiments were carried out and data were analyzed by C.M., I.-E.S., A.R., P.-A.L., A.-M.K.S., A.N.B. and M.S. Experimental tools and reagents were generated and characterized by D.Z., C.L., A.Ja., A.Je. and K.K. D.F. supervised D.Z. and A.Je. Mass spectrometry data were contributed by S.J., S.C.L. and M.L.N. The paper was written by M.S. with contributions from other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-13981-x.

Correspondence and requests for materials should be addressed to M.S.

Peer review information *Nature Communications* thanks Eric O'Neill and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020