Assessment of the functional impact of amino acid variants in DNA damage response proteins using orthologue complementation



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

Matthew Fawkes

Department Of Oncology
MRC Weatherall Institute of Molecular Medicine
St Peter's College

Hilary Term 2023

Supervisor: Dr. Andrew Blackford

Acknowledgements

Firstly I would like to thank my supervisor Dr. Andrew Blackford for his supervision during my DPhil. I have never encountered a more driven, hard-working and knowledgeable scientist, and I wish you every success for the future.

Secondly, I would like to thank members of the Blackford Lab for their friendship and support during my DPhil, as well as the contributions they made to the work performed. I am especially grateful to Dr. Samuel Jones, who found the time to supervise and guide me throughout my DPhil. You were never afraid of a reality check, even in my first lab meeting, where my delusions about it 'working first time' were met with laughter and a sigh. I would also like to thank Dr. Ann-Marie Shorrocks, Dr. Kaima Tsukada, Robert Horsley and Julius Bannister for their friendship and support during my DPhil. It has been a pleasure to work alongside you all.

Outside of the Blackford Lab I would like to thank members of the McHugh Lab for their assistance during my DPhil. A special mention goes out to Sanja Brolih for her endless provision of pastries and excellent work as Social Secretary for the McHugh, Blackford and Goriely Labs. I would also like to thank Malitha Ratnaweera for his computational expertise and frequent chats about cricket. I also want to thank the WIMM FACS facility for their time and patience with my cell sorts as well as Dr. Ryan Beveridge (previously of the WIMM virus screening facility) for his guidance on lentivirus production.

I would also like to thank my department and my funding body (Cancer Research UK), for the provision and funding of this DPhil. I would also like to thank St Peter's College for their support as well as the MCR and associated members, especially my old housemates Joe, PK and Will.

I would especially like to thank my partner Alex for her continued support throughout my DPhil, and for generally putting up with me. I will always appreciate the times you asked ‘how are your cells doing’, as I know you only wanted me to succeed and provide support if I needed it. You inspire me every day with your hard-work and passion for your research and I have no doubts that you will achieve everything you want to.

Lastly and most importantly I would like to thank all members of my family. Mum and Dad, I will always be thankful your continued support for whatever I choose to do in life. I would especially like to thank my grandad, Ron, to whom I dedicate this thesis. I know you were always fascinated by evolution, and upon reading your copy of The Selfish Gene, I was also inspired to learn more about this amazing feature of the natural world. I can only hope that the work I have performed here would have made you proud.

List Of Abbreviations

5-FU	5-fluoruracil
5-mC	5-methyl cytosine
6-4PPs	6-4 pyrimidine photoproducts
6-TG	6-thioguanine
8-oxo-G	8-oxoguanine
AA	Amino acid
ACMG	American College of Medical Genetics and Genomics
AID	Activation-induced deaminase
AML	Acute myeloid leukaemia
AMP	Association for Molecular Pathology
BER	Base excision repair
BPDE	Benzo(a)pyrene diol epoxide
BTR	BLM-TOP3A-RMI1-RMI2 complex
CHO	Chinese hamster ovary
CIMRA	Complete *in vitro* MMR Assay
ClinGen	Clinical Genome Resource Consortium
CMG	MCM2-7-Cdc45-GINS complex
COSMIC	Catalogue of Somatic Mutations In Cancer
CPDs	Cyclobutane pyrimidine dimers
CRC	Colorectal cancer
CSR	Class-switch recombination
DDC	Duplication-degeneration-complementation (model)
DDR	DNA damage response
dHJ	Double Holliday junction
DNA-PKcs	DNA-PK catalytic subunit
DPD	Divergence prior to duplication (model)
DSB(s)	Double-strand break(s)
dsDNA	Double-stranded DNA
dTNPs	Deoxynucleoside triphosphates
EAC	Escape from adaptive conflict (model)
ENIGMA	Evidence-based Network for the Interpretation of Germline Mutation Alleles
ExAC	Exome Aggregation Consortium
FA	Fanconi anaemia
FACS	Fluorescence activated cell sorting
GG-NER	Global-genomic nucleotide excision repair
gnomAD	Genome Aggregation Database
GWAS	Genome-wide association studies
HGMD	Human Gene Mutation Database
HNPCC	Hereditary nonpolyposis colorectal cancer
HR	Homologous recombination
HU	Hydroxyurea
ICG-HNPCC	International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer
ICL(s)	Inter-strand crosslink(s)
IDLs	Insertion/deletion loops

INSiGHT	International Society for Gastrointestinal Hereditary Tumours
IP	Immunoprecipitation
IR	Ionising radiation
LCPG	Leeds Castle Polyposis Group
LET	Linear energy transfer
LOF	Loss-of-function
LS	Lynch syndrome
^{Me}6-TG	6-methylthioguanine
MMC	Mitomycin C
MMEJ	Microhomology-mediated end-joining
MMR	Mismatch repair
MMS	Methyl methanesulfonate
MNNG	Methylnitronitrosoguanidine
MRN	MRE11-RAD50-NBS1 complex
MSI	Microsatellite instability
NCBI	Nation Centre for Biotechnology Information
NCCN	National Comprehensive Cancer Network
NER	Nucleotide excision repair
NGS	Next-generation sequencing
NHEJ	Non-homologous end-joining
OMIM	Online Mendelian Inheritance in Man
PAR	Poly(ADP-ribose)
PARP	Poly(ADP-ribose) polymerase
PARPi	PARP-inhibitors
PIKKs	Phosphatidylinositol 3-kinase like protein kinases
PTM(s)	Post-translational modification(s)
RNA-Pol II	RNA polymerase II
rNTPs	Ribonucleoside triphosphates
ROS	Reactive oxygen species
SCC	Squamous cell carcinoma
SCEs	Sister-chromatid exchanges
SCP	Small colony phenotype
SDSA	Synthesis-dependent strand annealing
SIFT	Sorting Intolerant From Tolerant
SNP(s)	Single nucleotide polymorphism(s)
SSA	Single-strand annealing
SSB(s)	Single-strand break(s)
ssDNA	Single-stranded DNA
TCGA	The Cancer Genome Atlas
TC-NER	Transcription-coupled nucleotide excision repair
UBC	Ubiquitin conjugating
UV	Ultraviolet (radiation)
VUS	Variant of Uncertain Significance
XP	Xeroderma pigmentosum
Y2H	Yeast-2-hybrid

Table of Contents

Acknowledgements ..... *i*

List Of Abbreviations ..... *iii*

List of Figures ..... *ix*

List of Tables ..... *xi*

1 Abstract *1*

2 Introduction *3*

2.1 Sources of DNA damage and the DNA damage response **3**

2.1.1 The DNA damage response 3
2.1.2 Endogenous sources of DNA damage 4
2.1.3 Exogenous sources of DNA damage 5
2.1.4 DNA damage detection, signalling and checkpoint activation 8
2.1.5 Single-strand break repair 11
2.1.6 Mismatch repair 11
2.1.7 Base excision repair 15
2.1.8 Nucleotide excision repair 15
2.1.9 Non-homologous end-joining 18
2.1.10 Homologous recombination 20
2.1.11 Additional DSB repair pathways 23
2.1.12 DSB repair pathway choice 23
2.1.13 Responding to DSBs in mitosis 24
2.1.14 Inter-strand crosslink repair 25

2.2 The DNA damage response in disease **28**

2.2.1 Immunodeficiency 28
2.2.2 Infertility 29
2.2.3 Premature ageing 30
2.2.4 Cancer 31
2.2.5 DDR syndromes with cancer predisposition 31
2.2.6 Targeting the DDR in cancer 34

2.3 Identifying, cataloguing and classifying variants **39**

2.3.1 Amino acid variants 39
2.3.2 Identifying amino acid variants 39
2.3.3 Cataloguing amino acid variants 43
2.3.4 Classifying amino acid variants 44
2.3.5 The importance of identifying amino acid variants and their functional impact 50

2.4 Methods for assessing the functional impact of variants **53**

2.4.1 In vitro assays 53
2.4.2 Yeast-based assays 54
2.4.3 Survival in response to drugs 55
2.4.4 Mechanism-specific assays 56
2.4.5 Protein stability and cellular localisation 57
2.4.6 Protein-protein interactions 58
2.4.7 High-throughput genome editing-based approaches 59
2.4.8 Computational approaches 61
2.4.9 Orthologue complementation 65

2.5 Aims and Objectives 66

3 Methods 69

3.1 Generation of sequence alignments 69

3.2 SDS-PAGE and western blotting 70

3.3 Plasmids and cloning 70

3.4 Cell lines and cell culture conditions 73

3.5 Immunoprecipitations 74

3.6 Lentiviral assembly and stable cell line generation 74

3.7 Colony survival assays 75

3.8 Variant and PTM database searching 76

4 Orthologue Complementation of XPA 77

4.1 Introduction 77

4.2 Aims and objectives 82

4.3 Results 83

4.3.1 Selection of vertebrate XPA orthologues for complementation of human cells 83

4.3.2 Vertebrate XPA orthologues can bind to XPC 85

4.3.3 Vertebrate XPA orthologues can restore protection against UV in human cells 86

4.3.4 Invertebrate XPA orthologues cannot complement human cells 87

4.3.5 A sequence alignment of complementing XPA orthologues contains known benign variants and does not contain known pathogenic variants 92

4.3.6 A sequence alignment of complementing XPA orthologues reveals functional amino acid variants in the human protein 94

4.3.7 A sequence alignment of complementing XPA orthologues reveals non-essential PTM sites 99

4.3.8 Comparison to bioinformatic predictions of variant effects 102

4.4 Discussion 103

4.4.1 Summary of the work presented in this chapter 103

4.4.2 A discussion of the findings presented in this chapter 103

4.5 Future experiments based on the work presented in this chapter 111

5 Orthologue complementation of XRCC4 113

5.1 Introduction 113

5.2 Aims and objectives 119

5.3 Results 120

5.3.1 Selection of vertebrate XRCC4 orthologues for complementation of human cells 120

5.3.2 Vertebrate XRCC4 orthologues can all bind to LIG4 122

5.3.3 Fish XRCC4 cannot restore protection against IR in human cells 123

5.3.4 A sequence alignment of complementing XRCC4 orthologues contains known benign variants but also a known pathogenic variant 126

5.3.5 A sequence alignment of complementing XRCC4 orthologues does not contain ClinVar XRCC4 VUS 130

5.3.6 A Sequence alignment of complementing XRCC4 orthologues reveals multiple non-essential PTM sites 130

5.3.7 Comparison to bioinformatic predictions of variant effects 134

5.4 Discussion 135
5.4.1 A summary of the work presented in this chapter 135
5.4.2 A discussion of the work presented in this chapter 136
5.5 Future work based on the work presented in this chapter 141
6 Orthologue complementation of MLH1 143
6.1 Introduction 143
6.2 Aims and objectives 150
6.3 Results 151
6.3.1 Selection of vertebrate MLH1 orthologues 151
6.3.2 MLH1 orthologues display variable affinity for known binding partners in human cells 153
6.3.3 MLH1 orthologues can restore PMS2 stability in 293T cells 155
6.3.4 Frog MLH1 cannot restore sensitivity to 6-TG but restores PMS2 expression in HAP1 ΔMLH1 cells 157
6.3.5 Frog MLH1 cannot restore sensitivity to 6-TG or PMS2 stability in HCT116 cells 160
6.3.6 A sequence alignment of complementing MLH1 orthologues does not contain known benign or pathogenic MLH1 variants 164
6.3.7 Selection of variants for assessment based on sequence alignments and database searches 174
6.3.8 MLH1 constructs containing variants present in complementing MLH1 orthologues can restore 6-TG sensitivity 186
6.3.9 MLH1 constructs containing variants present in complementing MLH1 orthologues can restore PMS2 stability 190
6.3.10 Prediction of requirement of PTMs reported in PhosphoSitePlus 193
6.4 Discussion 195
6.4.1 Summary of the work presented in this chapter 195
6.4.2 A discussion of the findings presented in this chapter 196
6.4.3 Future experiments based on the work presented in this chapter 206
7 Orthologue complementation of FANCL 208
7.1 Introduction 208
7.2 Aims and Objectives 215
7.3 Results 216
7.3.1 Selection of vertebrate FANCL orthologues for complementation of human cells 216
7.3.2 FANCL orthologues show decreasing affinity to the human FA core complex 218
7.3.3 FANCL orthologues show reduced FANCD2 ubiquitination activity 220
7.3.4 FANCL orthologues do not restore protection against cisplatin to WT levels 222
7.4 Discussion 224
7.4.1 Summary of the work presented in this chapter 224
7.4.2 A discussion of the work presented in this chapter 224
7.4.3 Future work based on the findings presented in this chapter 228
8 Discussion 229
8.1 Summary of the work presented in this thesis 229
8.2 An assessment of orthologue complementation 230
8.2.1 Strengths of orthologue complementation 230
8.2.2 Weaknesses of orthologue complementation 232
8.3 Future work 234

9 References 235

10 Appendices 298

10.1 FACS plots from vertebrate XPA orthologue UV survival assay 298

10.2 FACS plots from invertebrate XPA orthologue UV survival assay 301

10.3 FACS plots from XPA variants UV survival assay 303

10.4 FACS plots from XRCC4 vertebrate orthologue IR survival assay 306

10.5 FACS plots from XRCC4 6A variant IR survival assay 308

10.6 FACS plots from HAP1 ΔMLH1 MLH1 orthologue 6-TG survival assay 309

10.7 FACS plots from HCT116 MLH1 orthologue 6-TG survival assay 311

10.8 FACS plots from MLH1 variant constructs 6-TG survival assay 313

10.9 Sequence alignment of vertebrate and invertebrate MLH1 orthologues highlighting A681 316

10.10 Structural diagram of MLH1 with focus on alpha helix containing A681 317

10.11 FACS plots showing GFP % of stable pools of FANCL orthologue cell lines 318

List of Figures

Figure 2-1: The cellular response to DNA damage 3
Figure 2-2: Types of DNA lesion and their associated repair pathway(s) 10
Figure 2-3: The eukaryotic post-replicative MMR pathway 14
Figure 2-4: The nucleotide excision repair pathway 17
Figure 2-5: The non-homologous end-joining pathway 19
Figure 2-6: The homologous recombination pathway 22
Figure 2-7: The inter-strand crosslink repair pathway 27
Figure 2-8: Orthologue complementation of human cells as a method to identify functional amino acid variants in the human protein 68
Figure 4-1: Sequence alignment of vertebrate XPA orthologues used in this study with conservation shading 84
Figure 4-2: Vertebrate XPA orthologues can bind to human XPC 85
Figure 4-3: Vertebrate XPA orthologues can restore protection against UV in human cells 86
Figure 4-4: Sequence alignment of vertebrate and invertebrate XPA orthologues used in this study with conservation shading 89
Figure 4-5: Invertebrate XPA orthologues cannot complement human cells 91
Figure 4-6: Variants present in complementing orthologues do not perturb XPA function in response to UV radiation 98
Figure 4-7: Variation to non-modifiable residues in complementing orthologues reveals non-essential PTM sites 101
Figure 5-1: Sequence alignment of vertebrate XRCC4 orthologues used in this study with conservation shading 121
Figure 5-2: Vertebrate XRCC4 orthologues can all bind to LIG4 122
Figure 5-3: Fish XRCC4 cannot restore protection against IR in human cells 125
Figure 5-4: Sequence alignment of vertebrate XRCC4 orthologues which complement human cells 127
Figure 5-5: Sequence alignments of complementing XRCC4 orthologues reveal non-essential PTM sites in XRCC4 133
Figure 6-1: Sequence alignment of vertebrate MLH1 orthologues used in this study with conservation shading 152
Figure 6-2: MLH1 orthologues show varying affinity for known binding partners in human cells 154
Figure 6-3: MLH1 orthologues can restore PMS2 stability in 293T cells 156
Figure 6-4: Frog MLH1 cannot restore sensitivity to 6-TG in HAP1 ΔMLH1 cells 158
Figure 6-5: Vertebrate GFP-MLH1 orthologues restore PMS2 stability in HAP1 ΔMLH1 cells 159
Figure 6-6: Frog MLH1 cannot restore sensitivity to 6-TG in HCT116 cells 161
Figure 6-7: Frog MLH1 partially restores PMS2 expression in HCT116 cells 162
Figure 6-8: Sequence alignment of vertebrate MLH1 orthologues able to complement human cells with conservation shading 163
Figure 6-9: Schematic showing variants introduced into human MLH1 187
Figure 6-10: MLH1 constructs containing variants present in complementing orthologues can restore 6-TG sensitivity 188
Figure 6-11: Statistical analysis of MLH1 variant effects in response to 4μM 6-TG 189
Figure 6-12: MLH1 constructs containing variants present in complementing orthologues can restore PMS2 stability 191
Figure 6-13: Immunoprecipitation of PMS2 with A681S and A681T MLH1 192
Figure 7-1: Sequence alignment of vertebrate FANCL orthologues used in this study with conservation shading 217
Figure 7-2: FANCL orthologues show decreasing affinity to the human FA core complex with increasing sequence divergence to human 219

Figure 7-3: FANCL orthologues show reduced or absent FANCD2 ubiquitination ability 221
Figure 7-4: FANCL orthologues do not restore protection against cisplatin in human cells..... 223
Figure 10-1: Representative FACS plots from HAP1 WT and HAP1 ΔXPA cells either untransfected or transfected with GFP and GFP-tagged vertebrate XPA orthologues..... 300
Figure 10-2: Representative FACS plots from HAP1 WT and HAP1 ΔXPA cells transfected with GFP and GFP-tagged vertebrate and invertebrate XPA orthologues..... 302
Figure 10-3: Representative FACS plots from HAP1 WT and HAP1 ΔXPA cells either untransfected or transfected with GFP and GFP-tagged XPA variant constructs. 305
Figure 10-4: Representative FACS plots from HAP1 WT and HAP1 ΔXRCC4 cells either untransfected or transfected with GFP and GFP-tagged vertebrate XRCC4 orthologues 307
Figure 10-5: Representative FACS plots from HAP1 WT and HAP1 ΔXRCC4 cells transfected with GFP and GFP-tagged XRCC4 WT or variant constructs..... 308
Figure 10-6: Representative FACS plots from HAP1 WT and HAP1 ΔMLH1 cells constitutively expressing GFP or GFP-tagged MLH1 orthologues used for 6-TG sensitivity assays and PMS2 stability blots. 310
Figure 10-7: Representative FACS plots from HCT116 cells constitutively expressing GFP or GFP-tagged MLH1 orthologues used for 6-TG sensitivity assays and PMS2 stability blots. 312
Figure 10-8: Representative FACS plots from HAP1 ΔMLH1 cells constitutively expressing GFP or GFP-tagged MLH1 variant constructs used for 6-TG sensitivity assays and PMS2 stability/interaction blots. 316
Figure 10-9: Sequence alignment of vertebrate and invertebrate MLH1 orthologues with A681 highlighted 316
Figure 10-10: Structural prediction diagram for MLH1 with focus on alpha helix containing A681 317
Figure 10-11: FACS plots from analysis of HAP1 ΔFANCL cells transduced with lentiviruses encoding either GFP or mClover3-tagged FANCL orthologues 319

List of Tables

Table 2-1: Sources of endogenous and exogenous DNA mutagens and their associated lesions. 8
Table 2-2: Description of the ACMG/AMP criteria used for variant interpretation 47
Table 2-3: ACMG/AMP scoring system used for variant classification 48
Table 2-4: Description of scoring system for ClinVar submissions 49
Table 2-5: % of variants in DDR genes which are VUS on the ClinVar database as of Jan 2023 64
Table 3-1: Accession numbers of orthologues used in this study 69
Table 3-2: Names and sequences of primers used in this study 73
Table 4-1: Accession numbers and % identity to human orthologue for vertebrate XPA orthologues used 84
Table 4-2: Accession numbers and % identity to human orthologue for vertebrate and invertebrate XPA orthologues used 89
Table 4-3: Results of searching the sequence alignment of complementing orthologues for homozygous gnomAD XPA variants 93
Table 4-4: Results of searching the sequence alignment of complementing orthologues for benign XPA variants on ClinVar 93
Table 4-5: Results of searching the sequence alignment of complementing orthologues for pathogenic XPA variants on ClinVar 93
Table 4-6: Results from searching the alignment of complementing XPA orthologues for ClinVar XPA variants not classified as either ‘benign’ or ‘pathogenic’ 95
Table 4-7: Amino acid positions of XPA exons with amino acids encoded across exon boundaries 96
Table 4-8: Results from searching the sequence alignment of complementing XPA orthologues for variation at PTM sites in XPA reported on the PhosphoSitePlus database 99
Table 4-9: Computational predictions for XPA variants tested 102
Table 5-1: Accession numbers and % identity to human orthologue for vertebrate XRCC4 orthologues used 121
Table 5-2: Results of searching the sequence alignment of complementing XRCC4 orthologues for homozygous XRCC4 variants on gnomAD 128
Table 5-3: Results of searching the sequence alignment of complementing orthologues XRCC4 for benign XRCC4 variants on ClinVar 128
Table 5-4: Results of searching the sequence alignment of complementing XRCC4 orthologues for pathogenic XRCC4 variants on ClinVar 129
Table 5-5: Amino acid positions of XRCC4 exons and the amino acids encoded across exon boundaries 129
Table 5-6: Results from searching the sequence alignment of complementing XRCC4 orthologues for variation at PTM sites in XRCC4 reported on the PhosphoSitePlus database 132
Table 5-7: Computational predictions for XRCC4 variants tested 134
Table 6-1: Accession numbers and % identity to human orthologue for vertebrate MLH1 orthologues used 153
Table 6-2: Summary of the complementation ability of MLH1 orthologues in human cells 162
Table 6-3: Results from searching the sequence alignment of complementing MLH1 orthologues for homozygous MLH1 variants present in gnomAD. 165
Table 6-4: Results of searching the sequence alignment of complementing MLH1 orthologues for gnomAD variants with allele frequencies > 0.01% 166
Table 6-5: Results of searching the sequence alignment of complementing MLH1 orthologues for all MLH1 gnomAD variants 168
Table 6-6: Results of searching the sequence alignment of complementing MLH1 orthologues for benign ClinVar MLH1 variants 169

Table 6-7: Results of searching the sequence alignment of complementing MLH1 orthologues for pathogenic MLH1 variants present on the ClinVar database. 172
Table 6-8: Amino acid positions of MLH1 exons and the amino acids encoded across exon boundaries 173
Table 6-9: Results of searching the sequence alignment of complementing MLH1 orthologues for ClinVar variants without ‘benign’ or ‘pathogenic’ classifications. 185
Table 6-10: Results from searching the sequence alignment of complementing MLH1 orthologues for variation to non-modifiable residues at reported PTM sites. 194
Table 7-1: Accession numbers and % identity to human orthologue for vertebrate FANCL orthologues used 217
Table 7-2: Conservation of residues required for intramolecular and intermolecular interactions of FANCL 227

1 Abstract

The DNA in every cell of the human body is subject to a barrage of exogenous and endogenous mutagens, causing tens of thousands of DNA lesions per cell, per day. To counteract this assault, cells have evolved a network of proteins that work to detect and repair the DNA lesions, collectively known as the DNA damage response (DDR). Despite the actions of the DDR, DNA damage can persist and may lead to changes to the genetic code, in the form of missense, nonsense and silent mutations and chromosomal rearrangements. Mutagenesis in this manner contributes to diseases such as cancer and accelerated ageing. Therefore, preserving genomic integrity is of paramount importance for the health of the organism. In this thesis, I focus on the functional consequences of missense mutations, defined as base pair substitutions altering the amino acid coded for by the codon. The alternative amino acid incorporated into the protein, termed a 'variant', may cause disruption to protein activity depending on its location and characteristics. Disruption to protein activity can contribute to carcinogenesis, and is often used to stratify patients forming the basis for specific lines of treatment. Currently, a plethora of methods exist for interpreting the consequences of variants on protein function, but often lack a physiological readout or are limited in their throughput. Here, I describe a new method that provides an efficient, high-throughput means of assessing the impact of variants on protein function. It is based on complementation of a specific human gene knockout cell phenotype, by expression of a series of vertebrate orthologues with increasing evolutionary distance from human. Due to the natural variation of amino acid residues observed in orthologue sequences, full complementation with any given orthologue could reveal variants tolerated by the human protein at certain positions. Here, I test this methodology using proteins from the DDR, due to their extensive cross-species conservation, mutation in diseases such as cancer and relevance to synthetic lethality-based treatments. I show that this method can identify functional variants

in DDR genes, as well as determining the importance of previously identified post-translational modification (PTM) sites. Collectively, this method is able to provide insight into impact of amino acid variants on protein function, which offers clinical benefits as well as furthering our understanding of protein regulation.

2 Introduction

2.1 Sources of DNA damage and the DNA damage response

2.1.1 The DNA damage response

The DNA within the cells of our body is continuously exposed to a large number of endogenous and exogenous mutagens, which generate distinct types of DNA lesions (Ciccia and Elledge, 2010) Error! Reference source not found.**.** It has been estimated that the DNA within each cell experiences approximately 70,000 lesions per day (Ciccia and Elledge, 2010). To counteract this assault, cells have evolved a distinct set of biochemical pathways which induce a wide variety of cellular processes to maintain the genomic integrity of the organism (Jackson and Bartek, 2009, Ciccia and Elledge, 2010). Collectively this network is termed the DNA damage response (DDR) (Ciccia and Elledge, 2010, Jackson and Bartek, 2009) **Figure 2-1**.

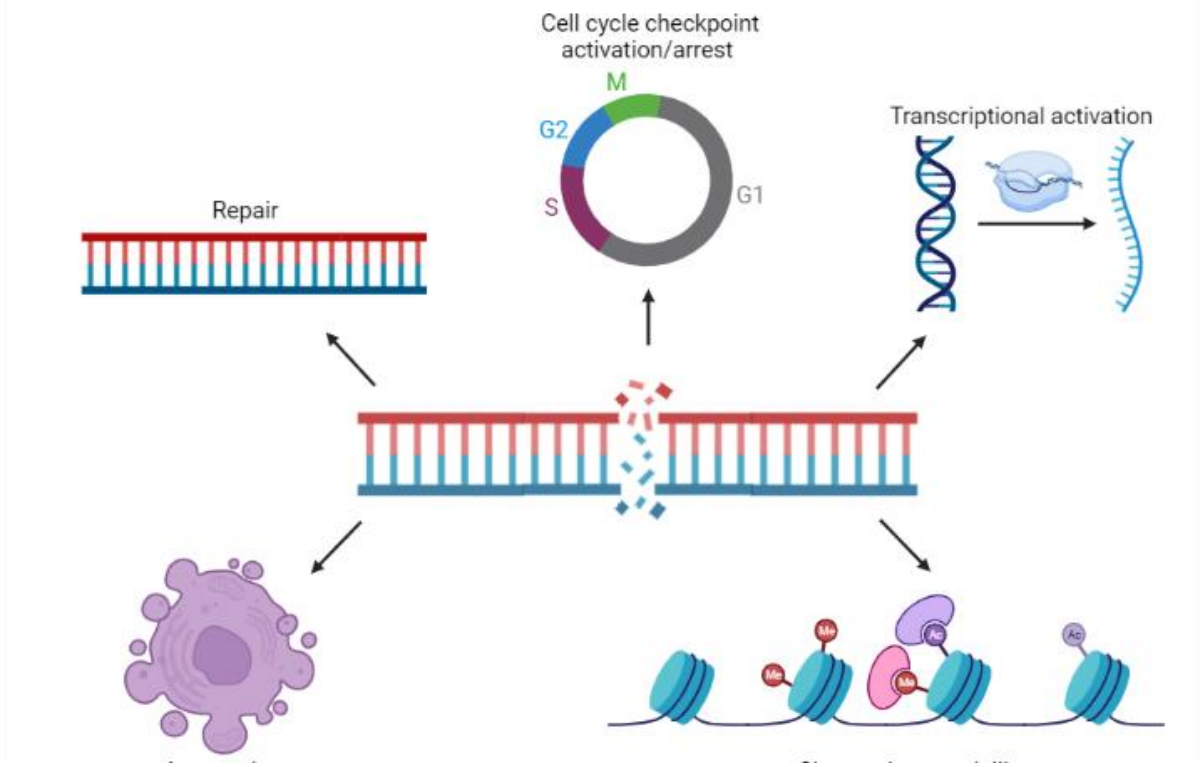


Figure 2-1: The cellular response to DNA damage

DNA damage induces a wide array of cellular processes including DNA repair, chromatin remodelling, cell cycle checkpoint activation/arrest, apoptosis and transcriptional activation which work to preserve the genomic integrity of the organism. Created with BioRender.com

2.1.2 Endogenous sources of DNA damage

Endogenous DNA damage can arise via oxidation during which DNA molecules engage in oxidative reactions with reactive oxygen species (ROS) within cells (Kryston et al., 2011, Chatterjee and Walker, 2017). Typically, the ROS species produced are superoxide radicals ($\bullet O_2^-$), hydrogen peroxide (H_2O_2) and the hydroxyl radical (•OH) generated via the electron transport chain during oxidative phosphorylation (Henle and Linn, 1997, Chatterjee and Walker, 2017). Hydroxyl radicals can react with DNA bases (Winterbourn, 2008), with a typical example being the hydroxylation of the C-8 residue of guanine to produce 8-oxoguanine (8-oxo-G) (Cadet et al., 2011). 8-oxo-G pairs incorrectly with adenine instead of cytosine, often resulting in guanine to thymine base substitutions in the next round of DNA replication (Cheng et al., 1992). It is estimated that approximately 400 - 1500 8-oxo-G lesions are generated per cell, per day (Ciccia and Elledge, 2010). Other than ROS, endogenous DNA damage can arise through other sources such as spontaneous base deamination, replication errors and DNA base mismatches (Tubbs and Nussenzweig, 2017). Spontaneous base deamination is a major source of mutagenesis in human cells, where cytosine (C), adenine (A), guanine (G), and 5-methyl cytosine (5-mC) lose their exocyclic amine to become uracil (U), hypoxanthine, xanthine and thymine (T), respectively (Yonekura et al., 2009). As a consequence, deamination can result in mutation, e.g. deamination of cytosine turns the native C:G base pair to a U:A base pair in the first round of replication, which in the next round of replication results in a C-T mutation (Yonekura et al., 2009). It is estimated that approximately 200 spontaneous deamination events occur per cell, per day (Ciccia and Elledge, 2010). Replication errors and DNA base mismatches result in single base pair additions, deletions or substitutions (Tubbs and Nussenzweig, 2017). These can be introduced during DNA replication and repair by error-prone DNA polymerases (Loeb, 2016, Loeb and Monnat, 2008). It is estimated that the replicative DNA polymerases δ and ε, make an error every 10^4-10^5 nucleotides, resulting in

100,000 polymerase errors per cell division (Bębenek and Ziuzia-Graczyk, 2018, Bebenek and Kunkel, 2004). To compensate, polymerases have evolved proofreading activity which improves the fidelity to an error every 10^8-10^{10} nucleotides (Bębenek and Ziuzia-Graczyk, 2018). Replicative polymerases may also incorrectly incorporate uracil into DNA depending on the relative and absolute concentrations of deoxynucleoside triphosphates (dNTPs) and ribonucleoside triphosphates (rNTPs) in the nucleus (Kumar et al., 2011, Clausen et al., 2013, Potenski and Klein, 2014). Additionally, during DNA replication, the template and primer strands can can dissociate and re-anneal incorrectly at microsatellites, defined as repeated-sequence motifs e.g. $[A]_n$ or $[CA]_n$ (Yamamoto and Imai, 2015). Such an event results in partnerless nucleotide sequences termed insertion/deletion loops (IDLs), and is termed microsatellite instability (MSI) (Jiricny, 2006).

The byproducts of metabolic pathways such as lipid oxidation and glycation, as well as ethanol processing, give rise to endogenous aldehydes (Garaycoechea et al., 2018, Vernole et al., 2011). These produce several types of DNA lesions including base adducts, DNA-DNA inter and intra-strand crosslinks, and DNA-protein crosslinks, which can form a barrier to DNA replication (Vernole et al., 2011, Garaycoechea et al., 2018). Lastly, during transcription, the formation of an RNA-DNA hybrid in complex with a displaced single-stranded DNA (ssDNA) strand, termed an R-loop (Huertas and Aguilera, 2003), can stall replication fork progression leading to replication fork collapse and double-strand break formation (Santos-Pereira and Aguilera, 2015).

2.1.3 Exogenous sources of DNA damage

In addition to the endogenous mutagens, cells also encounter a wide range of exogenous DNA damaging agents. These include ionising radiation (IR), ultraviolet (UV) radiation and crosslinking agents (Huang and Zhou, 2021). Exogenous mutagens such as these are able to induce single and double-strand DNA breaks (SSBs and DSBs), as well as mutagenic DNA

crosslinks, dimers and adducts. IR is classified by its effect (direct or indirect) combined with the amount of energy transferred to matter, referred to as linear energy transfer (LET) (Mavragani et al., 2019). IR is abundant in the environment with α, β, γ, neutrons, and X-rays being produced from diverse sources (Chatterjee and Walker, 2017). Most of an individual's annual IR exposure derives from a combination of high-LET cosmic rays and alpha particles emitted by decaying radon gas of geological origin (Moore et al., 2014). IR can induce multiple types of DNA lesions, including DSBs (Hutchinson, 1985), SSBs and base lesions such as 8-oxo-G (Henner et al., 1982). Such lesions arise either through the direct exposure of the DNA to IR, or via indirect radiolysis of the surrounding water in cells, which generates a cluster of hydroxyl radicals (Desouky et al., 2015). The proximity and nature of such lesions is largely dependent on the LET of the IR in question (Hada and Georgakilas, 2008). Both high-LET and low-LET IR produce the same spectrum of lesions (Moore et al., 2014), but high-LET IR results in greater lesion clustering due to the confined nature of the energy deposition, creating as many as 25 lesions per 1-2 helical turns (Semenenko and Stewart, 2004).

UV radiation is classified into three categories depending on its wavelength, specifically UV-A (320-400 nm), UV-B (290-320 nm) and UV-C (190-290 nm) (Rastogi et al., 2010). DNA absorbs maximal UV radiation at 260 nm (UV-C), beyond which the photo-absorption drops dramatically (Kciuk et al., 2020). Due to their wavelengths, UV-B and UV-C are most damaging to DNA (Kciuk et al., 2020). However most of the UV-B and UV-C radiation is filtered out by the ozone layer, preventing it from reaching the earth's surface (McKenzie et al., 2003). Upon absorption of UV radiation by DNA, covalent linkages between two adjacent pyrimidines are generated, forming DNA lesions in the form of cyclobutane pyrimidine dimers (CPDs) and 6-4 pyrimidine photoproducts (6-4PPs) (Sinha and Häder, 2002, Rastogi et al., 2010). These lesions generate distortions in the DNA helix, and if not repaired, impact on

processes such as DNA replication and transcription, often leading to mutagenesis (Saini et al., 2021).

Crosslinking agents are commonly used in chemotherapy with the most notable example being cisplatin (Kelland, 2007). Furthermore, crosslinking agents are found in tobacco smoke and also arise via metabolism of alcohol (Ceccaldi et al., 2016). Crosslinking agents form covalent linkages between two strands of DNA, thereby creating a block to DNA replication and transcription (Ceccaldi et al., 2016). Unrepaired crosslinks can also result in DNA breakage and chromosomal rearrangements (Deans and West, 2011). In the case of cisplatin, the water molecule in its chemical structure displaces N-heterocyclic bases on DNA (preferentially by guanine), linking it to the DNA (Eastman, 1987). Subsequently, DNA crosslinking occurs via displacement of one of cisplatin's chlorine atoms by another guanine base (Eastman, 1987).

Chemotherapeutic compounds such as 6-thioguanine (6-TG) and 5-fluorouracil (5-FU) are classed as antimetabolites and produce mismatches during DNA replication (Christensen et al., 2019, Swann et al., 1996, Waters and Swann, 1997). 6-TG can be incorporated into the DNA, and is then methylated by S-adenosylmethionine forming 6-methylthioguanine (^{Me}6-TG) which is capable of mispairing with thymine (Swann et al., 1996; Waters and Swann, 1997). Similarly, 5-FU results in the incorporation of 5-fluoro-2'-deoxyuridine into the DNA which can mispair with guanine (Christensen et al., 2019).

Endogenous mutagen	**DNA lesion generated**	**Estimated number of lesions/cell/day**
Oxidation	8-oxo-G	400-1500
Spontaneous base deamination	C → U A → hypoxanthine G → xanthine 5-mC → T	200
Replication errors of polymerases δ and ε	Misincorporation of DNA bases	100,000
Exogenous mutagen	**DNA lesion generated**	**Estimated number of lesions/cell/day**
Sunlight (UV radiation)	CPDs, 6-4PPs	100,000 (during peak hour sunlight)
External beam therapy (X-rays)	DSBs	72-80
Chest X-ray	DSBs	0.0008
Cigarette smoke	Aromatic DNA adducts	50-1000
Chernobyl accident	DSBs	12

Table 2-1 Sources of endogenous and exogenous DNA mutagens and their associated lesions.

Table describing the types and frequencies of DNA lesions associated with endogenous and exogenous DNA mutagens. Adapted from (Ciccia and Elledge, 2010).

2.1.4 DNA damage detection, signalling and checkpoint activation

In order for DNA damage to be repaired, the cell must first detect the lesion and then signal to the appropriate DNA repair pathway to carry out repair (Blackford and Jackson, 2017, Jackson and Bartek, 2009). In response to damage, signalling cascades are activated which can induce cell cycle checkpoints, chromatin alterations and gene expression, to facilitate the repair of the lesion (Blackford and Jackson, 2017). If lesions persist, these signalling pathways can induce senescence, apoptosis or cellular differentiation, to minimise the chances of damaged DNA being passed on to cellular progeny (Simoneau and Zou, 2021, Shibata and Jeggo, 2021). Three phosphatidylinositol 3-kinase like protein kinases (PIKKs), ATM, ATR and DNA-PK control the majority of DNA damage signalling within human cells (Blackford and Jackson, 2017, Savitsky et al., 1995, Bentley et al., 1996, Cimprich et al., 1996, Carter et al., 1990, Lees-Miller et al., 1990). ATM and DNA-PK primarily respond to DSBs, whereas ATR responds to

multiple types of lesion including DSBs, ssDNA regions and base adducts (Blackford and Jackson, 2017, Simoneau and Zou, 2021, Shibata and Jeggo, 2021).

In G1 and S/G2 phases, DSBs are bound by the MRE11-RAD50-NBS1 (MRN) complex, which recruits ATM via interaction with NBS1 (Dupré et al., 2006, Lee and Paull, 2005, Falck et al., 2005). ATM can phosphorylate CHK2 (Ahn et al., 2000, Brown et al., 1999, Matsuoka et al., 1998, Matsuoka et al., 2000), which phosphorylates p53 (Canman et al., 1998), leading to transcription of the CDK inhibitor p21 and subsequent inhibition of the Cyclin A/E-CDK2 complex, preventing entry into S-phase (Kastan and Bartek, 2004). In S/G2-phases, ATM promotes the resection of DSB ends via phosphorylation of CtIP, producing ssDNA overhangs at the site of DSBs (Adams et al., 2006, Cuadrado et al., 2006, Sartori et al., 2007, Shibata et al., 2011, Wang et al., 2013). The ssDNA ends provide a substrate for RPA (Raderschall et al., 1999) to which ATR is recruited via ATRIP (Zou and Elledge, 2003) and subsequently activated by TOPBP1 (Kumagai et al., 2006) or ETAA1 (Bass et al., 2016, Feng et al., 2016, Haahr et al., 2016). Upon activation, ATR phosphorylates and activates CHK1 (Guo et al., 2000, Hekmat-Nejad et al., 2000, Liu et al., 2000, Zhao and Piwnica-Worms, 2001). Activated CHK1 phosphorylates and inhibits CDC25 (Mailand et al., 2000, Sanchez et al., 1997), which in its active state removes the inhibitory phosphorylation on CDK1 and CDK2 deposited by the WEE1 kinase (Gu et al., 1992, Heald et al., 1993) and the WEE1-like kinase PKMYT1 (Booher et al., 1997, Liu et al., 1997). Furthermore, CHK1 activates WEE1, promoting inhibitory phosphorylation of CDC25 by WEE1 (Beck et al., 2010, Beck et al., 2012). Collectively, this prevents entry into mitosis as Cyclin A/B-CDK1/2 complexes are responsible for progression through the G2/M transition (Barnaba and LaRocque, 2021). In addition to preventing premature mitotic entry, ATR also prevents the collapse of replication forks in cells experiencing replication stress (Blackford and Jackson, 2017, Simoneau and Zou, 2021). Inhibition of CDKs via ATR restrains replication origin firing (Costanzo et al., 2003) which

can limit the exhaustion of nucleotide pools and repair factors such as RPA (Toledo et al., 2013). Furthermore, ATR targets the SMARCAL1 helicase, preventing formation of fork configurations which can be cleaved by the SLX4/MUS81 complex (Couch et al., 2013, Forment et al., 2011, Ragland et al., 2013). Lastly, ATR can increase the availability of nucleotides by upregulating the ribonuclease reductase subunit RRM2 (Buisson et al., 2015, Pfister et al., 2015, D'Angiolella et al., 2012). After detection of DNA damage, the cell must carry out the repair of the lesion in order to maintain the genomic integrity of the organism. The wide array of lesions that a cell experiences means that multiple pathways have evolved to repair distinct types of lesion **Figure 2-2.** These pathways are discussed below.

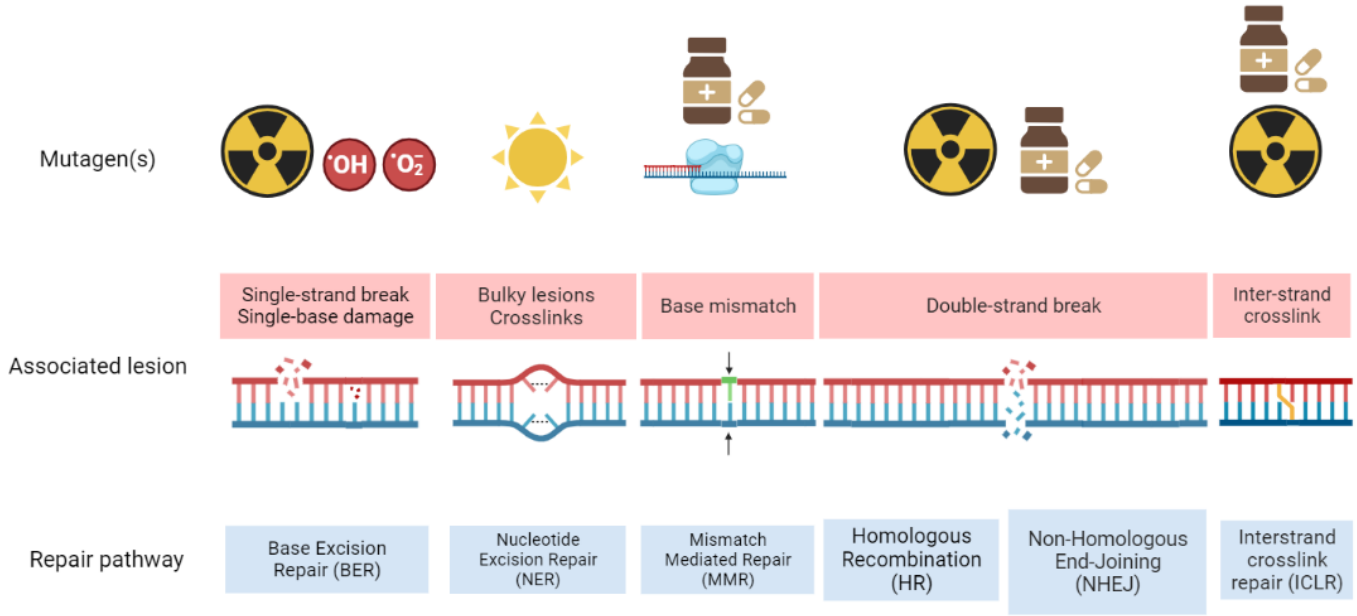


Figure 2-2: Types of DNA lesion and their associated repair pathway(s)

As cells experience a wide array of DNA lesions induced by various exogenous and endogenous mutagens, they have evolved multiple distinct repair pathways which work to repair specific types of DNA damage. Created with BioRender.com.

2.1.5 Single-strand break repair

SSBs are recognised by members of the poly(ADP-ribose) polymerase (PARP) family, primarily PARP1 or PARP2 (Caldecott, 2014). PARP1/2 use NAD^+ to add poly(ADP-ribose) (PAR) chains onto adjacent proteins and PARP1 itself (Schreiber et al., 2006). Such chains facilitate the recruitment of the scaffold protein XRCC1 (El-Khamisy et al., 2003), permitting recruitment of LIG3 (Leppard et al., 2003) and other accessory factors which work to re-ligate the break (Caldecott, 2014).

2.1.6 Mismatch repair

IDLs and mismatched DNA bases arising via means such as spontaneous base deamination and DNA polymerase replication errors, are detected and repaired via the mismatch repair (MMR) pathway (Jiricny, 2006). This pathway was first described in 1975, with the finding that *E. coli* can recognise and process heteroduplex DNA constructs containing mismatched bases (Wildenberg and Meselson, 1975, Wagner and Meselson, 1976). This repair depends on the products of the *mutS, mutL, mutH* and *uvrD* genes *in vivo* (Nevers and Spatz, 1975, Rydberg, 1978, Glickman and Radman, 1980) and *in vitro* (Lu et al., 1983, Su and Modrich, 1986). Additional proteins such as DNA ligase, DNA polymerase III, SSB and exonuclease 1 are also required, and their discovery led to the *E. coli* MMR system being fully reconstituted *in vitro* (Lahue et al., 1989). Mechanistically, mutS recognises the mismatch (Su and Modrich, 1986) and mutL is recruited to the mutS heteroduplex (Grilley et al., 1989, Au et al., 1992). The transient unmethylated state of adenine in newly replicated prokaryotic DNA provides the strand discrimination necessary for the endonuclease activity of mutH which acts at unmethylated d(GATC) sites (Lu et al., 1983, Welsh et al., 1987). mutS and mutL are required for activation of mutH endonuclease activity at the unmethylated strand 5’ or 3’ to the mismatch (Au et al., 1992). mutS and mutL load DNA helicase II at the mutH strand break, allowing unwinding of the DNA towards the mismatch (Dao and Modrich, 1998, Yamaguchi

et al., 1998). If the nick created by mutH is 5' to the mismatch, the displaced DNA is degraded by either exoVII or recJ (Cooper et al., 1993, Burdett et al., 2001, Viswanathan et al., 2001). If 3', then exoI or exoX are responsible (Viswanathan et al., 2001, Cooper et al., 1993).

Alongside prokaryotes, mismatches are also corrected in eukaryotic cells (Glazer et al., 1987, Brown and Jiricny, 1988). Some patient-derived colorectal cancer cell lines were found to demonstrate MSI (Aaltonen et al., 1993, Ionov et al., 1993), a phenotype typical of *E. coli* MMR mutants (Levinson and Gutman, 1987). Subsequently, these cell lines were shown to be MMR-deficient (Parsons et al., 1993, Veigl et al., 1998), due to a lack of proteins comprising MutSα - a heterodimer comprised of two MutS homologues MSH2 and MSH6 (Drummond et al., 1995, Iaccarino et al., 1996, Palombo et al., 1996), or MutLα - a heterodimer of the two MutL homologues MLH1 and PMS2 (Li and Modrich, 1995). Additional proteins involved in eukaryotic MMR include the MSH2-MSH3 heterodimer (MutSβ) (Palombo et al., 1996, Acharya et al., 1996, Genschel et al., 1998), EXO1 (Szankasi and Smith, 1995, Tishkoff et al., 1997, Genschel et al., 2002), RPA (Lin et al., 1998, Ramilo et al., 2002), PCNA (Umar et al., 1996), RFC (Dzantiev et al., 2004) and DNA polymerase δ (Longley et al., 1997). Eukaryotic MMR is initiated by the binding of either the MutSα complex or the MutSβ complex to mismatched bases or insertion/deletion loops (Genschel et al., 1998, Palombo et al., 1996). MutSα preferentially recognises base-base mismatches and smaller insertion/deletion loops of 1-3 extrahelical residues (Drummond et al., 1995, Palombo et al., 1996, Genschel et al., 1998). MutSβ recognises larger insertion/deletion loops of around 2-10 bases (Acharya et al., 1996, Palombo et al., 1996, Genschel et al., 1998). Mismatch recognition by MutSα is mediated by MSH6 via base-flipping, whereas MutSβ recognition is mediated by MSH3 via insertion of residues between the two DNA strands (Groothuizen et al., 2015). Such interactions distort the DNA by approximately 45° helping to discriminate mismatches from intact DNA (Groothuizen et al., 2015). MSH6 can bind and hydrolyse ATP (Antony et al., 2006), enhancing binding of

ATP by MSH2 (Heinen et al., 2011). Once hydrolysed, MSH6 releases the ADP product whereas MSH2 retains ADP, preventing further ATP binding by MSH6 (Heinen et al., 2011, Mazur et al., 2006). This conformation allows free diffusion of MutSα complexes along the DNA to enable detection of mismatches (Gorman et al., 2007). Mismatch detection is also mediated by coupling of MutSα and MutSβ to PCNA, allowing detection in close proximity to replication (Clark et al., 2000, Flores-Rozas et al., 2000, Kleczkowska et al., 2001, Indiani et al., 2005, Hombauer et al., 2011). Upon detection of a mismatch, MutS family members bind ATP inducing conformational changes which permit the binding of MutL homologues (Groothuizen et al., 2015, Mendillo et al., 2009, Fernandez-Leiro et al., 2021, Trojan et al., 2002, Plotz et al., 2003, Plotz et al., 2006, Plotz et al., 2002). The endonuclease activity of MutLα is then activated via interaction with PCNA in an RFC-dependent manner (Kadyrov et al., 2006). The strand specificity of the endonuclease activity of MutLα is ensured by the mirror image loading of PCNA on the leading and lagging strands (Pluciennik et al., 2010). The endonuclease activity of MutLα can cut distally, but also proximally, relative to the gap at which PCNA was loaded by RFC (Dzantiev et al., 2004, Pluciennik et al., 2010). After the incision step, EXO1 is recruited and activated via MSH2 and/or MLH1 (Tran et al., 2004, Nielsen et al., 2004). EXO1 then works to remove the mismatch containing DNA in a 5'- 3' direction (Genschel and Modrich, 2003, Dzantiev et al., 2004). Mispair excision can also occur in the absence of EXO1 (Goellner et al., 2015). MutL homologues are capable of performing multiple incisions either side of the mismatch via PCNA stimulation which result in nick-directed strand degradation (Goellner et al., 2015). Furthermore, such nicks can enable processing by DNA Polδ-dependent strand displacement (Kadyrov et al., 2009) or by the 3′ – 5′ exonuclease activity of DNA polymerases (Tran et al., 1999). Following removal of the mismatch containing strand, Polδ fills in the gap using the intact strand as a template with the

remaining nick sealed by DNA ligase I (Prindle and Loeb, 2012, Longley et al., 1997) **Figure 2-3.**

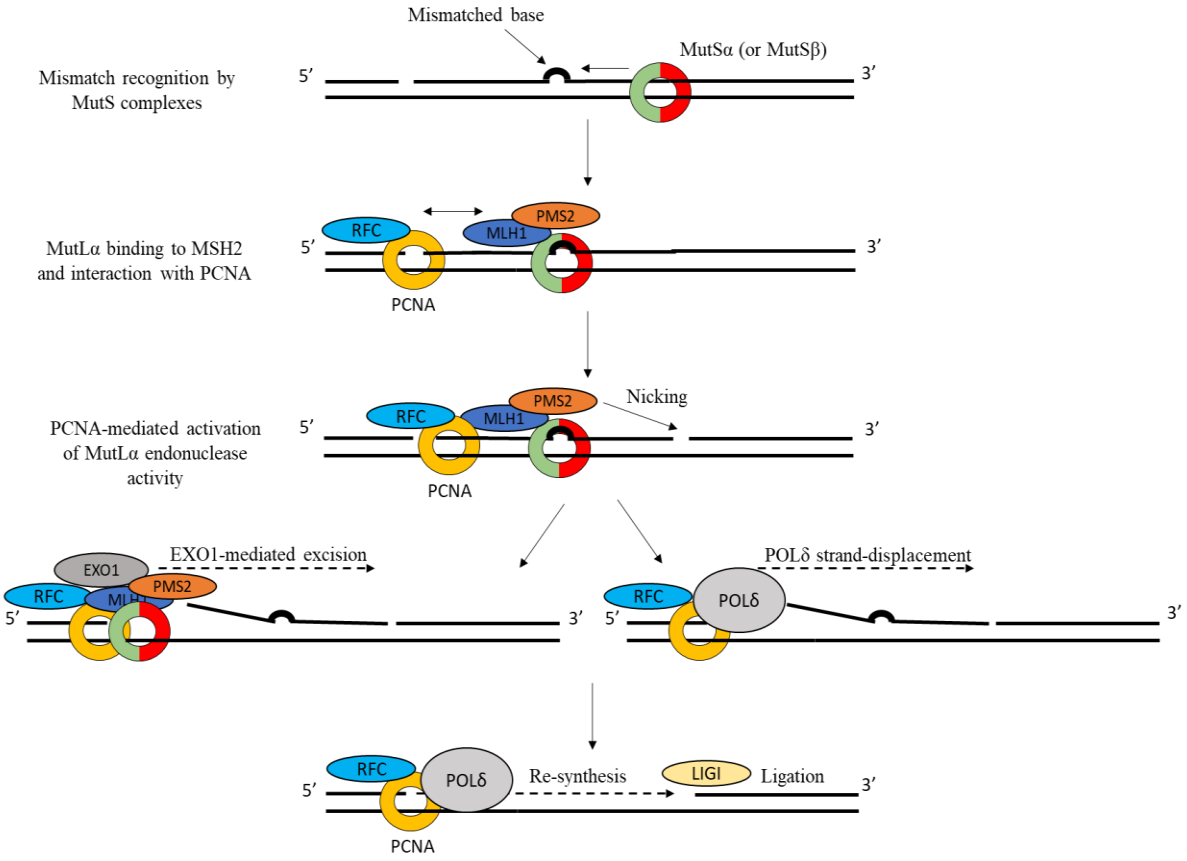


Figure 2-3: The eukaryotic post-replicative MMR pathway

Eukaryotic MMR is initiated by the binding of either MutSα or MutSβ heterodimers to the mismatched bases. ATP induced conformational changes permit the binding of the MutLα heterodimer consisting of the MLH1 and PMS2 proteins. Interaction of MutLα with PCNA loaded by RFC activates the endonuclease activity of PMS2 which nicks the mismatch containing DNA. Subsequently, the mismatch containing strand is removed by one of multiple mechanisms including EXO1-mediated 5' - 3' excision or POLδ-mediated strand displacement. Lastly, the DNA is re-synthesised by POLδ with the remaining gap sealed by LIGI.

2.1.7 Base excision repair

Chemical alterations to DNA bases arising from oxidation, deamination and alkylation are repaired via the base excision repair (BER) pathway (Krokan and Bjørås, 2013, Beard et al., 2019). Damaged bases are removed via the action of a DNA glycosylase, specific to the chemical alteration to the base (Krokan and Bjørås, 2013, Beard et al., 2019). This generates an abasic site, which provides a substrate for the endonuclease APE1 generating a SSB at this region (Mol et al., 2000). Gap filling and DNA ligation ensues, via short-patch repair involving XRCC1, LIG3 and POLβ, or long-patch repair requiring PCNA, LIG1, POLβ, POLδ and FEN1 (Krokan and Bjørås, 2013, Beard et al., 2019).

2.1.8 Nucleotide excision repair

UV radiation induced lesions such as CPDs and 6-4PPs are repaired by the nucleotide excision repair (NER) pathway (Marteijn et al., 2014, Schärer, 2013, Duan et al., 2021). The NER pathway is also involved in the repair of adducts formed by environmental mutagens such as benzo[a]pyrene, aromatic amines, certain oxidative endogenous lesions e.g. cyclopurines, and adducts formed by chemotherapeutic drugs such as cisplatin (Marteijn et al., 2014, Schärer, 2013, Duan et al., 2021). NER can be divided into two main pathways: global-genomic NER (GG-NER) and transcription-coupled NER (TC-NER) (Spivak, 2015, Marteijn et al., 2014, Duan et al., 2021). Both pathways comprise of a lesion-recognition, lesion-verification, lesion-excision and lesion-repair step. The pathways differ in their lesion-recognition step (Marteijn et al., 2014, Kusakabe et al., 2019). In GG-NER, UV-induced lesions are recognised by the UV-DDB complex, consisting of the DDB1 and DDB2 proteins (Wakasugi et al., 2001, Scrima et al., 2008). This facilitates the binding of XPC (Sugasawa et al., 1998), stabilised by its association with RAD23B and CENT2 (Nishi et al., 2005, Masutani et al., 1994, Volker et al., 2001). In TC-NER, lesion-recognition occurs when RNA polymerase II (RNA-Pol II) is stalled during transcript elongation, resulting from a lesion in the template strand (Vermeulen and

Fousteri, 2013). Stalled RNA-Pol II shows greater affinity for the TC NER specific CSB protein (Boom et al., 2004), which then recruits the CSA protein to form the CSA-CSB complex (Spivak, 2015, Duan et al., 2021). The CSA-CSB complex likely facilitates the backtracking of RNA-Pol II, to reveal the lesion responsible for its stalling (Sigurdsson et al., 2010, Marteijn et al., 2014), and subsequently recruits the NER machinery to mediate repair of the lesion (Weegen et al., 2020). Following lesion-recognition in both GG-NER and TC-NER, TFIIH is recruited to the lesion site (Yokoi et al., 2000, Volker et al., 2001, Weegen et al., 2020) and this is where the mechanisms of the two pathways converge (Marteijn et al., 2014). TFIIH consists of 10 subunits including the ATPases/helicases XPB and XPD (Compe and Egly, 2012). XPB and XPD helicases have opposite polarities and work to extend the open DNA configuration around the lesion (Tapias et al., 2004, Compe and Egly, 2012). However, only the ATPase activity of XPB is required for NER, whereas both the ATPase and helicase activities of XPD are required (Coin et al., 2007, Winkler et al., 2000). The helicase activity of XPD is thought to be blocked by DNA lesions, providing a lesion-verification step during NER (Liu et al., 2008, Fan et al., 2008, Mathieu et al., 2013). Stalling of the XPD helicase at the lesion is succeeded by the individual recruitment of XPA, RPA and the XPG endonuclease (Volker et al., 2001, Rademakers et al., 2003), followed by binding of the XPF-ERCC1 endonuclease to XPA (Li et al., 1994, Tsodikov et al., 2007). Subsequently, XPF-ERCC1 makes a 5'-incision (Staresincic et al., 2009), generating a free 3' hydroxyl group used by the replication machinery to initiate DNA synthesis by PCNA, RFC and polymerases δ, ε, or κ (Shivji et al., 1995, Ogi et al., 2010). Lastly, the lesion containing DNA is displaced and XPG performs a 3'-incision to release the DNA (Staresincic et al., 2009, O'Donovan et al., 1994), with the gap subsequently repaired via LIG1 or LIGIII in a cell cycle-dependent manner (Marteijn et al., 2014) **Figure 2-4**.

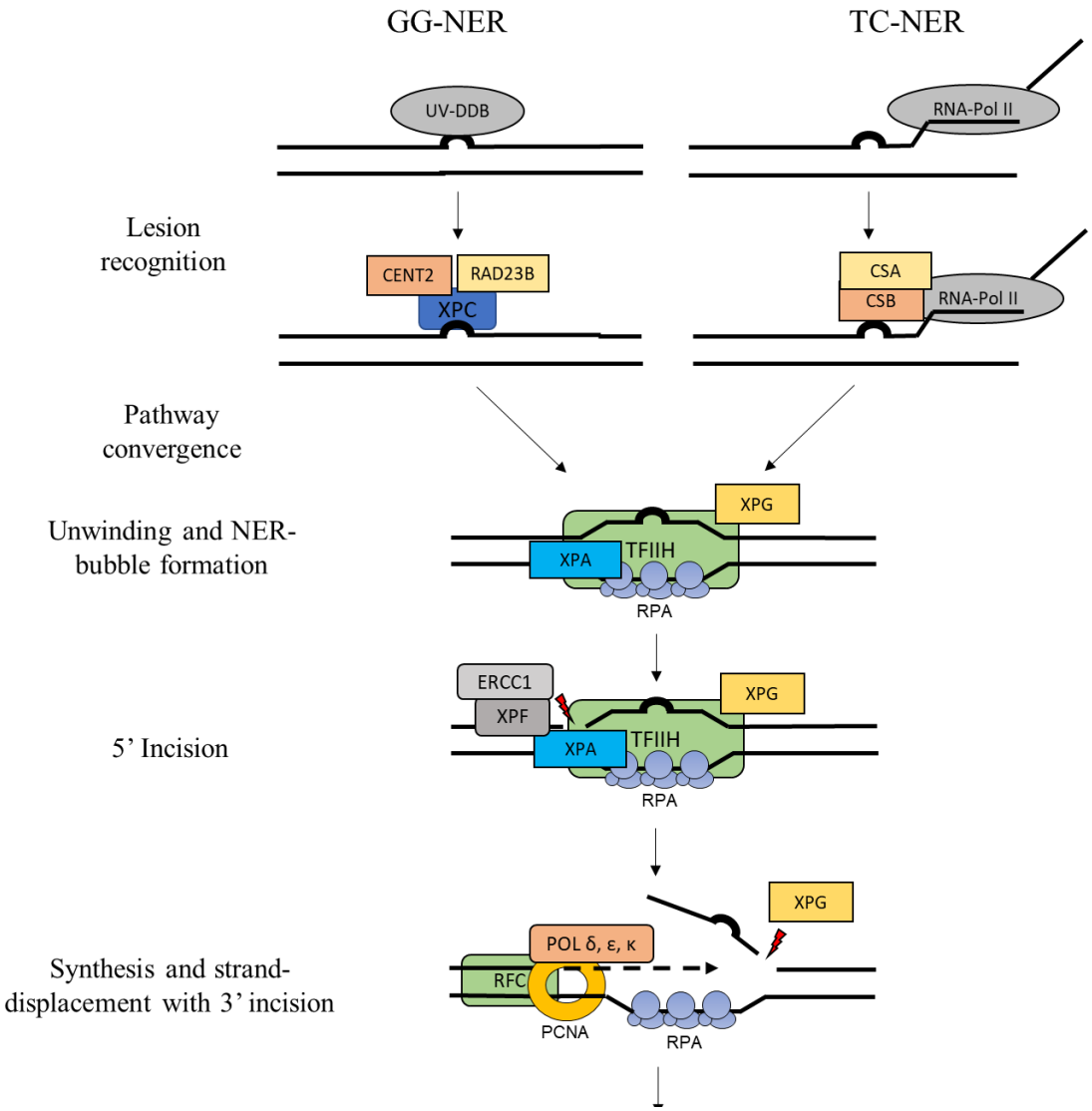


Figure 2-4: The nucleotide excision repair pathway

The initial step of both GG-NER and TC-NER is lesion-recognition. In GG-NER, UV-DDB recognises the lesion and facilitates the binding of XPC in complex with CENT2 and RAD23B. In TC-NER, RNA-Pol II is stalled upon collision with a lesion which permits the binding of first CSB then CSA. Subsequently, the two pathways converge upon binding of TFIIH, which unwinds the DNA surrounding the lesion using its XPB and XPD subunits and is succeeded by the individual recruitment of XPA, RPA and the XPG endonuclease to form the NER bubble. The ERCC1-XPF endonuclease is then recruited to XPA and performs an incision 5' to the lesion. PCNA loaded by RFC then permits POLδ, κ or ε to synthesise new DNA using the intact strand as a templated. This also displaces the strand containing the lesion which is then subsequently removed via XPG-mediated 3' incision. Lastly, the final gap at the termination of DNA synthesis is sealed by either LIGI or LIGIII.

2.1.9 Non-homologous end-joining

DSBs arising via various exogenous or endogenous mutagens are highly toxic DNA lesions and multiple pathways exist to repair them. The primary pathway for repairing DBSs is non-homologous end-joining (NHEJ). NHEJ is initiated by the binding of the Ku heterodimer (Ku70 & Ku80) to DSB ends (Walker et al., 2001). Ku is an extremely abundant protein with high-affinity for DSB ends (Mimori and Hardin, 1986). Ku serves two primary roles in NHEJ: the first is to protect DSB ends from resection (Sun et al., 2012), the second is to serve as a platform for the recruitment of downstream proteins in the NHEJ pathway (Frit et al., 2019). Ku-bound DNA ends recruit the DNA-PK catalytic subunit (DNA-PKcs) (Carter et al., 1990, Lees-Miller et al., 1990), which together form the active DNA-PK holoenzyme (Gottlieb and Jackson, 1993). The kinase activity of this complex is largely directed towards itself, promoting DNA end-synapsis (Graham et al., 2016, Chen et al., 2021), but also enables the dissociation of this complex from DSB ends once repair is completed (Dobbs et al., 2010, Hammel et al., 2010). Subsequently, XRCC4, LIG4 and XLF or PAXX are recruited which form a protein-protein interaction network forming the long-range synaptic complex to bridge two DNA-PK bound DSB ends (Chen et al., 2021). Following dissociation of DNA-PKcs, conformational changes in the remaining proteins forms the short-range synaptic complex (Chen et al., 2021, Graham et al., 2016, Graham et al., 2018a) positioning two LIG4 molecules in close proximity to carry out repair of the lesion (Critchlow et al., 1997, Grawunder et al., 1997, Chen et al., 2021). Finally, the NHEJ factors dissociate from the DNA and repair is completed (Chen et al., 2021). Additional NHEJ factors include nucleases such as Artemis and APLF (Ma et al., 2002, Iles et al., 2007), polymerases such as Pol μ and Pol λ and TdT (Davis and Chen, 2013) as well as kinase/phosphatase proteins such as PNKP (Koch et al., 2004). These assist in the repair of a subset of DSBs, but are not essential for the NHEJ process (Zhao et al., 2020) **Figure 2-5.**

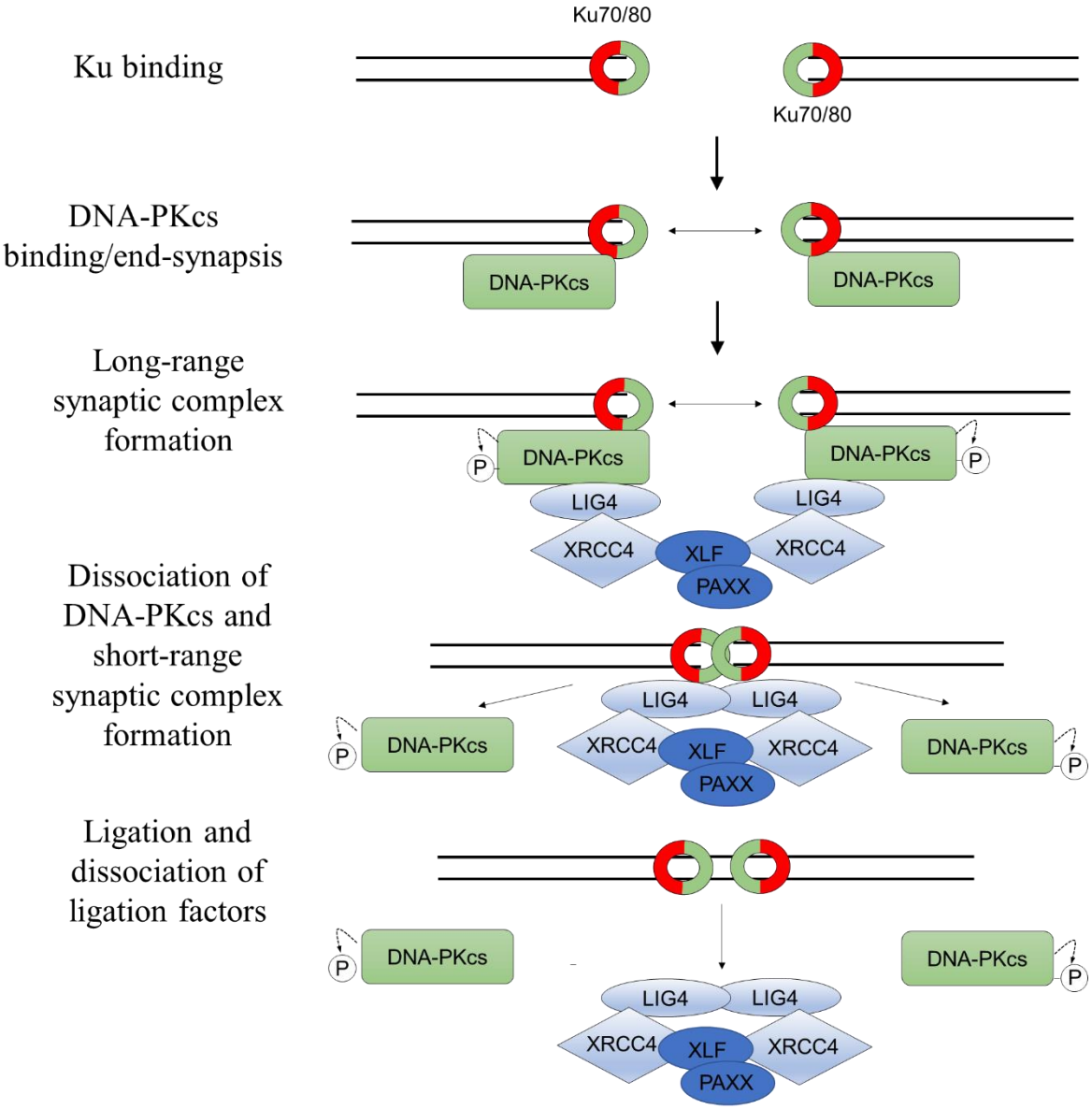


Figure 2-5: The non-homologous end-joining pathway

NHEJ is initiated by the binding of the Ku heterodimer (Ku70 & Ku80) to each end of the DSB. This forms a platform for the recruitment of DNA-PKcs which with Ku, forms the DNA-PK holoenzyme. DNA-PK then recruits the LIG4, XRCC4 and XLF or PAXX proteins to form the long-range synaptic complex to bridge two DSB ends closer together. DNA-PKcs then dissociates via autophosphorylation and conformational changes in the remaining protein complex forms the short-range complex, bringing two DSB ends in close enough proximity for ligation by LIG4. Following ligation, the remaining NHEJ factors dissociate and repair is completed.

2.1.10 Homologous recombination

Use of the NHEJ pathway to repair one-ended DSBs arising via replication fork collapse can result in chromosomal translocation and apoptosis (Balmus et al., 2019). Therefore, the homologous recombination (HR) pathway of DSB repair has evolved as the primary repair pathway for such lesions (Chapman et al., 2012, Tarsounas and Sung, 2020) **Figure 2-6**. HR is initiated by the resection of one strand of the double-stranded DNA present at the DSB end, to produce ssDNA overhangs (Cejka and Symington, 2021, Liu and Kong, 2021). DNA end-resection occurs in two steps, short-range and long-range, with the former being required for initiation of the latter (Cejka and Symington, 2021). Short-range resection is initiated by the removal of Ku, and binding of MRN in complex with phosphorylated CtIP to DSB ends (Liu and Kong, 2021). CtIP is phosphorylated by Cyclin/CDK complexes present in S/G2-phases only, restricting the initiation of end-resection to these cell cycle phases (Sartori et al., 2007, Himmels and Sartori, 2016, Huertas, 2010). MRE11 within the MRN-CtIP complex possesses endonuclease activity, and can generate a nick adjacent to the DSB (Anand et al., 2016, Cannavo and Cejka, 2014, Deshpande et al., 2016). This provides an entry point for the exonuclease activity of MRE11, which removes the DNA in a 3' - 5' direction towards the DSB, producing a short ssDNA overhang (Anand et al., 2016, Cannavo and Cejka, 2014, Deshpande et al., 2016). After this short-range resection, long-range resection can occur via two distinct pathways. The first pathway is dependent on the activity of the EXO1 nuclease (Mimitou and Symington, 2008, Garcia et al., 2011, Zhu et al., 2008), the second dependent on the BLM helicase in complex with the DNA2 nuclease (Zhu et al., 2008, Gravel et al., 2008, Mimitou and Symington, 2008). Long-range resection produces extensive ssDNA tracts, which are bound by the RPA protein complex (Raderschall et al., 1999). RPA is then replaced by the RAD51 recombinase, dependent on the activity of the BRCA1-PALB2-BRCA2 complex (Wong et al., 1998, Yang et al., 2002a, Jensen et al., 2010, Liu et al., 2010a, Thorslund et al.,

2010, Prakash et al., 2015, Kowalczykowski, 2015). RAD51 loaded ssDNA can invade the sister-chromatid to form a D-loop structure (Baumann and West, 1998), enabling the use of the sister-chromatid as a template for repair (Wright et al., 2018). As D-loop structures comprise interacting sister-chromatids, they must be resolved prior to mitosis to ensure the correct segregation of genetic material (Wright et al., 2018). Two primary pathways exist to resolve D-loop structures, namely, the synthesis-dependent strand annealing (SDSA) pathway and the double Holliday junction (dHJ) dissolution pathway (Wright et al., 2018).

SDSA is the preferred pathway of D-loop resolution in somatic cells (Bzymek et al., 2010). This process is carried out via the displacement of the invading strand which anneals to the other resected DSB end followed by DNA synthesis and ligation, resulting in a non-crossover event (Nassif et al., 1994, Nimonkar and Kowalczykowski, 2009). In the case of second-end capture, a dHJ forms, via lack of strand-displacement of the invading strand followed by annealing of the invading strand to the other resected DSB end (Nimonkar and Kowalczykowski, 2009). DHJs must be resolved prior to mitosis to prevent chromosome mis-segregation (Manthei and Keck, 2013). The primary mechanism by which cells process dHJs is via the BLM helicase, TOP3A, RMI1 and RMI2 (BTR) dissolution complex (Manthei and Keck, 2013), which produces non-crossover events (Singh et al., 2008, Wu and Hickson, 2003, Wu et al., 2006a). Alternatively, the structure-specific endonucleases SLX1-SLX4, MUS81-EME1 and GEN1 are able to resolve dHJs in human cells via DNA cleavage (Fekairi et al., 2009, Ip et al., 2008, Taylor and McGowan, 2008, Wyatt et al., 2013). These form two genetically distinct pathways dependent on SLX1-SLX4 and MUS81-EME1, or GEN1, and can result in crossover products dependent on the orientation of cleavage of the dHJ (Wyatt et al., 2013). Concomitantly, loss of BLM gives rise to an increase in sister-chromatid exchanges (SCEs) (Cunniff et al., 2017).

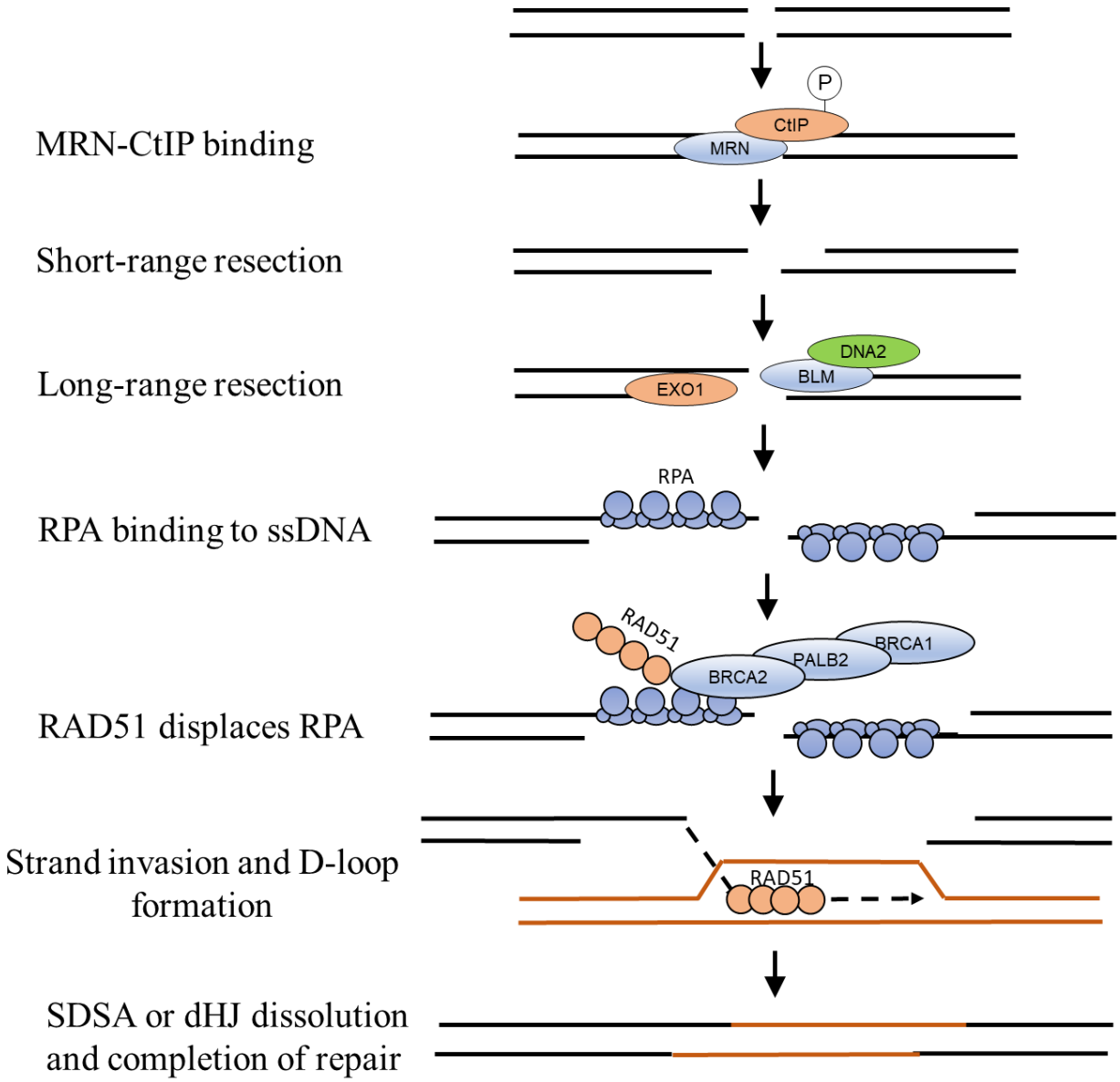


Figure 2-6: The homologous recombination pathway

HR is initiated by the binding of MRN in complex with phosphorylated CtIP to DSB ends. This complex performs short range resection driven by the nuclease activity of MRE11 in the MRN complex. The resulting short ssDNA overhangs provide a substrate for long-range resection, performed by one of EXO1 or BLM-DNA2. The extended ssDNA overhangs are then bound by RPA which is then replaced by the RAD51 recombinase dependent on the BRCA1-PALB2-BRCA2 complex. RAD51 bound ssDNA can then invade the intact sister chromatid to use as a template for repair, forming a D-loop. D-loop structures are then resolved by one of SDSA or the dHJ resolution or dissolution pathways.

2.1.11 Additional DSB repair pathways

In the absence of HR or NHEJ, alternative 'back-up' pathways of DSB repair can carry out repair; specifically, single-strand annealing (SSA) and microhomology-mediated end-joining (MMEJ). SSA can occur in the case of extensive resection at DSB ends promoted via BLM, EXO1 and DNA2 (Mendez-Dorantes et al., 2018). Extensive resection can reveal homology between two ssDNA regions. The two homologous stretches of DNA are then annealed via RAD52 (Gottifredi and Wiesmüller, 2020), which is followed by endonucleolytic cleavage of the non-homologous 3' ssDNA tails by ERCC1/XPF, and subsequent ligation (Motycka et al., 2004). MMEJ requires end resection via the MRN-CtIP complex to reveal 2-20bp of homology for annealing (Sfeir and Symington, 2015). Once annealed, the 3' ssDNA flaps are processed by FEN1 (Mengwasser et al., 2019) followed by POLQ polymerase activity filling in the gaps flanking the annealed region, with subsequent ligation by LIG1 or LIG3 (Sallmyr and Tomkinson, 2018).

2.1.12 DSB repair pathway choice

The decision to perform HR or NHEJ is controlled by resection at DNA ends as well as modifications to the chromatin surrounding the DSB (Scully et al., 2019). In response to DSBs, ATM phosphorylates the histone H2A variant, H2AX, on serine 139, forming γH2AX (Rogakou et al., 1998). MDC1 then binds to this phosphorylated residue via its C-terminal BRCT domains (Stucki et al., 2005). MDC1 is also constitutively phosphorylated by CK2, and this phosphorylation mark is recognised by the FHA-BRCT domains of NBS1, promoting the retention of the NBS1-MDC1 on damaged chromatin (Chapman and Jackson, 2008, Melander et al., 2008, Spycher et al., 2008). ATM then phosphorylates MDC1, with this mark recognised by the E3 ubiquitin ligase RNF8 (Huen et al., 2007, Kolas et al., 2007, Mailand et al., 2007). RNF8 acts in combination with its E2-conjugating enzyme UBC13, to ubiquitinate K63 on histone H1 (Thorslund et al., 2015). This mark is recognised by the E3 ubiquitin ligase

RNF168, which ubiquitinates K15 of H2A to form H2AK15ub (Mattiroli et al., 2012). 53BP1 binds to a combination of this mark and methylated H4 K20 (H4K20me1/2), formed during late G2-phase (Pellegrino et al., 2017, Simonetta et al., 2018, Botuyan et al., 2006, Fradet-Turcotte et al., 2013). 53BP1 is then phosphorylated by ATM (Jowsey et al., 2007), permitting the binding of the downstream effector proteins RIF1 (Chapman et al., 2013, Escribano-Díaz et al., 2013, Virgilio et al., 2013, Zimmermann et al., 2013) and the Shieldin complex, consisting of REV7, SHLD1, SHLD2 and SHLD3 (Boersma et al., 2015, Xu et al., 2015, Dev et al., 2018, Findlay et al., 2018b, Gao et al., 2018, Ghezraoui et al., 2018, Gupta et al., 2018, Noordermeer et al., 2018, Tomida et al., 2018). This complex counteracts DNA end-resection directing the cell towards the NHEJ pathway of DSB repair (Mirman et al., 2022, Ochs et al., 2019). In S phase, the dilution of methylation on H4K20 permits the recruitment of the BARD1-BRCA1 complex over 53BP1, via interaction of BARD1 with a combination of H2AK15ub and H4K20me0 (Becker et al., 2021, Hu et al., 2021, Krais et al., 2021, Nakamura et al., 2019). This enables the production of short ssDNA overhangs by the nuclease activity of MRN, in combination with CtIP (Sartori et al., 2007), directing the cell towards homology-based repair mechanisms (Cejka and Symington, 2021, Liu and Kong, 2021).

2.1.13 Responding to DSBs in mitosis

In contrast to G1, S and G2-phases, 53BP1 and BRCA1-BARD1 are not recruited to DSBs in mitosis (Blackford and Stucki, 2020). ATM driven phosphorylation of H2AX persists, promoting recruitment of MDC1 (Blackford and Stucki, 2020). In mitosis, CK2 mediated phosphorylation of MDC1 promotes recruitment of TOPBP1 (Leimbacher et al., 2019). This is recruited with CIP2A, to form filamentous structures which bridge two MDC1 foci together allowing for accurate chromosome segregation (Adam et al., 2021, Leimbacher et al., 2019).

2.1.14 Inter-strand crosslink repair

The repair of inter-strand crosslinks (ICLs) formed by endogenous aldehydes or crosslinking agents such as cisplatin is triggered when two converging replication forks stall approximately 20-40 base pairs away from an ICL (Räschle et al., 2008). Subsequently, TRAIP ubiquitinates the MCM2-7, Cdc45, GINS (CMG) replicative helicase complex to determine pathway choice for ICL repair (Wu et al., 2019). The addition of short ubiquitin chains promotes the recruitment of the NEIL3 DNA glycosylase, which acts to unhook one of the two *N*-glycosyl bonds forming the cross-link, permitting translesion synthesis and adduct removal (Semlow et al., 2016). Addition of longer ubiquitin chains promotes unloading of the CMG complexes by the p97 ATPase, which enables binding of the FANCM in complex with FAAP24, to the stalled replication forks (Kim et al., 2008). This interaction is stabilised by the histone fold proteins MHF1 and MHF2 (Singh et al., 2010, Yan et al., 2010). FANCM serves as the platform for the recruitment of the remainder of the Fanconi anaemia (FA) core complex (Kim et al., 2008) comprising FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, UBE2T, FAAP100 and FAAP20 (Shakeel et al., 2019).

At the ICL, the FANCL subunit acts as an E3 ubiquitin ligase in complex with the corresponding E2 ubiquitin-conjugating enzyme UBE2T, to monoubiquitinate the FANCD2-FANCI heterodimer on K561 and K523 respectively (Garcia-Higuera et al., 2001, Smogorzewska et al., 2007, Machida et al., 2006, Meetei et al., 2003, Meetei et al., 2004). Ubiquitination of the FANCI-FANCD2 heterodimer locks the FANCI-FANCD2 structure onto damaged DNA (Wang et al., 2020a, Alcón et al., 2020). Following this, SLX4 is recruited (Yamamoto et al., 2011) along with the XPF-ERCC1 and MUS81-EME1 structure-specific endonucleases (Fekairi et al., 2009, Muñoz et al., 2009). These work to perform incisions either side of the ICL, to unhook the ICL from one DNA strand (Silva et al., 2000, Niedernhofer et al., 2004, Klein Douwel et al., 2014, Hanada et al., 2006, McHugh, 2020).

The process of unhooking the DNA leaves the crosslinked nucleotide tethered to one strand, and a DSB in the other, with the crosslinked nucleotide needing to be bypassed (Räschle et al., 2008, Niraj et al., 2019). In order to do this, the FAAP20 subunit recruits the translesion polymerase REV1 (Budzowska et al., 2015, Kim et al., 2012), via an interaction between the UBZ4 domain of FAAP20 and the BRCT domain of REV1 (Lim et al., 2015). REV1 bypasses the crosslinked nucleotide using its deoxycytidyl transferase activity to insert dCMPs opposite an ICL (Haracska et al., 2001); or via its ability to recruit additional translesion polymerases, such as Pol ζ and Pol κ (Sharma and Canman, 2012, Guo et al., 2003). Following bypass, the nascent DNA strand is extended via a REV1-Pol ζ complex (Räschle et al., 2008, Budzowska et al., 2015) forming an intact DNA duplex. The unhooked ICL remnant will be removed by the NER pathway and the resulting DNA duplex can be used as a template by the HR pathway to repair the DSB in the sister strand (Niraj et al., 2019). Once repair is complete, USP1 in conjunction with UAF1 deubiquitinates FANCD2, and the pathway terminates (Nijman et al., 2005, Cohn et al., 2007, Ceccaldi et al., 2016) **Figure 2-7**.

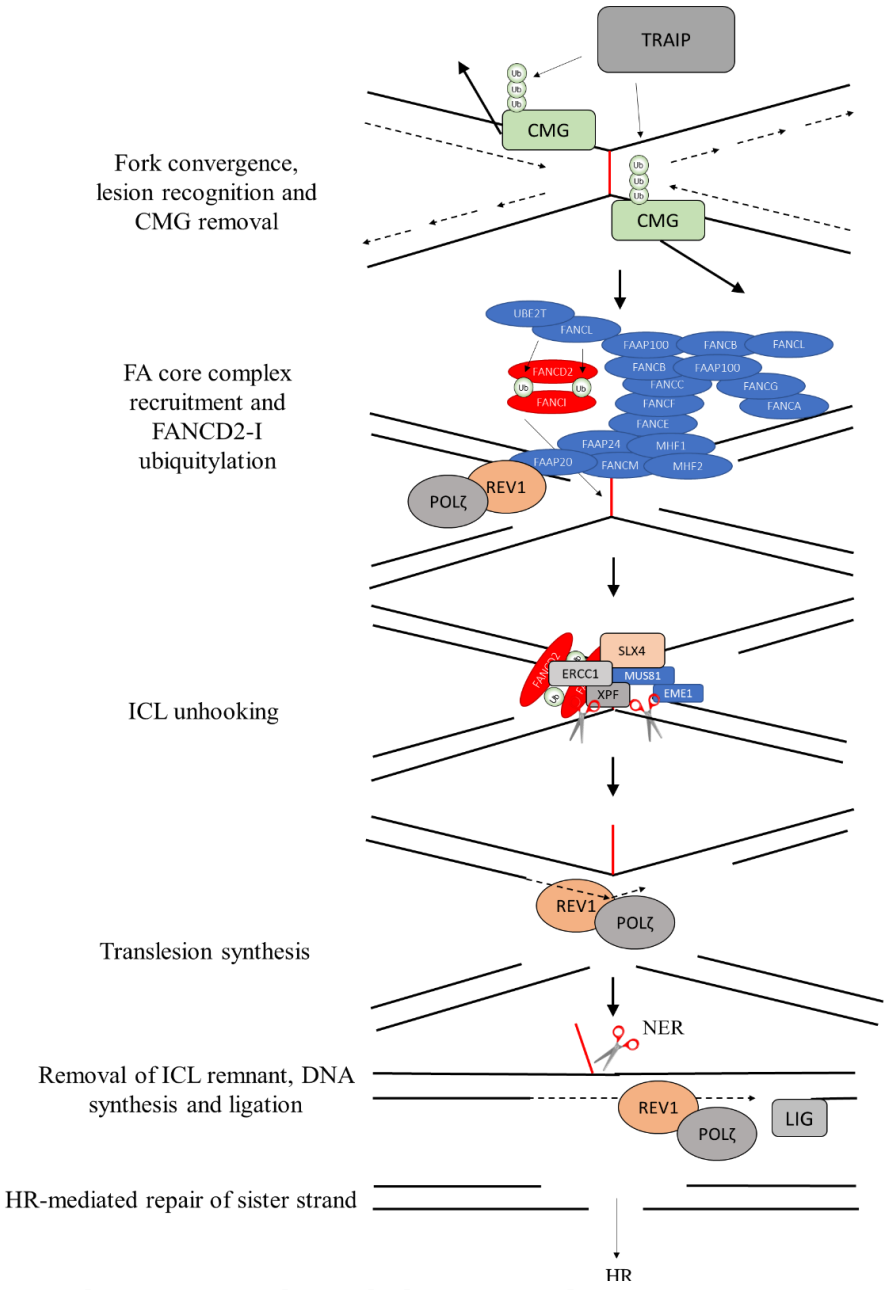


Figure 2-7: The inter-strand crosslink repair pathway

ICL repair is initiated by the convergence and stalling of two CMG helicase complexes at an ICL. The addition of long ubiquitin chains onto the CMG complexes by TRAIP leads to their removal. This enables to binding of the FA core complex via FANCM and FAAP24 to the ICL. FANCL of the FA core complex then ubiquitinates the FANCD2-FANCI heterodimer which locks it to the DNA at the site of the ICL. This enables the recruitment of SLX4 and the structure specific endonucleases XPF-ERCC1 and MUS81-EME1 which cleave the DNA either side of the ICL, unhooking it from one strand. The crosslinked nucleotide is bypassed by REV1 via its deoxycytidyl transferase activity or via its ability to recruit additional translesion polymerases such as Pol ζ. Subsequently, the nascent DNA strand is extended via a REV1-Pol ζ complex and ligated to form an intact DNA duplex and the ICL remnant is removed by the NER pathway. The DSB in the sister strand generate during unhooking is then repaired by homology based repair mechanisms such as HR using the intact sister chromatid as a template.

2.2 The DNA damage response in disease

Germline or somatic mutations in DDR genes can result in a wide range of pathologies including immunodeficiency, premature ageing, infertility and cancer (Jackson and Bartek, 2009, Ciccia and Elledge, 2010, Tiwari and Wilson, 2019). Furthermore, germline mutations in DDR genes give rise to multiple hereditary syndromes, characterised by the phenotypes mentioned above. (Jackson and Bartek, 2009, Tiwari and Wilson, 2019). Despite this, multiple chemotherapeutic compounds which disrupt the DDR have significantly improved prognosis in the clinic (Groelly et al., 2022). Therefore, whilst perturbation of DDR function is the underlying cause of multiple pathologies, it also provides treatment opportunities.

2.2.1 Immunodeficiency

The DDR is responsible for the processes of V(D)J recombination and class-switch recombination (CSR), which generate diversity within the adaptive immune system, facilitating the recognition of a wide-range of pathogens (Alt et al., 2013, Roth, 2014, Xu et al., 2012, Christie et al., 2022). V(D)J recombination is the process required for generation of diversity in the variable-region of the heavy-chain (Roth, 2014). This process centres around the fusion of random combinations of variable (V), diversity (D) and joining (J) coding segments (Alt et al., 2013, Christie et al., 2022). This is initiated by cleavage of recombination-signal sequences flanking the V, D and J coding segments by the RAG1 and RAG2 nucleases (Alt et al., 2013, Chi et al., 2020). The subsequent joining of V, D and J coding segments is carried out by the NHEJ pathway (Alt et al., 2013, Woodbine et al., 2014, Chi et al., 2020).

Class-switch recombination is the process responsible for determining the class of the immunoglobulin, e.g. IgM to IgG, by altering the constant-region of the antibody (Xu et al., 2012, Shang and Meng, 2021). This process involves the removal of the IgM defining (C)μ-region with a downstream Cα, Cγ or Cε coding for IgA, IgG or IgG isotypes respectively

(Stavnezer et al., 2008). This is achieved by the deamination of cytosine to uracil by activation-induced deaminase (AID) in the switch-regions flanking either end of the C-region (Xu et al., 2012, Shang and Meng, 2021). Uracil is then removed by uracil N-glycosylase, which in combination with APE1, generates a DSB (Xu et al., 2012, Stavnezer et al., 2008). This allows removal of a C-region from the Ig locus with the subsequent free DNA ends joined by the NHEJ machinery (Stavnezer et al., 2008, Woodbine et al., 2014). Concomitantly, loss-of-function (LOF) of NHEJ components such as Ku, DNA-PKcs and LIG4 result in immunodeficiency in patients and mouse models (Woodbine et al., 2014, Alt et al., 2013, Christie et al., 2022).

2.2.2 Infertility

The DDR plays a prominent role in the exchange of genetic material during meiotic recombination (Hunter, 2015). Specifically, proteins from the DDR mediate the crossing-over of chromosome homologs, permitting the exchange of genetic information (Hunter, 2015). Furthermore, they resolve structures arising from crossover events, ensuring correct chromosome segregation during the first meiotic division (Manhart and Alani, 2016). Meiotic crossing-over is initiated by the formation of DSBs by the SPO11 nuclease (Szekvolgyi et al. 2015). Subsequently, the endonuclease activity of the MRN complex in combination with CtIP, followed by EXO1 exonuclease activity, resects DNA ends producing 5’ overhangs on both sides of the DSB (Baudat et al., 2013). The strand-exchange proteins RAD51 and DMC1 then bind to overhangs in a BRCA2-dependent manner (Sharan et al., 2004), facilitating strand invasion of the homologous chromosome and D-loop formation (Marston and Amon, 2004). Two proteins from the MutS family, MSH4 and MSH5, play an important role in the formation and stabilisation of crossing-over intermediates and are specifically induced in meiosis (Ross-Macdonald and Roeder, 1994, Hollingsworth et al., 1995, Novak et al., 2001, Argueso et al., 2004). Studies in mouse spermatocytes and human oocytes show that MSH4 is present at all

DSB sites prior to crossover formation, with MSH4-MSH5 complexes present at the DSBs which will undergo crossover formation (Lenzi et al., 2005). In the absence of either MSH4 or MSH5, significant reductions in crossing-over are observed (Ross-Macdonald and Roeder, 1994, Hollingsworth et al., 1995, Novak et al., 2001, Argueso et al., 2004). Concomitantly, both *Msh4* and *Msh5* null mice are viable, but infertile (Edelmann et al., 1999, Kneitz et al., 2000). In addition to MSH4 and MSH5, MLH1 in partnership with MLH3 plays a fundamental role in crossover formation and double-Holliday junction resolution (Manhart and Alani, 2016). Yeast strains deficient in *Mlh1* show the same crossing-over defects as strains deficient in *Mlh1* and either *Msh4* or *Msh5* (Argueso et al., 2004, Hunter and Borts, 1997, Santos et al., 2003). Furthermore, Mlh1 foci colocalise with Msh4 foci at sites of crossover formation (Kolas et al., 2005). In addition to this, Mlh1-Mlh3 complexes are present at sites of crossover resolution (Lipkin et al., 2002, Moens et al., 2002).

2.2.3 Premature ageing

It has been argued that accumulation of DNA damage is the primary cause of ageing within organisms (Schumacher et al., 2021). However, individuals with germline mutations disrupting the proofreading activity of the replicative polymerases Pol ε and Pol δ, do not display a premature ageing phenotype, despite significant mutational burden (Robinson et al., 2021). Despite this, DNA damage is known to give rise to additional phenotypes associated with ageing such as stem cell exhaustion, mitochondrial dysfunction, inflammation, telomere dysfunction and cellular senescence (Schumacher et al., 2021; Yousefzadeh et al., 2021). Furthermore, premature ageing is a phenotype of multiple DDR syndromes including; Bloom syndrome (Manthei and Keck, 2013), Cockayne syndrome (Wang et al., 2020b), ataxia telangiectasia (Shiloh and Lederman, 2017), FA (Brosh et al., 2017) and Werner syndrome (Tsuge and Shimamoto, 2022).

2.2.4 Cancer

Disruption to DDR protein function can accelerate carcinogenesis, with genomic instability being one of the hallmarks of cancer (Hanahan and Weinberg, 2011). Genomic instability can result in increased mutational burden, increasing the likelihood of oncogene activation or loss of tumour suppressor function (Hanahan and Weinberg, 2011). Concomitantly, increased cancer predisposition is a prominent feature of multiple DDR syndromes (Kuhlen and Borkhardt, 2018). However, disruptions to DDR protein function can also provide treatment opportunities with direct inhibition of DDR protein function increasingly being utilised in the clinic (Groelly et al., 2022). Furthermore, disruption to DDR protein function in cancer cells provides opportunities for synthetic-lethality based approaches (Setton et al., 2021).

2.2.5 DDR syndromes with cancer predisposition

Predispositions to specific types of cancer arise depending on the DDR gene that is mutated. Deleterious mutations to MMR genes *MLH1*, *MSH2*, *MSH6* and *PMS2* give rise to Lynch syndrome (LS) (Papadopoulos et al., 1994, Bronner et al., 1994, Fishel et al., 1993, Leach et al., 1993, Nicolaides et al., 1994, Miyaki et al., 1997, Liu et al., 2001), also known as hereditary nonpolyposis colorectal cancer (HNPCC) (Lynch et al., 1966, Lynch et al., 2015). First described by Henry T. Lynch in the mid-1960's, LS is an autosomal-dominant disorder characterised by heightened risk of predominantly colorectal cancer (CRC), but also cancer of the endometrium, ovaries, stomach, kidneys, bladder and other tissues (Baretti and Le, 2018, Lynch et al., 1966). Approximately 1 in 100 CRC patients have LS (Chen et al., 2006). The two-hit hypothesis dictates that inheritance of one faulty allele means that only one additional hit (in the healthy allele) is required for total loss of protein function (Knudson, 1985). Concomitantly, LOF of the remaining wild-type *MLH1* allele is characteristic of tumour cells derived from LS patients (Hemminki et al., 1994). Thus, a definitive LS diagnosis is obtained by identification of a heterozygous deleterious germline mutation to any of the aforementioned

MMR genes(Edwards and Monahan, 2022) . Due to aberrant MMR protein function, tumours of LS patients typically display MSI and/or loss of MMR protein expression (Roudko et al., 2021). Concomitantly, such tumours display a 'mutator phenotype', characterised by increased mutation frequency and increased progression of oncogenesis (Loeb et al., 1974, Kunkel and Erie, 2005). Analysis of CRC reveals MMR-deficient tumours contain hundreds to thousands of somatic point mutations, including both frameshift and nucleotide substitutions, well in excess of those observed in MMR-proficient tumours (Yamamoto and Imai, 2015).. The National Comprehensive Cancer Network (NCCN) guidelines now endorse the MMR and/or MSI testing of all patients who have a personal history of CRC (Benson et al., 2021). Ultimately, diagnosis of LS centers around screening of the MMR genes *MLH1*, *PMS2*, *MSH2* and *MSH6* and the presence of MSI (Lynch et al., 2015, Umar et al., 2004, Giardiello et al., 2014, Baretti and Le, 2018, Edwards and Monahan, 2022). Diagnosis of LS results in frequent screening for CRCs, biennially from age 18 through 40 and then annually thereafter (Lynch et al., 2015). Frequent screening decreases overall mortality by 65% in HNPCC families (Järvinen* et al., 2000, Stoffel and Chittenden, 2010). Furthermore, LS patients displaying CRC are recommended to undergo subtotal colectomy, to reduce their incidence of metachronous CRC (Natarajan et al., 2010). Therefore, clarity regarding whether or not mutations to MMR genes are disruptive is of paramount importance to both diagnosis and subsequent management of LS (Georgiou et al., 2023).

Autosomal biallelic germline inactivation of any one of the 22 currently known FA genes (*FANCA-FANCW*) apart from *FANCB* which is X-chromosomal, results in FA (Niraj et al., 2019). FA patients display remarkable variability at the clinical level with regards to symptoms. There are patients with a multitude of congenital malformations, patients with early childhood leukaemia, medulloblastoma or Wilms tumour, and there are patients who reach adulthood with clinical and phenotypic features that are reminiscent of premature ageing (Moreno et al., 2021).

There are also adult patients with only subtle or no malformations, diagnosed as FA only because of infertility and short stature, or because they develop acute myeloid leukemia (AML) or squamous cell carcinomas (SCCs). Some patients are diagnosed only because they show an abnormally severe reaction to chemotherapy of adult-onset malignancy (Neveling et al., 2009).

Of the FA genes, *FANCA* is most commonly found mutated in FA patients comprising 57% of all identified and is followed by *FANCC* (15%) and *FANCG* (11%) (Gueiderikh et al., 2022). Additionally, heterozygous mutations in FA genes in non-FA patients can predispose to other types of neoplasm (Niraj et al., 2019). The most prominent example is monoallelic LOF mutations in either *BRCA1 (FANCS*) or *BRCA2* (*FANCD1*), characterised by a 40-80% increased risk of developing breast and ovarian cancer over a lifetime (Easton et al., 1995, Schubert et al., 1997, Burke et al., 1997). Furthermore, LOF mutations in *BRIP1* (*FANCJ*) predispose to ovarian cancer (Rafnar et al., 2011) and LOF mutations in *PALB2* (*FANCN*) or *RAD51C* (*FANCO*) predispose to breast cancer (Hofstatter et al., 2011, Vaz et al., 2010).

Other examples of DDR syndromes which predispose to cancer include xeroderma pigmentosum (XP) caused by biallelic mutations in *XP* genes and characterised by increased predisposition to skin cancer due to hypersensitivity to UV radiation (Piccione et al., 2021). Lastly, Bloom syndrome, caused by biallelic LOF of *BLM* causes predisposition to all types of cancer (German, 1993, Manthei and Keck, 2013, Cunniff et al., 2017).

2.2.6 Targeting the DDR in cancer

Cancer cells typically display relaxed DNA damage-sensing and/or repair mechanisms, and defective cell cycle checkpoints (Matthews et al., 2022). Concomitantly, in response to DNA damage, cancer cells may replicate and propagate damaged DNA, increasing the likelihood of cell death (Wang et al., 2021). Because of this, DNA damaging agents have been utilised in the treatment of cancers for over 100 years (Gianfaldoni et al., 2017). Well known examples include ionising radiation, cisplatin and etoposide which are all capable of inducing DSBs (Baskar et al., 2014, Cheung-Ong et al., 2013, Montecucco et al., 2015). Cancer cells are continuously proliferating and often display defective G1/S checkpoints and/or overexpression of Cyclin D1 or Cyclin E (Macheret and Halazonetis, 2015), meaning that they begin DNA replication with insufficient resources to do so (O'Connor, 2015). As a result, cancer cells often display replication stress (Macheret and Halazonetis, 2015, Zeman and Cimprich, 2014). This can result in the formation of extended tracts of ssDNA, which are coated by RPA, leading to subsequent ATR activation and replication fork stalling (Simoneau and Zou, 2021). Replication stress can be exacerbated via antimetabolites such as gemcitabine, 5-FU and 6-TG, which are frequently used in the clinic (Munshi et al., 2014, Sethy and Kundu, 2021). Such compounds reduce nucleotide availability and induce chain-termination, exacerbating replication stress by further reducing the speed of DNA synthesis (Ewald et al., 2008). This can result in replication fork stalling and subsequent collapse, leading to DSB formation and cell death via the ATM-CHK2-p53 axis (Ewald et al., 2008, Karnitz et al., 2005, Toledo et al., 2017). Antimetabolites can also incorporate into the DNA as nucleotide analogues, driving the formation of mismatched bases (Longley et al., 2003, Luengo et al., 2017, Swann et al., 1996, Waters and Swann, 1997, Ewald et al., 2008). After mismatch recognition via the MMR pathway, this pathway triggers cell cycle arrest and apoptosis via the ATR/CHK1 axis (Duckett et al., 1996, Cejka et al., 2003, Jiricny, 2006, Meyers et al., 2005). In MMR-deficient cells,

there is no attempt to process such mismatches, resulting in the survival of the cell at the cost of extensive mutagenesis (Jiricny, 2006). Concomitantly, multiple studies have shown that the presence of high-MSI inversely correlates to the effectiveness of 5-FU (Ribic et al., 2003, Sargent et al., 2010, Tejpar et al., 2011).

The treatments described thus far display significant off-target effects, damaging healthy tissue (Cheung-Ong et al., 2013, Baskar et al., 2014). This has led to the pursuit of treatments involving the DDR which show minimal toxicity to healthy tissue, whilst maximising killing of cancer cells (O'Connor, 2015, Topatana et al., 2020). Perhaps the most prominent example is LOF mutations in either *BRCA1* or *BRCA2*, sensitising cells to PARP-inhibitors (PARPi) (Farmer et al., 2005, Bryant et al., 2005). As explained previously, germline LOF mutations in either *BRCA1* or *BRCA2* predisposes to breast and ovarian cancer (Easton et al., 1995, Schubert et al., 1997, Burke et al., 1997). Typically, BRCA1 and BRCA2 are required for HR-based repair of one-ended DSBs which arise via replication fork collapse (Roy et al., 2012). In the absence of either protein, NHEJ is employed to repair such lesions, giving rise to chromosomal translocations and cell death (Balmus et al., 2019, Rose et al., 2020). It was widely believed that PARPi selectively killed BRCA-deficient cells by causing SSB accumulation via inhibition of PARP-dependent SSB repair, leading to DSB formation during replication (Bryant et al., 2005, Farmer et al., 2005). However, alternative mechanisms are beginning to emerge. PARPi are able to trap PARP complexes onto DNA, and their ability to do so correlates with their clinical potency (Murai et al., 2012). Furthermore, loss or inhibition of factors which decrease PARP-trapping such as ALC1, RNase H2 or p97 are synthetic-lethal with BRCA loss (Zimmermann et al., 2018, Hewitt et al., 2021, Juhász et al., 2020, Verma et al., 2021, Krastev et al., 2022). Furthermore, PARPi increase the number of ssDNA gaps behind the replication fork due to defective Okazaki fragment maturation, resulting in the formation of one-ended

DSBs in the subsequent S-phase (Vaitsiankova et al., 2022, Hanzlikova et al., 2018, Cong et al., 2021).

Despite their clinical efficacy, resistance to PARPi is a frequent occurrence in the clinic (Rose et al., 2020). Typically, this arises via reversion mutations in *BRCA1* or *BRCA2* which restore the reading frame and protein function (Lin et al., 2018, Pettitt et al., 2020). However, PARPi resistance can arise via suppression of NHEJ at one-ended DSBs (Noordermeer and Attikum, 2019). LOF mutations in *53BP1* or downstream factors *RIF1*, *REV7* or the Shieldin complex results in PARPi resistance in *BRCA1*-deficient cells (Bunting et al., 2010, Bouwman et al., 2010, Cao et al., 2009, Xu et al., 2015, Chapman et al., 2013, Noordermeer et al., 2018). These factors work to prevent resection at one-ended DSBs allowing NHEJ to occur in the absence of BRCA1 (Noordermeer and Attikum, 2019). Loss of any of these factors permits resection at one-ended DSBs preventing NHEJ and subsequent genomic instability. Resistance to PARPi via LOF mutations in *53BP1* or *REV7* in *BRCA1*-deficient cell lines can be overcome via inhibition of ATM (Bunting et al., 2010, Xu et al., 2015). Furthermore, ATM inhibition sensitises cells to PARPi in a manner dependent on toxic NHEJ at one-ended DSBs (Balmus et al., 2019). Loss of ATM delays end-resection at one-ended DBSs allowing a subset to be repaired via toxic NHEJ (Balmus et al., 2019). Importantly, *ATM* is often mutated in multiple cancers including lung, breast, brain or pancreatic cancer (Cremona and Behrens, 2014). Therefore, PARPi in combination with *ATM*-deficiency or inhibition is being widely explored in the clinic (Priya et al., 2023). In addition to ATM inhibition, inhibition of other DDR kinases ATR, DNA-PKcs, CHK1, WEE1 and PKMYT1 is also showing promise in the clinic (Groelly et al., 2022). Due to DNA-PKcs's role in NHEJ, inhibition of its activity sensitises cells to DSB-inducing agents (Fok et al., 2019). DNA-PKcs loss is synthetic-lethal with loss of ATM (Gurley and Kemp, 2001, Callén et al., 2009, Riabinska et al., 2013). Therefore, inhibition of DNA-PKcs in ATM-deficient tumours could provide an effective chemotherapeutic strategy.

In the absence of BRCA1, BRCA2 or ATM, POLQ mediated end-joining serves as a backup pathway for the repair of DSBs (Mateos-Gomez et al., 2015, Ceccaldi et al., 2015, Shima et al., 2004). To this end, loss or inhibition of POLQ is synthetic-lethal with loss of BRCA1, BRCA2 or ATM (Mateos-Gomez et al., 2015, Ceccaldi et al., 2015, Shima et al., 2004, Zhou et al., 2021, Zatreanu et al., 2021). Importantly, POLQ inhibition kills BRCA1-deficient cells in which 53BP1 is lost, overcoming PARPi resistance (Zatreanu et al., 2021).

Cancer cells displaying defective G1-S checkpoints are dependent on ATR signalling to manage replication stress in S-phase (Cliby et al., 1998, Nghiem et al., 2001, Toledo et al., 2013). Therefore, inhibition of ATR in cancer cells of this nature results in cytotoxicity derived from fork stalling and collapse (Mei et al., 2019). Furthermore, ATR and ATM show some degree of functional redundancy, leading to ATR-inhibition selectively killing ATM-deficient cancer cells (Tu et al., 2018). ATR-inhibition is also effective in killing PARPi-resistant BRCA1-deficient tumour cells, by preventing the BRCA1-independent loading of RAD51 to DSBs (Yazinski et al., 2017). Inhibition of CHK1, a downstream target of ATR also produces cytotoxicity in cancer cells experiencing high levels of replication stress (Brooks et al., 2013, Ferrao et al., 2012). CHK1 phosphorylates WEE1 to allow it to phosphorylate CDK1/CDK2, preventing entry into mitosis as well as upregulating RRM1/2 and suppressing dormant origin firing (Tominaga et al., 2006, Beck et al., 2010, Beck et al., 2012, Pfister et al., 2015). WEE1-inhibition and CHK1-inhibition act synergistically to increase apoptosis and levels of γH2AX (Carrassa et al., 2012, Guertin et al., 2012). The mechanism proposed is that loss of CHK1 and WEE1 increases DNA damage whilst abrogating the normal schedule of mitotic entry, leading to cell death via mitotic catastrophe (Carrassa et al., 2012, Guertin et al., 2012). Cancers deficient in SETD2 have reduced levels of RRM2, leading to a shortage of dNTPs and replication stress (Pfister et al., 2015). Inhibition of WEE1 in SETD2-deficient cells causes further RRM2 loss, resulting in critical dNTP depletion, S-phase arrest and apoptosis (Pfister

et al., 2015). Furthermore, inhibition of the WEE1 like kinase PKMYT1 is synthetic-lethal in tumours with elevated levels of CCNE1 expression (Gallo et al., 2022). Inhibition of PKMYT1 induces unscheduled CDK1 activation, premature mitotic entry and catastrophic genomic instability (Gallo et al., 2022).

The WRN helicase has recently been identified as a synthetic lethal target in cells displaying MSI (Chan et al., 2019). Mechanistically, WRN resolves cruciform structures arising from large regions of TA repeats in cells displaying MSI (Wietmarschen et al., 2020). In the absence of WRN, the MUS81-EME1-SLX4 endonuclease complex cleaves these structures resulting in chromosomal fragmentation and apoptosis (Wietmarschen et al., 2020). Additionally, frameshift mutations in protein coding regions arising via MSI can lead to the formation of neoantigens (Baretti and Le, 2018). Indeed, inactivation of MLH1 in colorectal, breast and pancreatic mouse cancer cells causes persistent renewal of neoantigens over time (Germano et al., 2017). Concomitantly, PD-1 blockade is an effective treatment for tumours displaying MSI or deficient MMR (Le et al., 2015, Le et al., 2017).

Overall, loss of DDR protein function is a common feature of oncogenesis, but also provides a multitude of therapeutic opportunities. Therefore, determining whether or not the function of a DDR protein has been disrupted is of paramount importance for determining both treatment and patient outcome. One way in which protein function can be lost is via missense mutations, resulting in the appearance of alternative amino acids, termed ‘variants’, in the polypeptide chain.

2.3 Identifying, cataloguing and classifying variants

2.3.1 Amino acid variants

Variant amino acids arise via missense mutations, defined as base pair substitutions that alter the amino acid coded for by the codon (Stefl et al., 2013). Missense mutations can arise via several mechanisms including spontaneous base deamination, hydroxylation of DNA bases by ROS, DNA polymerase errors during DNA replication and DNA repair by error-prone DNA polymerases (Yonekura et al., 2009, Cadet et al., 2011, Loeb and Monnat, 2008). Reductions in the time and cost of whole-genome sequencing have led to a plethora of variants being identified in both coding, and non-coding areas of the genome (Niroula and Vihinen, 2016, Shendure et al., 2017). It is now estimated that each human genome contains approximately 3 million single nucleotide polymorphisms (SNPs) resulting in on average 11,000 amino acid variants in comparison to the reference genome (Niroula and Vihinen, 2016, 2010). Recently, it was claimed that there will be a variant for every amino acid position (asides from lethal LOF mutations) somewhere in the human population of almost 8 billion (Frazer et al., 2021).

2.3.2 Identifying amino acid variants

Missense mutations are identified via DNA sequencing (Logsdon et al., 2020). Early attempts to sequence DNA relied on primer-extension which was extremely cumbersome and took months to sequence 10-20 bases (Wu and Kaiser, 1968). In the 1970's, the advent of Sanger and Gilbert sequencing based on gel-electrophoresis enabled 100s of bases to be sequenced in a single afternoon (Sanger et al., 1977, Gilbert and Maxam, 1973, Maxam and Gilbert, 1977). In the 1980's, Sanger sequencing machines were developed allowing the sequencing of 1000 bases per day (Smith et al., 1986). This lead to the curation of databases such as GenBank and search tools like BLAST allowing researchers to store and search sequencing data obtained (Altschul et al., 1990). Such technology eventually enabled the curation of the genomes of several organisms including *Haemophilus influenza* (Fleischmann et al., 1995), *Caenorhabditis*

elegans (Consortium*, 1998) and finally *Homo sapiens* (Tanaka et al., 1989, 2001b, 2001a, Consortium, 2004). At the time the human genome project was completed, DNA sequencing technology had improved to a point where 600-700 base pairs of DNA could be sequenced for $1 (Shendure et al., 2017). The technology continued to improve with the advent of next-generation sequencing (NGS) in the mid 2000's (Shendure et al., 2005). Such technology moved away from the electrophoretic methods employed in Sanger sequencing; instead relying on the immobilisation of a complex library of DNA templates on a 2D surface, followed by sequencing by ligation or sequencing by synthesis (Goodwin et al., 2016, Shendure et al., 2017, Shendure et al., 2005). To date, NGS is capable of obtaining a terabase of sequence over 2 days for a cost of ~$5000, exceeding the total amount of data obtained for the human genome project by a factor of 40 (Shendure et al., 2017, Goodwin et al., 2016).

Among the first variants identified to be disease-causing were missense mutations in the *CFTR* gene (Kerem et al., 1990), giving rise to cystic fibrosis, and in *BRCA1* for breast cancer risk (Hall et al., 1990). The improvement in the speed, depth and quality of DNA sequencing means DNA sequencing is widely used in the clinic to facilitate the diagnosis of many more diseases (Prokop et al., 2018, Nangalia and Campbell, 2019). Initially, this was restricted to the gene of interest, e.g. *CFTR*, but then progressed to whole-exome sequencing (WES) of an array of genes known to be involved in disease progression (Bamshad et al., 2011, Choi et al., 2009, Nangalia and Campbell, 2019). Use of exome sequencing in the clinic continues to grow (Gilissen et al., 2011, Yang et al., 2014). To date, it is estimated that WES allows the diagnosis of ~25% of rare diseases (Prokop et al., 2018, Jacob et al., 2013, Yang et al., 2014). Further reductions in the cost of sequencing have resulted in an increased use of whole-genome sequencing (WGS), which enables a more accurate determination of protein-coding variants (Martin et al., 2014, Saunders et al., 2012). Perhaps the most prominent example of genome

sequencing to identify causal variants in cancer, rare diseases and infectious diseases on a large scale, is the recently completed 100,000 genomes project (Turnbull et al., 2018).

In the clinic, rigorous pipelines are in place to ensure accurate identification of both inherited and de novo germline variants which may be pathogenic (Krusche et al., 2019, Koboldt, 2020). Specifically, trio sequencing offers a powerful diagnostic tool to narrow down potentially pathogenic germline mutations in the affected individual (Koboldt, 2020, Gabriel et al., 2022). Trio sequencing centres around the sequencing of the maternal, paternal and affected progenies DNA, be it via gene panel, WES or WGS (Koboldt, 2020). Variants only present in the maternal or paternal sequencing data can assumed as benign. However, variants only present in the affected individuals DNA are classified as de novo variants and may be disease causing. It is estimated that the human de novo mutation rate results in approximately 70 de novo variants on a background of 50,000 inherited variants (Besenbacher et al., 2015, Jónsson et al., 2017). Whilst NGS technologies have an accuracy of >99%, it is estimated that there will be approximately 50 false positive calls for each genuine de novo mutation (Koboldt, 2020). Despite this, trio sequencing offers a 10-15% increase in diagnostic yield compared to just sequencing the affected individual (Tan et al., 2019, Farwell et al., 2015). Furthermore, a variant present in one allele of maternal and paternal DNA but present in both alleles of the progeny may also be considered as potentially disease causing (Kuhlen and Borkhardt, 2018).

Whilst trio analysis offers a powerful method of identifying potentially pathogenic germline mutations, it cannot be utilised to identify somatic variants occurring specifically in a patients tumour (Giardina et al., 2018). To identify such variants, some laboratories have focused purely on sequencing DNA from the tumour sample (Pritchard et al., 2014, Wagle et al., 2012). More modern approaches rely on sequencing both the tumour sample but also healthy tissue from the affected individual (Koboldt, 2020, Horak et al., 2022). Following this, algorithms such as MuTect2 (Cibulskis et al., 2013) or VarScan2 (Koboldt et al., 2012) are employed to determine

variants present in the tumour DNA but not the healthy tissue. Due to the genetic heterogeneity within tumours, mutations may be present at low frequency and thus be harder to detect (Chen et al., 2020). A recent analysis of sequencing depth vs mutation frequency revealed that for mutations present at ≥ 20% within the tumour, a sequencing depth of ≥ 200 is sufficient for identifying 95% of the mutations (Chen et al., 2020). Some studies have revealed that the algorithms used for calling somatic variants differ widely in their performance, especially for mutations present at ≤ 10% frequency (Jin et al., 2022, Xu, 2018). Therefore, it is recommended that a combination of algorithms are used to ensure robustness of variant calling (Koboldt, 2020, Horak et al., 2022). More recently, sequencing analysis of tumour tissue and healthy tissue from the same individual has revealed thousands of somatic variants present in tumour samples (Pleasance et al., 2010b, Pleasance et al., 2010a, Ley et al., 2008). Such analyses have also paved the way for identification of 'mutational signatures', defined as combinations of mutation types present in tumour cells arising due to inherent defects in genome stability or exposure to genotoxins (Forbes et al., 2017). Recently, identification of mutational signatures within *MLH1*, *MSH2* or *MSH6* null cells allowed the creation of a mutational signature, which accurately detects MMR deficiency in a panel of ~7500 whole-genome sequenced cancers (Zou et al., 2021).

Another way in which variants are identified is through genome-wide association studies (GWAS) (Tam et al., 2019) First performed in 2005 (Klein et al., 2005), GWAS sequence entire genomes or exomes from individuals with a disease and disease free individuals (Visscher et al., 2017). The overall aim of this approach is to identify SNPs present at higher frequencies in disease affected individuals compared to healthy controls (Uffelmann et al., 2021). To date, more than 5700 GWAS have been carried out (Prokop et al., 2018, Uffelmann et al., 2021). Results from GWAS are used to formulate polygenic risk scores, which hold the same predictive power as monogenic risk prediction using highly penetrant SNPs (Khera et al.,

2018). Due to the sheer number of SNPs identified in GWAS, the exact contributions of individual SNPs to pathogenesis often remains unclear (Tam et al., 2019). Studies have also been performed to identify and curate variants in the general, healthy population. Identification of common variants in this way enables their exclusion as disease causing (Gudmundsson et al., 2022, Richards et al., 2015). The first study of this manner was the 1000 genomes project (2010). This was succeeded by a slightly larger study, termed the Exome Sequencing Project, where 6500 individuals of European and African American descent underwent whole-exome sequencing (Fu et al., 2013). Subsequently, a ground-breaking study collated the exome sequencing data from 60,000 individuals from a diverse range of genetic backgrounds to form Exome Aggregation Consortium (ExAC) database, providing a reference point for diverse genetic information in healthy individuals (Consortium et al., 2016). This was expanded upon with the addition of genomic data, to form the Genome Aggregation Database (gnomAD) with sequencing data from over 195,000 healthy individuals to date (Karczewski et al., 2020).

2.3.3 Cataloguing amino acid variants

The capability to sequence entire genomes within a matter of days has led to the requirement of multiple databases to curate this information. Their overall aim is to provide easy access to variants discovered and the associated information (Peterson et al., 2013, Landrum et al., 2017, Landrum and Kattman, 2018). One of the first databases of genetic variation to be curated was the dbSNP from the National Centre for Biotechnology Information (NCBI) (Sherry et al., 2001). This catalogues all variation, regardless of disease association or genomic location (Sherry et al., 2001). To aid clinicians and researchers in diagnosis of disease, additional databases were curated to catalogue variants identified in disease-affected individuals. Prominent examples include Online Mendelian Inheritance in Man (OMIM) (McKusick, 2007), ClinVar (Landrum et al., 2014), UniProt (Boeckmann et al., 2003) and the Human Gene Mutation Database (HGMD) (Stenson et al., 2012), but a multitude of other databases also exist

(Licata et al., 2023). In addition to pan-genome variant databases, databases which curate information for specific diseases and/or their associated genes also exist These include MITOMAP for mitochondrial genes (Ruiz-Pesini et al., 2007), PDmutDB for Parkinson's disease (Nuytemans et al., 2010), IDbases for immunodeficiency (Piirilä et al., 2006) and the Catalogue of Somatic Mutations In Cancer (COSMIC) (Forbes et al., 2011) and the Cancer Genome Atlas (TCGA) (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga) for cancer associated mutations. Importantly, databases specific to variants in DDR genes have also been curated. In 1997, the International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer (ICG-HNPCC) established the international database of mutations identified in HNPCC patients (Peltomaki and Vasen, 1997). The majority of mutations were present in MSH2 and MLH1. Intriguingly, whilst the majority of MSH2 mutations were frameshift (60%) or nonsense (23%), MLH1 demonstrated missense mutations in 31% of cases (Peltomaki and Vasen, 1997). Further efforts to catalogue variants in MMR genes led to the formation of the International Society for Gastrointestinal Hereditary Tumours (INSiGHT), a merger of ICG-HNPCC with the Leeds Castle Polyposis Group (LCPG) (Peltomäki and Vasen, 2004). This led to the curation of the INSiGHT database, which at the time contained 448 mutations in MMR genes, 50% of which were found in MLH1, with 44% being missense (Peltomäki and Vasen, 2004). Additionally, other databases such as the BRCA exchange have been set up to curate variants in *BRCA1* and *BRCA2* (Cline et al., 2018). As discussed above, databases also exist to catalogue variants identified in healthy individuals, such as gnomAD (Karczewski et al., 2020).

2.3.4 Classifying amino acid variants

It is essential to determine the functional impact of variants identified both in the clinic and through research studies such as GWAS, to aid patient stratification and subsequent genetic counselling (Stein et al., 2019). To facilitate this, groups of experts have designed guidelines

to provide a clinical classification for any given variant. Initially, it was recommended that upon detection of a variant, clinicians should determine whether or not this variant is already known to be disease-causing by searching databases such as HGMD (Kazazian et al., 2000). Furthermore, it was suggested that clinicians could establish whether or not the amino acid was evolutionarily conserved, or present at consensus splice sites within the gene (Kazazian et al., 2000). This was expanded upon to include predictions from bioinformatic algorithms, and results from lab-based functional assays (Kazazian et al., 2000, Richards et al., 2008). This culminated in the creation of the standards and guidelines for the interpretation of sequence variants by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP), in 2015 (Richards et al., 2015). This consortium designed a 5-tier classification system based on scores achieved by combining data from sources such as population-based studies, trio/familial sequencing, lab-based functional assays and computational data (Richards et al., 2015) **Table 2-2**, **Table 2-3**. Laboratories submitting variants identified in patients to databases such as ClinVar, are recommended to follow such guidelines when determining the functional impact of the variant identified (Richards et al., 2015, Landrum et al., 2017, Landrum and Kattman, 2018). These guidelines continue to be used to form the basis of more specific guidelines for variant interpretation. Recently, this framework was used to design a classification scoring system specifically for somatic mutations identified in the clinic (Horak et al., 2022). Importantly, evidence generated in *in vivo* laboratory-based assays is considered as 'strong' data for classification within this framework (Horak et al., 2022).

Whilst fundamental to modern variant classification, the ACMG/AMP guidelines and their derivatives are a generic framework which may miss gene-specific nuances with disease association. To improve the application of the criteria, the Clinical Genome Resource (ClinGen) Consortium developed panels of experts associated with a particular disease or

variant classification techniques (Rehm et al., 2015) This has led to the formation of improved overall criteria for variant interpretation (Brnich et al., 2019, Biesecker et al., 2018, Ghosh et al., 2018) as well variants in specific disease contexts and/or genes (Oza et al., 2018, Mester et al., 2018, Kelly et al., 2018). To highlight whether or not an expert panel was involved in the assessment of a variant, ClinVar developed a scoring system to demonstrate the level of interpretation a variant had received (Landrum and Kattman, 2018) **Table 2-4**. Importantly, consortia have been developed which study the impact of variants in DDR genes. Notable examples include the Evidence-based Network for the Interpretation of Germline Mutation Alleles (ENIGMA), who focus on interpretation of variants in *BRCA1*, *BRCA2* and other known breast cancer-predisposing genes (Spurdle et al., 2012) and the InSiGHT consortium who work to classify variants in LS associated genes such as *MLH1* (Thompson et al., 2014). Importantly, both the ENIGMA and InSiGHT consortia are now considered expert panels by ClinGen. It must be mentioned that the InSiGHT team developed their own classification system for interpretation of variants in LS associated genes (Thompson et al., 2014). These have since been implicated into the ACMG-AMP guidelines for variant interpretation.

Strength		
	Pathogenic	**Benign**
Stand alone		- Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium (now gnomAD).
Very strong	- Null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease.	
Strong	- Well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product. - *De novo* (both maternity and paternity confirmed) in a patient with the disease and no family history.	- Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age. - Well-established *in vitro* or *in vivo* functional studies show no damaging effect on protein function or splicing.
Moderate	- Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation. - Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before.	
Supporting	- Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease. - Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.)	- Multiple lines of computational evidence suggest no impact on gene or gene product (evolutionary conservation, splicing impact, etc.) Caveat:, each algorithm cannot be counted as an independent criterion. - Missense variant in a gene for which primarily truncating variants are known to cause disease.

Table 2-2: Description of the ACMG/AMP criteria used for variant interpretation

This table depicts criteria devised by the ACMG/AMP for variant interpretation (adapted from Richards et al., 2015). Criteria fall under different strength categories which are then given certain weighting when it comes to variant classification. Note, not all criteria listed by the ACMG/AMP are described here, just criteria that were deemed relevant for this thesis.

Table 2-3: ACMG/AMP scoring system used for variant classification

Classification	Criteria needed
Pathogenic	1 Very strong AND - ≥ 1 strong OR - ≥ 2 moderate OR - 1 moderate and 1 supporting OR - ≥ 2 supporting ≥ 2 Strong 1 Strong AND - ≥3 moderate OR - 2 moderate AND ≥ 2 supporting OR - 1 moderate AND ≥4 supporting
Likely pathogenic	1 Very strong AND 1 moderate 1 Strong AND 1-2 moderate 1 Strong and ≥ 2 supporting ≥ 3 Moderate 2 Moderate AND ≥ 2 supporting 1 Moderate and ≥ 4 supporting
Likely benign	1 Strong and 1 Supporting ≥ 2 Supporting
Benign	1 Stand alone ≥ 2 Strong
Uncertain significance/Conflicting interpretations	Criteria shown are not met. Criteria for variant from benign and pathogenic categories are contradictory.

This table depicts the scoring system devised by the ACMG/AMP for variant classification (adapted from Richards et al., 2015). Criteria in **Table 2-2** fall under different strength categories which are then given certain weighting when it comes to variant classification depicted here.

Number of stars	Review status	Description
Four	Practice guidelines.	Practice guidelines have been met as well as expert panel and criteria provided – single submitter criteria.
Three	Reviewed by an expert panel.	Variant is reviewed based on ACMG/AMP guidelines by an expert panel of scientists, representing clinical, diagnostic laboratory and basic research areas. Membership should include representation from three or more institutions and will encompass disease/gene expert members as well as biocurators. Submitters must also define how to resolve and review newly submitted conflicting assertion criteria within 4 months of being notified of the conflict. This must all be reviewed and confirmed by the Clinical Genome Resource programme (ClinGen) before acceptance.
Two	Criteria provided, multiple submitters, no conflicts.	Two or more submitters with assertion criteria provided the same interpretation.
One	Criteria provided, conflicting interpretations.	Multiple submitters provided assertion criteria and evidence but there are conflicting interpretations.
One	Criteria provided, single submitter.	One submitter provided an interpretation with assertion criteria.
None	No assertion for the individual variant.	The variant was not interpreted directly in any submission; it was submitted to ClinVar only as a component of a haplotype or a genotype.
None	No assertion criteria provided.	The variant was included in a submission with an interpretation but without assertion criteria and evidence.
None	No assertion provided.	The variant was included in a submission that did not provide an interpretation.

Table 2-4: Description of scoring system for ClinVar submissions

Description of criteria needed to obtain certain classifications on ClinVar. The higher the number of stars given the higher quality the submission is deemed. It provides a way for users to assess the quality of the variants interpretation on which to base their assessment of patient samples or any new research findings. From ClinVar (Landrum et al., 2014).

2.3.5 The importance of identifying amino acid variants and their functional impact

Variant amino acids can disrupt protein activity via several means, such as alteration of protein conformation, or loss of a PTM site required for protein regulation and/or function (Vihinen, 2021) Concomitantly, whether or not a variant disrupts protein activity is largely dependent on the biochemical nature and location of the variant within the polypeptide chain (Stefl et al., 2013). In the clinic, the use of databases containing information on the functional impact of variants is becoming increasingly common and is frequently included in best practice guidelines (Marshall et al., 2020, Koboldt, 2020, Horak et al., 2022, Landrum and Kattman, 2018). As disease is often driven by disruption to protein function, identifying which amino acid variants perturb protein activity can prove crucial for diagnosis and subsequent treatments but also the genetic counselling of affected families (Peterson et al., 2013, Manolio et al., 2017, Stein et al., 2019, Lappalainen and MacArthur, 2021).

The identification of germline variants and understanding their functional impact can have significant ramifications on both parents and progeny. In the case of trio sequencing, the identification of a variant in an unaffected parent is considered strong evidence of benignity by the ACMG/AMP framework (Richards et al., 2015). On the other hand, identification of a de novo variant in the affected proband is considered as strong data for pathogenicity (Richards et al., 2015). However, in order to achieve ‘pathogenic’ or ‘benign’ status on ClinVar, these findings must be combined with an additional strong criterion such as *in vivo* functional evidence of variant effect via lab-based functional assays (Richards et al., 2015). In addition to identifying de novo variants, determining the inheritance of two copies of a pathogenic variant via trio sequencing can benefit both the proband and parents (Kuhlen and Borkhardt, 2018). In the context of the DDR, this is highlighted by the example of children diagnosed with constitutional mismatch repair deficiency, resulting from biallelic pathogenic germline mutations in one of the four MMR genes (MLH1, MSH2, MSH6 or PMS2) (Wimmer et al.,

2014). Whilst a diagnosis will provide a clear treatment plan for the affected proband, it will also mean that the transmitting parents are at risk of tumours on the LS spectrum, as LS is diagnosed by the presence of a heterozygous germline pathogenic MMR variant (Edwards and Monahan, 2022). Upon diagnosis, annual colonoscopies from the age of 25 are often recommended (Giardiello et al., 2014). Therefore, identifying which variants result in LOF via trio sequencing combined with lab-based functional assays would enable more accurate diagnosis and genetic counselling, whilst also preventing unnecessary invasive screenings from a young age (Duraturo et al., 2019). In addition to MMR related disorders, this is also highlighted by disorders such as FA, which can arise via homozygous LOF mutations in *BRCA1* (*FANCS*) or *BRCA2* (*FANCD1*) (Niraj et al., 2019). Whilst FA patients are severely affected, transmitting non-FA parents are diagnosed with hereditary breast/ovarian cancer syndrome, which again can result in preventative surgeries/frequent tumour surveillance (Yoshida, 2021). Overall, trio sequencing combined with experimental understanding of variant effect provides significant power in both diagnosis and prognosis, whilst facilitating genetic counselling and treatment plans/tumour surveillance.

The identification of somatic variants and understanding their functional impact is also of paramount importance in the clinic (Giardina et al., 2018). Whilst LOF variants can be inherited, they can also occur somatically and drive tumour formation and progression (Wang et al., 2022b, Elze et al., 2021, Walker et al., 2023). One example in the context of the DDR is the finding that the majority of MMR-deficient colorectal and endometrial cancers which did not contain somatic MLH1 promoter hypermethylation or pathogenic MMR germline variants are caused by somatic MMR mutations (Elze et al., 2021, Walker et al., 2023). These findings demonstrated a high diagnostic yield of somatic MMR testing and offers the relatives of patients assurance that the MMR deficiency in the affected individual is driven by somatic mutations and not germline, meaning an LS diagnosis is unlikely (Elze et al., 2021, Walker et

al., 2023). In addition to providing details on potential familial risk to cancer, the identification of somatic mutations can also inform the treatments given to affected individuals. As discussed in section 2.2.6, the effective application of certain cancer therapies targeting the DDR are dependent on LOF in DDR genes (Groelly et al., 2022). Therefore identifying mutations occurring specifically in tumour cells will enable the effective application of DDR-based therapies targeting LOF and reducing off-target effects. Lastly, identifying disruptive amino acid variants can also provide insight into protein function, regulation and protein-protein interactions. which can further our understanding of how cellular processes are regulated, revealing new targets for therapeutic intervention (Meng et al., 2021, Lu et al., 2020). Overall, it is crucial to gain understanding of how variants impact protein function to inform patient prognosis and treatment but also provide accurate genetic counselling to the affected individuals and family members. Significant efforts have been made in this regard using lab based functional assays and a discussion on this is provided in the next section.

2.4 Methods for assessing the functional impact of variants

As evidence associated with a variant enables its classification (Richards et al., 2015), significant efforts have been made by researchers to understand the functional impact of variants. Researchers utilise multiple methods ranging from *in vivo* functional studies to *in silico* predictions (Stein et al., 2019, Ghosh et al., 2017, Wang et al., 2022a). All methods present advantages and disadvantages and their findings can differ significantly with one another (Stein et al., 2019, Ghosh et al., 2017, Peterson et al., 2013). Here a discussion is provided on some of the methods used, with a focus on attempts to interpret variants in DDR proteins.

2.4.1 In vitro assays

In vitro assays have been used extensively to study variants in MMR genes. These assays centre around transfection of cells with either WT or variant containing cDNA for the protein of interest. Cell lysates are then prepared and incubated with a mismatch containing substrate. Proficient activity is marked by removal of the mismatch *in vitro,* restoring an endonuclease site which can be detected via gel-electrophoresis (Plotz et al., 2006, Borràs et al., 2012, Drost et al., 2010, Kosinski et al., 2010, Nyström-Lahti et al., 2002, Raevaara et al., 2005, Hinrichsen et al., 2013). The strength of such an *in vitro* assay was validated recently with comparison of the complete *in vitro* MMR assay (CIMRA) to cellular data for *MLH1, MSH6, MSH2* and *PMS2* (Drost et al., 2019, Drost et al., 2020, Rayner et al., 2022). The authors found high concordance between the CIMRA assay and cell-based studies of MMR variant effects. Limitation of such approaches include lack of assessment of cellular localisation and an inability to assess splicing defects (Rayner et al., 2022). However, the speed and ease of such an assay makes it an attractive option for variant assessment in MMR proteins. It must be noted that *in vitro* findings are not always consistent with one another and with studies from *in vivo* systems. MMR appears to operate differently *in vitro* and *in vivo,* highlighted by the *in vitro*

proficiency, but *in vivo* deficiency, of MutS mutants unable to form sliding clamps (Graham et al., 2018b).

2.4.2 Yeast-based assays

Yeast-based assays have been used extensively to study variant effects in DDR proteins including MLH1, BRCA1, MSH2, RAD52, BRCA2, KU70 and POL ε and POL δ (Cervelli et al., 2020). Introduction of mutations into the correspondent alleles of human *RAD52* in yeast *Rad52*, followed by assessment of the response to methyl methanesulfonate (MMS) accurately identifies pathogenic human variants in this gene (Lee et al., 2015b). Expression of constructs containing the C-terminus of BRCA1 in yeast results in the 'small colony phenotype' (SCP) (Humphrey et al., 1997, Coyne et al., 2004). The BRCTs are the region required for this phenotype with inactivating mutations in these domains permitting wild-type growth levels (Coyne et al., 2004). This technique enables the testing of multiple variants within *BRCA1* (Humphrey et al., 1997, Coyne et al., 2004, Thouvenot et al., 2016). Furthermore, expression of human *BRCA2* in yeast increases both intra-chromosomal and inter-chromosomal HR (Spugnesi et al., 2013). This was exploited to study *BRCA2* variant impact showing concordance with human data (Spugnesi et al., 2013). Additionally, yeast-2-hybrid (Y2H) assays are used to study the effect of variants within the N-terminus of BRCA1 and its interaction with BARD1 (Morris et al., 2006). Furthermore Y2H screens have been used to study the effect of variants on interactions between MLH1 and PMS2/EXO1 (Kondo et al., 2003).

Expression of human *MLH1* in yeast results in a dominant-negative effect on yeast MMR activity, allowing the study of multiple variants in human *MLH1* (Shimodaira et al., 1998, Takahashi et al., 2007). Furthermore, introducing disruptive mutations into yeast *mlh1* or *msh2* has a significant impact on the overall mutation rate in this system, producing a mutator

phenotype (Cervelli et al., 2020). This has been used extensively to study the impact of mutations to *mlh1* or *msh2* in yeast at positions corresponding to the human alleles (Ellison et al., 2001, Shcherbakova and Kunkel, 1999, Wanat et al., 2007, Vogelsang et al., 2009, Gammie et al., 2007, Drotschmann et al., 1999) Interestingly, where the corresponding yeast codon cannot be accurately identified due to weak homology between yeast and human, hybrid human-yeast fusion proteins enable the interpretation of variants in these regions (Ellison et al., 2001). Of note, C-terminal fusion proteins (yeast 1-505-human 505-756 and yeast 1-50-human 505-595) cannot complement yeast MMR function, whereas N-terminal fusions can (Ellison et al., 2001). Importantly, this finding suggests that the N-terminal portion or MLH1 is less tolerant to amino acid changes to retain function between species. Furthermore, the effects of *MLH1* mutations have different effects depending on the yeast strain used (Wanat et al., 2007).

Overall, yeast present a cost-effective way of assessing variant impact (Cervelli et al., 2020). Furthermore, deletion of genes, tagging proteins and mutagenesis of your gene of interest are easier and faster than performing the same processes in human cells (Mohammadi et al., 2015). However, findings in yeast do not always comply with findings in the human system, and with one another. Analysis of results from yeast assays which determined an MMR variant as benign, show disagreement for 8/19 variants, compared to 1/18 in mammalian assays (Thompson et al., 2014). Therefore, the use of functional data from yeast assays is not sufficient to provide a definitive classification for any given MMR variant according to the InSiGHT criteria (Thompson et al., 2014)

2.4.3 Survival in response to drugs

Loss of protein function can sensitise or induce cellular resistance to a chemical compound and can therefore be used study the impact of variants on protein function The impact of variants on BRCA1 protein function in response to ionising radiation, PARPi and cisplatin is well

established, and provides an accurate method for determining variant effect (Scully et al., 1999, Bouwman et al., 2013, Chang et al., 2009). The impact of variants in additional HR proteins such as BRCA2, PALB2 and BARD1 on their ability to restore survival in response to PARPi and cisplatin is also an accurate method (Kuznetsov et al., 2008, Boonen et al., 2019, Wiltshire et al., 2020, Becker et al., 2021, Ikegami et al., 2020). In addition to HR proteins, resistance to alkylating agents such as methylnitronitrosoguanidine (MNNG) and antimetabolites such as 6-TG has been used to great effect to study the impact of variants in MMR proteins including MLH1, MSH2, MSH6 and PMS2 (Blasi et al., 2006, Bouvet et al., 2019, Houlleberghs et al., 2017, Houlleberghs et al., 2020, Jia et al., 2021, Rayner et al., 2022, Rath et al., 2022). Overall, analysis of the response of cells containing variants provides an accurate, *in vivo* methodology to assess variant impact (Stein et al., 2019). Indeed, results obtained via assessment of MLH1 variant impact in response to 6-TG showed high concordance with a set of known pathogenic and benign variants from the InSIGHT database (Houlleberghs et al., 2020). Furthermore, as drugs are frequently used in the clinic, it provides an accurate picture for the effect a variant might have on patient outcome.

2.4.4 Mechanism-specific assays

In addition to drug responses, multiple assays have been invented which specifically assay the mechanistic function of the protein. This potentially circumvents off-target effects that drugs may have, providing a direct readout of variant impact. For DDR proteins, assays which examine HR and MMR are well established. The use of plasmid-based MMR assays was discussed in section 2.4.1 and provides an effective *in vitro* method for assessing variant effect. The DR-GFP assay allows *in vivo* assessment of HR activity via the restoration of the GFP open reading frame (Pierce et al., 1999). This has been used successfully to study the impact of variants in BRCA1 (Bouwman et al., 2013, Starita et al., 2018), BRCA2 (Farrugia et al., 2008, Mesman et al., 2019, Richardson et al., 2021), BARD1 (Lee et al., 2015a, Adamovich et

al., 2019) and PALB2 (Boonen et al., 2019, Wiltshire et al., 2020). Additional features of BRCA1 have also been exploited such as the transcriptional-activation activity of its C-terminus (Lee et al., 2010, Phelan et al., 2005, Vallon-Christersson et al., 2001, Ostrow et al., 2004) and its E3 ubiquitin ligase activity (Morris et al., 2006). Whilst offering an accurate means of assessing variant impact, mechanism-specific assays may miss the additional roles of the protein in question, preventing a pan-analysis of the impact of the variant on protein function.

2.4.5 Protein stability and cellular localisation

Variants may disrupt a protein's ability to fold correctly leading to decreased protein stability and aberrant cellular localisation (Sanavia et al., 2020). Therefore, readouts of stability and cellular localisation are often used to determine variant impact (Sanavia et al., 2020). The overall stability and folding of DDR proteins such as MLH1 and BRCA1, has been used as a readout of variant impact in multiple publications (Raevaara et al., 2005, Borràs et al., 2012, Andersen et al., 2012, Lee et al., 2010, Rowling et al., 2010, Williams et al., 2003). Furthermore**,** proficient nuclear localisation of MLH1 has been used as a readout to study the effect of variants in this protein (Raevaara et al., 2005, Borràs et al., 2012, Andersen et al., 2012). Recently, a specific technique termed VAMP-seq, has been developed to allow high-throughput measurement of the impact of variants on the steady-state cellular abundance of a protein (Matreyek et al., 2018). In this method, a library of individual variants is created using genome-editing with each variant fused to a GFP-tag (Matreyek et al., 2018). The stability and thus abundance of each variant-containing protein can then be measured using the GFP signal, via FACS (Matreyek et al., 2018). A single VAMP-seq experiment can generate quantitative data regarding the impact of thousands of individual variants on the stability of a protein (Matreyek et al., 2018)However, it is limited to this readout alone, and lacks mechanistic detail

as to how protein stability is disrupted by the variant in question (Stein et al., 2019, Matreyek et al., 2018).

2.4.6 Protein-protein interactions

Disruption to protein-protein interactions may result in loss of protein stability and/or protein function, making it an important readout for the functional impact of variants (Laddach et al., 2018). Such defects have been assessed as part of analysis of variants in DDR proteins e.g. the Y2H screens performed on MLH1 and PMS2, and BRCA1 and BARD1 as previously discussed (Kondo et al., 2003, Morris et al., 2006). In addition to Y2H approaches, *in vitro* approaches to assess protein-protein interaction have also been used. Variants in the BRCT domains of BRCA1 were examined for their ability to disrupt binding to peptides containing a pSer-X-X-Phe motif, which is a known functional target of the BRCA1 BRCT-domain (Lee et al., 2010). Furthermore**,** GST-pulldown based assays have been used to successfully study the impact of variants within MLH1, assaying for interaction with EXO1 and PMS2 (Guerrette et al., 1999, Andersen et al., 2012, Jäger et al., 2001). Notably, MLH1 variants displaying attenuated binding to PMS2 *in vitro*, showed high levels of MSI when present as mutations in patient-derived cell lines (Andersen et al., 2012), suggesting this can be an accurate predictor of variant effect *in vivo*. The interaction between MLH1 and PMS2 is essential for stability of PMS2 (Trojan et al., 2002). Concomitantly, restoration of PMS2 stability measured via co-transfection of MLH1 and PMS2 into 293T cells and western blotting is an effective readout of MLH1 variant impact (Trojan et al., 2002, Raevaara et al., 2005). Furthermore, transfection of MLH1 variant constructs into MLH1-deficient HCT116 cells also identifies variants incapable of restoring PMS2 stability (Brieger et al., 2002). Co-immunoprecipitation studies also reveal which variants in MLH1 disrupt binding to PMS2 (Raevaara et al., 2005, Kosinski et al., 2010). However, assaying one protein-protein interaction may restrict the analysis to one specific role for the protein, potentially missing other important functions the protein performs.

2.4.7 High-throughput genome editing-based approaches

Classically, variant effects were analysed via transient transfection or lentiviral integration of cDNA constructs encoding the protein of interest followed by functional analysis using one of the assays previously described. These approaches rely on the expression of cDNA from a plasmid or lentivirus which often utilise promoters which are constitutively active. This can result in overexpression of the protein of interest, masking potential variant effects (Ratz et al., 2015, Prelich, 2012, Starita et al., 2017). Furthermore, cDNAs contain no intronic DNA, preventing the assessment which variants may have on pre-mRNA splicing (Lord and Baralle, 2021). Whilst *in silico* and *in vivo* tools to predict splicing exist (Tournier et al., 2008, Scalzitti et al., 2021), assessment of variant impact in a genomic context would provide more accurate information (Starita et al., 2017, Hayward and Ciccia, 2021). The advent of CRISPR-Cas9 technology and its derivatives has enabled the high-throughput assessment of variants in a genomic context (Hayward and Ciccia, 2021). Variants can be introduced into the genome via CRISPR-Cas9 DSB induction and subsequent repair by HR using an ectopically expressed single-stranded oligo DNA nucleotide as a template (Cong et al., 2013, Paquet et al., 2016). This approach has been used to study variants in several proteins including MLH1 and MSH2 (Rath et al., 2022, Rath et al., 2019). This approach enabled the multiplexed assessment of MLH1 variant effect including assays such as immunoblotting for MLH1 and PMS2, PCR-based assessment of splicing, sensitivity to MNNG measured by MTT assay and MSI analysis using PCR amplification of microsatellite regions (Rath et al., 2022). Notably, levels of MLH1 and PMS2 were low for some of the benign control variants, K618A and V716M, suggesting that MLH1 protein levels do not necessarily correspond to MLH1 protein function levels in the cell (Rath et al., 2022). Furthermore, MSI assessment fails to differentiate between benign and pathogenic controls suggesting the MSI markers used may not be a good readout for functionality in some cases (Rath et al., 2022).

Such approaches tend to introduce predefined mutations in study the effects of specific variants of interest Approaches using CRISPR-Cas9 have been expanded to permit the high-throughput screening of variants (Fowler and Fields, 2014). Libraries of donor DNA templates containing individual mutations are provided to introduce every possible base-substitution within your gene of interest, via HR (Findlay et al., 2018a). Individual clones are then grown, often under selection pressure, which couples protein activity to cell growth, stability or ligand binding (Fowler and Fields, 2014, Stein et al., 2019). Subsequently, variants in the library change in frequency depending on how well they survive and/or cope under the selection pressure (Findlay et al., 2018a). The variant frequency is determined using DNA sequencing of the surviving clones, and this provides a readout of how detrimental the variant is for the function of the protein in question (Hayward and Ciccia, 2021). This technique, termed saturation genome-editing, has been used to assay proteins within the DDR, including BRCA1 (Findlay et al., 2018a). This technology is limited by low-efficiency of HDR and/or cell cycle arrest/death due to DSB formation (Hayward and Ciccia, 2021, Haapaniemi et al., 2018, Ihry et al., 2018). Adaptations to CRISPR-Cas9 technology have circumvented these issues to some extent Nucleotide substitutions can be generated in both a DSB and donor-free manner via CRISPR-dependent base-editors and prime-editing (Hayward and Ciccia, 2021). CRISPR-dependent base-editors utilise the nickase Cas9 mutant tethered to an adenine or cytosine deaminase which generate A>G or C>T substitutions respectively, within a 6 base window (Komor et al., 2017, Gaudelli et al., 2017). This was recently utilised to perform a screen of multiple DDR genes (Cuella-Martin et al., 2021, Hanna et al., 2021) and also applied to BRCA1 specifically (Kweon et al., 2020). Prime-editing utilises a prime-editing gRNA (pegRNA) for localisation of the Cas9 nickase mutant (Anzalone et al., 2019). Reverse-transcriptase is then used to insert the sequence of interest from the extended pegRNA (Anzalone et al., 2019). Prime-editing is capable of introducing all 12 possible base

substitutions as well as insertions and deletions (Anzalone et al., 2019). This was recently utilised to study the impact of variants in BRCA2 (Erwood et al., 2022). Base-editor screens are limited by the activity and windows of the base-editors utilised, as well as the location of PAM sites required for Cas9 nuclease activity (Hayward and Ciccia, 2021). Whilst prime-editing may circumvent this, the efficiency of this technique differs widely between cell lines (Anzalone et al., 2019).

2.4.8 Computational approaches

In addition to the functional assays outlined above, a large number of bioinformatic algorithms exist to estimate the functional impact of individual variants (Peterson et al., 2013, Ghosh et al., 2017, Livesey and Marsh, 2022, Wang et al., 2022a). These are typically publicly available, and integrate several different parameters to determine the consequences of variants on protein function (Peterson et al., 2013, Niroula and Vihinen, 2016). Since the creation of the Sorting Intolerant From Tolerant (SIFT) algorithm (Ng and Henikoff, 2003), the number of such predictors has continuously increased (Livesey and Marsh, 2022). An analysis of the frequency of use of *in silico* evidence for pathogenic and benign variant assertion in online databases revealed frequencies of 39% and 18% respectively (Amendola et al., 2016). Metrics employed by such algorithms for predicting variant impact include evolutionary sequence conservation, amino acid physiochemical properties, secondary structure, known residue function, splice-site location and comparisons to known functional and population data (Wang et al., 2022a). Notably, algorithms have been extensively to predict the impact of variants in DDR proteins including BRCA1 (Abkevich et al., 2012, Karchin et al., 2007, Nielsen et al., 2017), MSH2 (Nielsen et al., 2017, Frazer et al., 2021, Thompson et al., 2013) and MLH1 (Arnold et al., 2009, Chao et al., 2008, Abildgaard et al., 2019, Chan et al., 2007, Thompson et al., 2013). Indeed, a combination of evolutionary sequence conservation of MLH1 orthologues combined with existing functional and clinical data can accurately determine MLH1 variant impact

(Arnold et al., 2009, Chao et al., 2008, Chan et al., 2007). Combining predictions from both MAPP (Stone and Sidow, 2005) and PolyPhen2 (Adzhubei et al., 2010) algorithms is the most accurate *in silico* determinant of MMR variant impact (Chao et al., 2008). Scores for individual variants derived from a combination of these algorithms for MLH1 can be found on (http://priors.hci.utah.edu/PRIORS/BRCA/indexMLH1.php). This is recommended for use as the computational line of evidence in variant interpretation according to the INSiGHT consortium https://www.insight-group.org/content/uploads/2021/11/DRAFT_Nov_2021_TEMPLATE_SVI.ACMG_Specifications_InSiGHT_MMR_V1.pdf). Protein stability can also be predicted computationally and use of the FoldX energy function (Schymkowitz et al., 2005) can accurately predict variants which destabilise MLH1 *in vivo* (Abildgaard et al., 2019).

The obvious limitation of using algorithms to predict variant impact is the lack of functional readout**.** Furthermore such a large number of *in silico* algorithms based on a variety of different inputs and metrics makes it challenging to know which algorithm is best to use (Peterson et al., 2013, Wang et al., 2022a, Livesey and Marsh, 2022). To further complicate matters, different bioinformatic prediction algorithms can differ significantly in their conclusions (Ghosh et al., 2017). Indeed, a recent meta-analysis of over 25 algorithms, including SIFT and PolyPhen, using 14,819 pathogenic or benign missense variants from ClinVar (which had obtained at least one star status), revealed high levels of disagreement between algorithms (Ghosh et al., 2017). The authors observed only 5.2% of the benign and 39.2% of pathogenic variants had concordant assertions across all algorithms (Ghosh et al., 2017). To compound this issue, many researchers have found when performing functional testing that their data does not match predictions by algorithms. Indeed, when comparing classifications for PALB2 to the predictions of the PolyPhen and SIFT algorithms, little to no correlation was observed (Boonen et al., 2019). Notably, a gross over-representation of deleterious variants was observed when

using these algorithms compared to functional observations using HR and PARPi assays (Boonen et al., 2019). However, computational-studies have recently been shown to match or even out-perform results from functional assays, when interpreting variants in DDR proteins. Approaches modelling the distribution of sequence variation across organisms were able to match or even out-perform findings from saturation-genome editing approaches for determining the impact of variants in BRCA1 and MSH2 (Frazer et al., 2021, Findlay et al., 2018a, Jia et al., 2021).

Overall, significant efforts have been made, and continue to be made, by researchers to understand the functional impact of variants in DDR proteins. Despite this, the vast majority of variants in DDR genes remain classified as variants of uncertain significance (VUS) on ClinVar **Table 2-5** (Landrum et al., 2018). This presents a pressing need for additional functional information required to classify them (Richards et al., 2015). In this thesis I describe a novel methodology for assessing the functional impact of variants and apply it to proteins of multiple DDR pathways. This method is based on the ability of orthologues from other species being able to perform the same function when expressed in a cell type from another species/different organism, termed orthologue complementation. In the next section, I provide a discussion of this concept and how I plan to apply it to understanding the functional impact of amino acid variants.

Gene	Associated disease(s)	Total missense variants	VUS (%)
ATM	Ataxia-telangiectasia, Familial breast cancer, Hereditary predisposing cancer-syndrome.	5,673	5303 (93%)
ATR	Seckel Syndrome, Hereditary predisposing cancer syndrome.	451	421 (93%)
BRCA1	Familial breast and ovarian cancer.	5021	2886 (57%)
BRCA2	Familial breast and ovarian cancer.	6649	5628 (85%)
CHK2	Familial breast cancer.	1353	1284 (95%)
MLH1	LS, Hereditary nonpolyposis colorectal neoplasms.	1486	1216 (82%)
MSH2	LS, Hereditary nonpolyposis colorectal neoplasms.	2182	1954 (90%)
MSH6	LS, Hereditary nonpolyposis colorectal neoplasms.	3286	3057 (93%)
TP53	Li-Fraumeni Syndrome, Hereditary cancer predisposing syndrome.	1105	688 (62%)
PALB2	Familial breast and ovarian cancer.	2170	2016 (93%)

Table 2-5: % of variants in DDR genes which are VUS on the ClinVar database as of Jan 2023

2.4.9 Orthologue complementation

In the 1970's, it was determined that cytochrome C from different mammalian species is fully interchangeable *in vitro* when tested with mitochondrial cytochrome oxidase (Dickerson, 1971). Some years later, orthologue complementation was central to the Nobel Prize winning work demonstrating that human CDK1 could complement yeast defective for *Cdc2* (Lee and Nurse, 1987). More recent work replaced 414 yeast genes with their human orthologues, finding that approximately half could successfully restore the function of the missing yeast gene (Kachroo et al., 2015). Additionally, several DDR genes including *XPA* and *XRCC4* were first cloned via complementation of cell lines from one species with orthologous cDNA of another (Tanaka et al., 1989, Li et al., 1995b). This highlights that whilst species present clear phenotypic differences, and in some cases protein function has diverged between them (e.g. crocodile haemoglobin (Perutz et al., 1981)), orthologues show remarkable complementation ability between species. This is in line with the neutral theory (Kern and Hahn, 2018), suggesting that the vast majority of sequence differences between orthologues has no effect on protein function. To date, *in silico* predictions frequently utilise sequence alignments of protein orthologues to assess conservation of any given amino acid (Peterson et al., 2013). Conservation of any given residue has been shown to be a potent predictor of the functional importance of the residue in question (Peterson et al., 2013, Frazer et al., 2021). Furthermore, the fundamental conservation of protein function between organisms has allowed the frequent use of model organisms to study protein function (Hieter and Boycott, 2014). Recently, the Congruent clinical Variation Visualization Tool (ConVarT) was curated to allow users to search for variants present or generated in multiple species with phenotypic annotations (Pir et al., 2021). Importantly, variants have been introduced at equivalent positions in model organisms to study their impact on protein function (Hieter and Boycott, 2014, Houlleberghs et al., 2020, Houlleberghs et al., 2017). This was performed recently for the MLH1 and MSH6

proteins, whereby the authors introduced human VUS, pathogenic and benign controls into the corresponding position in the mouse protein. Importantly, known pathogenic variants in the human proteins also disrupted mouse MLH1 protein function when introduced at the corresponding position, whereas benign variants were tolerated (Houlleberghs et al., 2020, Houlleberghs et al., 2017). Overall, this highlights the potential that orthologue complementation has for determining the effects of variants in DDR proteins.

2.5 Aims and Objectives

Identifying whether amino acid variants in DDR proteins affect protein function is of paramount importance for the stratification of patients in the clinic and the application of appropriate therapies (Groelly et al., 2022, Stein et al., 2019, Richards et al., 2015). To date, a multitude of approaches have been developed which have sought to assess the impact of such variants, but the vast majority still remain classified as VUS on databases such as ClinVar. This highlights a significant problem, in that we still do not understand the impact the vast majority of currently identified variants have on DDR protein function. Furthering our understanding of such changes would facilitate improved prognosis in the clinic via improved diagnosis and treatment. Furthermore, it may further our understanding of how DDR proteins are regulated via assessment of whether variants at sites of reported PTMs impact protein function. The aim of this DPhil has been to develop a novel methodology which allows for the high-throughput functional assessment of amino acid variants. The findings discussed thus far suggest that orthologues from other species could be used to complement human cells. Furthermore, introduction of variants found in human at the corresponding positions in orthologues have the same effect as if introduced into the human protein (Houlleberghs et al., 2017, Houlleberghs et al., 2020). Due to extensive sequence divergence between species, complementation of human cells with any given orthologue may reveal amino acids capable of undergoing variation with no consequence to protein function. Variants present at the corresponding position in a

complementing orthologue would likely be tolerated by the human protein. To my knowledge, widespread assessment of the ability of orthologues to complement human cells has not been performed. Here, I aim to carry out the systematic complementation of human cells with orthologues of DNA repair proteins from a diverse array of species. Specifically, the methodology will be applied to XPA of the NER pathway, XRCC4 of the NHEJ pathway, MLH1 of the MMR pathway and FANCL of the FA/ICL repair pathway. All proteins chosen are fundamental to their respective pathways, all are mutated in disease and all exist as orthologues in a wide array of species. Complementation ability is measured using functional assays appropriate to the protein of interest. Fully complementing orthologues will have their sequences aligned to reveal amino acid variants present at the corresponding positions in these species. Subsequently, variant amino acids present in complementing orthologues will be introduced into the human protein to demonstrate their functionality. Lastly, my findings will be compared to computational predictions, public databases of variant effects and previous functional data to assess overall concordance. A schematic of the methodology is outlined in **Figure 2-8.** Due to its functional nature, data produced could be classified as strong by the ACMG/AMP framework, offering an advantage over predictions produced by *in silico* algorithms exploiting evolutionary sequence variation (Richards et al., 2015, Peterson et al., 2013, Frazer et al., 2021, Horak et al., 2022). Furthermore, the methodology is not restricted to PAM site location in contrast to CRISPR-based approaches (Cuella-Martin et al., 2021, Hayward and Ciccia, 2021). Lastly, due to the extensive sequence variation between species, it allows for an efficient, high-throughput way of testing the functional impact of multiple variants at a time with several different functional readouts.

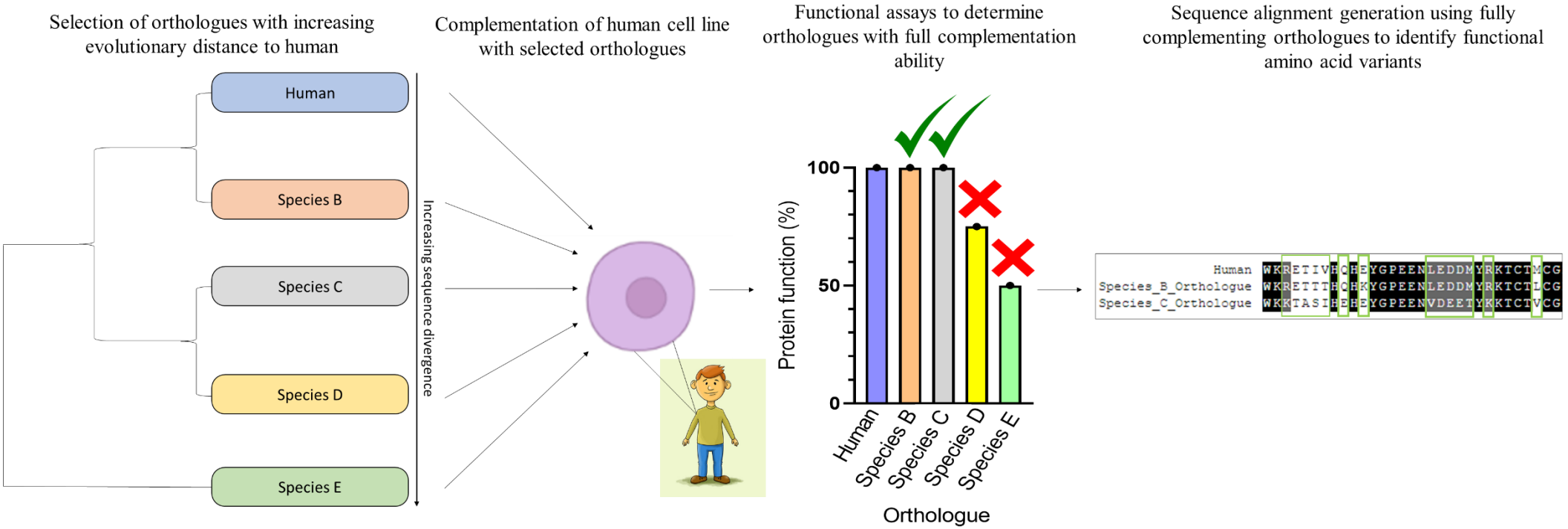


Figure 2-8: Orthologue complementation of human cells as a method to identify functional amino acid variants in the human protein

Orthologues are selected from a wide array of diverged species with increasing evolutionary distance to human. Subsequently, the orthologues are expressed in human cells and functional assays are performed to determine whether or not they function to the same extent as the human orthologue. Orthologues which exhibit the same complementation ability as the human orthologue then have their sequences aligned to identify areas of amino acid variation between species. The amino acids present in fully complementing orthologues should be tolerated by the human protein at the corresponding position. In this way, the methodology identifies functional amino acid variants in the human protein, providing functional information for the re-classification of VUS.

3 Methods

3.1 Generation of sequence alignments

Sequences from each protein of interest were acquired from NCBI and aligned using the T-Coffee algorithm (https://www.ebi.ac.uk/Tools/msa/tcoffee) (Notredame et al., 2000). Accession numbers for each orthologue can be seen in **Table 3-1**. Conservation of identical or similar residues was indicated using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html).

Gene	Orthologue	Common name	Accession Number
XPA	*Homo sapiens*	Human	NP_000371.1
XPA	*Mus musculus*	Mouse	NP_035858.2
XPA	*Gallus gallus*	Chicken	NP_990184.2 (no longer available, now NP_990184.3)
XPA	*Xenopus laevis*	Frog	NP_001081354.1
XPA	*Danio rerio*	Zebrafish	XP_005159871.1
XPA	*Drosophila melanogaster*	Fruit Fly	NP_476866.1
XPA	*Schizosaccharomyces pombe*	Fission Yeast	NP_595222.1
XRCC4	*Homo sapiens*	Human	NP_003392.1
XRCC4	*Mus musculus*	Mouse	Q924T3.1
XRCC4	*Gallus gallus*	Chicken	XP_015136170.1
XRCC4	*Xenopus tropicalis*	Frog	XP_012826245.1
XRCC4	*Danio rerio*	Zebrafish	NP_957080.1
MLH1	*Homo sapiens*	Human	NP_000240.1
MLH1	*Mus musculus*	Mouse	NP_081086.2
MLH1	*Gallus gallus*	Chicken	XP_418828.1
MLH1	*Xenopus laevis*	Frog	NP_001090545.1
FANCL	*Homo sapiens*	Human	NP_001108108.1
FANCL	*Mus musculus*	Mouse	NP_080199.1
FANCL	*Gallus gallus*	Chicken	NP_001029113.1
FANCL	*Xenopus tropicalis*	Frog	NP_001007482.1
FANCL	*Danio rerio*	Zebrafish	NP_998147.1

Table 3-1: Accession numbers of orthologues used in this study

Table displaying the gene, orthologue, common name and NCBI accession number with associated information of all the orthologues used in this study.

3.2 SDS-PAGE and western blotting

SDS-PAGE and western blotting were performed using Tris-Bicine gels with the SE400 and TE42 systems (Hoefer), or Bolt 4-12% Bis-Tris Plus gels (Thermo Fisher Scientific) using nitrocellulose membrane. Nitrocellulose membranes were first blocked in 5% milk in Tris-buffered Saline, 0.1% Tween (TBST) for 1 hour then incubated in primary antibody in 5% milk in TBST for 2 hours. The following primary antibodies were used at the indicated dilutions: CHK1 (sc-8408, Santa Cruz Biotechnology, 1:500), EXO1 (ab95068, Abcam, 1:1000), FANCA (ab201457, Abcam, 1:2000), FANCD2 (sc-20022, Santa Cruz Biotechnology, 1:500), GFP (11814460001, Roche, 1:1000, 1:5000), LIG4 (ab193353, Abcam, 1:1000), MLH1 (ab92312, Abcam, 1:2000), MSH6 (ab92471, Abcam, 1:3000), NBS1 (ab32074, Abcam, 1:2000), PMS2 (ab110638, Abcam, 1:1000), XPA (ab65963, Abcam, 1:2000), XPC (A301-122A, Bethyl Laboratories, 1:2000) and XRCC4 (611506, BD Transduction Labs, 1:2000). Membranes were washed five times in TBST before incubation in secondary antibody conjugated to HRP in 5% milk in TBST for 1 hour. The following secondary antibodies were used: goat anti-rabbit IgG HRP (P044801-2, Agilent Technologies) and goat anti-mouse IgG HRP (P044701-2, Agilent Technologies). Membranes were then treated with ECL (Amersham or Bio-Rad) as the chemiluminescent substrate for HRP. The signal was then detected using X-ray film (Scientific Laboratory Supplies).

3.3 Plasmids and 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). *Homo sapiens*, *Mus musculus*, *Gallus gallus*, *Xenopus laevis* and *Danio rerio XPA* cDNAs were synthesised by GeneArt (Thermo Fisher Scientific) and cloned into pEGFP-C1 via *Eco*RI and *Sal*I restriction digest and ligation by Dr. Samuel E. Jones of the Weatherall Institute of Molecular Medicine. *Drosophila melanogaster*,

Amphimedon queenslandica, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae XPA* cDNAs were synthesised and cloned into pEGFP-C1 by GenScript. *Homo sapiens XPA* fragments containing S173A, S196A and T233A were synthesised by GenScript, amplified by PCR and then cloned into pEGFP-C1 using InFusion Technology (Takara Bioscience). L191V and V234L were introduced into pEGFP-C1 *Homo sapiens XPA* via site-directed mutagenesis using the QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies) and primers designed with the QuikChange Primer Design Tool (Agilent) (https://www.agilent.com/store/primerDesignProgram.jsp). I68T and K141R were introduced into pEGFP-C1 *XPA* via site-directed mutagenesis using KLD enzyme mix (New England Biolabs) and primers designed with the NEB base changer (https://nebasechanger.neb.com/).

GFP-2X-FLAG-*Homo sapiens* XRCC4 in pEGFP-C1 FLAG was a gift from Sebastien Britton. *Mus musculus* XRCC4 in pBluescript was provided by Dr. Yijie Gao of Harvard University Medical School and then cloned into pEGFP-C1 FLAG via *Eco*RI and *Sal*I restriction digest and ligation by Dr. Samuel E. Jones. *Gallus gallus*, *Xenopus tropicalis* and *Danio rerio* XRCC4 cDNAs were synthesised by GeneArt (Thermo Fisher Scientific) and cloned into pEGFP-C1 FLAG via *Eco*RI and *Sal*I restriction digest and ligation by Dr. Samuel E. Jones. A *Homo sapiens* XRCC4 cDNA containing either S6A, T17A, S219A, T222A, S237A, T241A and T264A were synthesised and cloned into pEGFP-C1 by GenScript.

mClover3-*Homo sapiens, Mus musculus, Gallus gallus, Xenopus tropicalis and Danio rerio* FANCL cDNAs were synthesised and cloned into pcDNA5/FRT/TO by GeneArt. mClover3-FANCL cDNAs were amplified by PCR, subcloned into pDONR221 and cloned into pLenti-PGK (Addgene #19067, a gift from Prof Ross Chapman of The Weatherall Institute of Molecular Medicine) via Gateway cloning (Hartley et al., 2000) by Miss Roos Vincken, an Erasmus Masters student in the Blackford lab.

Homo sapiens MLH1 in pEGFP-C1 was a gift from Dr. Josep Forment of AstraZeneca. *Mus musculus* MLH1 was cloned from a pCI-FLAG vector (gifted by Dr. Marilia Cascalho of The University of Michigan) into pEGFP-C1 using *Sac*I and *Kpn*I restriction digestion and ligation by Dr. Samuel E. Jones. *Gallus gallus* MLH1 cDNA was synthesised by GeneArt and then cloned into pEGFP-C1 via *Eco*RI and *Sal*I restriction digestion and ligation by Dr. Samuel E. Jones. *Xenopus laevis* MLH1 was cloned from a pCMV-Sport6 vector purchased from Source BioScience into pEGFP-C1 via *Eco*RI and *Sal*I restriction digestion and ligation by Dr. Samuel E. Jones. All GFP-MLH1 orthologues were amplified by PCR, subcloned into pDONR221 and delivered into pLenti-PGK via gateway cloning. S46N, T82I, V185I, V185G, P208S, A619T, A681S and A681T were introduced into pDONR221 GFP-*Homo sapiens* MLH1 via site-directed mutagenesis using KLD enzyme mix (New England Biolabs) using primers designed with the NEB base changer. Variant GFP-MLH1 cDNAs were amplified by PCR, subcloned into pDONR221 and then delivered into pLenti-PGK via Gateway cloning. All primers used are listed in **Table 3-2**.

Primer name	Primer sequence
Homo sapiens GFP-XPA L191V_Fwd	CCCCACACTTCCACGCTCCGCTTCACG
Homo sapiens GFP-XPA L191V_Rev	CGTGAAGCGGAGCGTGGAAGTGTGGGG
Homo sapiens GFP-XPA V234L_Fwd	CTCTTCCACAGACTGCTCCGCACGGCT
Homo sapiens GFP-XPA V234L_Rev	AGCCGTGCGGAGCAGTCTGTGGAAGAG
Homo sapiens GFP-XPA I68T_Fwd	GCTCCTAAGACCATCGACACC
Homo sapiens GFP-XPA I68T_Rev	GGCTTTCACATTAGCCATG
Homo sapiens GFP-XPA K141R_Fwd	CTGATCACCAGGACCGAGGCC
Homo sapiens GFP-XPA K141R_Rev	CTTGTGCTTGTCGTCGGC
Homo sapiens GFP-XPA pEGFP-C1 backbone Fwd	GGTACCGCGGGCCCGGGAT
Homo sapiens GFP-XPA pEGFP-C1 backbone Rev	TTGTCGCAGGTAGGCAGGT
Homo sapiens GFP-XPA 3A insert Fwd	GCCTACCTGCGACAACTGCAGAGATGCTGAT GATAAACAC
Homo sapiens GFP-XPA 3A insert Rev	CGGGCCCGCGGTACCGTCGACTCACATTTTT TCATATGTC
Homo sapiens GFP-MLH1 S46N_Fwd	AAATCCACAAATATTCAAGTG
Homo sapiens GFP-MLH1 S46N_Rev	TGCATCTAAACAGTTCTC
Homo sapiens GFP-MLH1 T82I_Fwd	AGGTTCACTATTAGTAAACTGC
Homo sapiens GFP-MLH1 T82I_Rev	TTCACATACAATATCCAGATC

Homo sapiens GFP-MLH1 V185I_Fwd	CAGGTATTCAATACACAATGC
Homo sapiens GFP-MLH1 V185I_Rev	CCAACAACTTCCAAAATTTTC
Homo sapiens GFP-MLH1 V185G_Fwd	AGGTATTCAGGACACAATGCAG
Homo sapiens GFP-MLH1 V185G_Rev	GCCAACAACTTCCAAAATTTTC
Homo sapiens GFP-MLH1 P208S_Fwd	TAGGACACTATCCAATGCCTC
Homo sapiens GFP-MLH1 P208S_Rev	ACATCAGCTACTGTCTCTC
Homo sapiens GFP-MLH1 A619T_Fwd	GAAGAAGAAGACTGAGATGCTTG
Homo sapiens GFP-MLH1 A619T_Rev	AGAAACTCAACAATGTATTCAG
Homo sapiens GFP-MLH1 A681S_Fwd	TAAAGAATGCTCTATGTTCTATTCC
Homo sapiens GFP-MLH1 A681S_Rev	CTGAGGCTTTCAAAACATTC
Homo sapiens GFP-MLH1 A681T_Fwd	TAAAGAATGCACTATGTTCTATTCC
Homo sapiens GFP-MLH1 A681T_Rev	CTGAGGCTTTCAAAACATTC
pEGFP-C1 Gateway_Fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTC CCACCATGGTGAGCAAGGGCGAGG
Homo sapiens GFP-MLH1 Gateway_Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCT AACACCTCTCAAAGACTTT
Mus musculus GFP-MLH1 Gateway_Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTT AACACCGCTCAAAGACTTT
Gallus gallus GFP-MLH1 Gateway_Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTC AGCATCTCTCAAAAACTTT
Xenopus laevis GFP-MLH1 Gateway_Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCT AGCACCTTTCAAACACTTT

Table 3-2: Names and sequences of primers used in this study

3.4 Cell lines and 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). HAP1 and HAP1 Δ*FANCL* cells were purchased from Horizon Discovery. HAP1 Δ*XRCC4*, HAP1 Δ*XPA* and HAP1 Δ*MLH1* cells were a gift from Prof. Steve Jackson of The University of Cambridge. HAP1 cells were cultured in Iscove's Modified Dulbecco's Medium (IMDM, Sigma-Aldrich, I3390) supplemented with 10% FCS, 2 mM L-glutamine and 100 units/ml penicillin/100 µg/ml streptomycin. After either transfection or transduction with G418 resistant expression plasmids or lentivirus, 500 µg/ml G418 was added to media to maintain selection. HCT116 cells were a gift from Prof. Ross Chapman and cultured in DMEM supplemented with 10% FCS, 2 mM L-glutamine and 100 units/ml penicillin/100 µg/ml streptomycin. After transduction with lentivirus, 500 µg/ml Geneticin (G418; Thermo Fisher Scientific) was added to media to

maintain selection. 293FT and 293T were purchased from Thermo Fisher Scientific and cultured in DMEM supplemented with 10% FCS, 2 mM L-glutamine, 1% MEM non-essential amino acids (Thermo Fisher Scientific) and 500 μg/ml G418.

3.5 Immunoprecipitations

Plasmids were transfected into 293FT cells using Lipofectamine 2000 (Thermo Fisher Scientific) according to the 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 20 μl of GFP-Trap magnetic agarose beads (ChromoTek) for 2 hours 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.

3.6 Lentiviral assembly and stable cell line generation

To assemble lentiviruses, 293FT cells were transfected with pLenti PGK Neo DEST encoding GFP, GFP-*Homo sapiens* MLH1, GFP-*Mus musculus* MLH1, GFP-*Gallus gallus* MLH1, GFP-*Xenopus laevis* MLH1, mClover3-*Homo sapiens* FANCL, mClover3-*Mus musculus* FANCL, mClover3-*Gallus gallus* FANCL, mClover3-*Xenopus laevis* FANCL and mClover3-*Danio rerio* FANCL with pHDM-tat1b, pHDM-G, pRC/CMV-rev1b, and pHDM-Hgpm2 lentiviral assembly plasmids (all gifts from Prof. Ross Chapman) using Lipofectamine 2000 according to the manufacturer's instructions. Viral supernatants were harvested at 48 hours, pooled, passed through a 0.45-μm filter, and stored at −80 °C. For transduction**,** cells were cultured in media containing lentiviral supernatant for 48 hours. Transduced cells were cultured in

selection for 5 days then selected via Fluorescence Activated Cell sorting (FACS) to obtain a stable pool.

3.7 Colony survival assays

pEGFP-C1 constructs encoding GFP or orthologues of GFP-XPA were transfected into HAP1 WT and HAP1 Δ*XPA* cells using Lipofectamine 3000 and P3000 (Thermo Fisher Scientific) according to the manufacturer's instructions. Positive cells were then selected via FACS. 24 hours after seeding, media was removed, cells were irradiated using the UV Stratalinker 2400 (Stratagene) and media was then added back to cells. Cells were then left for 7 days to allow colonies to grow. Colonies were washed in PBS, stained in 0.1% Coomassie Brilliant Blue R 250, 7% acetic acid, 50% methanol at room temperature for 30 minutes, and washed in water before counting.

pEGFP-C1 constructs encoding GFP or orthologues of GFP-XRCC4 were transfected into HAP1 WT and HAP1 Δ*XRCC4* cells using Lipofectamine 3000 and P3000 (Thermo Fisher Scientific) according to the manufacturer's instructions. Positive cells were then selected via FACS. 24 hours after seeding, cells were irradiated with the indicated doses of IR. Cells were then left for 7 days to allow colonies to grow. Colonies were washed in PBS, stained in 0.1% Coomassie Brilliant Blue R 250, 7% acetic acid, 50% methanol at room temperature for 30 minutes, and washed in water before counting.

HAP1 WT, Δ*MLH1* or HCT116 cells stably expressing GFP or GFP-MLH1 orthologues were sorted via FACS and plated. 24 hours after seeding, cells were treated with the indicated doses of 6-TG in media. 24 hours later, media was removed and replaced with drug free media. Cells were then left for 7 days to allow colonies to grow. Colonies were washed in PBS, stained in 0.1% Coomassie Brilliant Blue R 250, 7% acetic acid, 50% methanol at room temperature for 30 minutes, and washed in water before counting.

HAP1 Δ*FANCL* cells stably expressing GFP or mClover3-FANCL orthologues were sorted via FACS and plated. 24 hours after seeding, cells were treated with the indicated doses of cisplatin in media. 4 hours later, media was removed and replaced with drug free media. Cells were then left for 7 days to allow colonies to grow. Colonies were washed in PBS, stained in 0.1% Coomassie Brilliant Blue R 250, 7% acetic acid, 50% methanol at room temperature for 30 minutes, and washed in water before counting.

3.8 Variant and PTM database searching

Data was downloaded from the publicly available databases: ClinVar (Landrum et al., 2014), Phosphositeplus (Hornbeck et al., 2015) and gnomAD (Karczewski et al., 2020). Computational analysis searching the sequence alignments of complementing orthologues for variants or residues found in these databases was performed in collaboration with Mr. Malitha Ratnaweera of The Weatherall Institute of Molecular Medicine.

4 Orthologue Complementation of XPA

4.1 Introduction

XPA was first cloned via selection of UV-resistant complementation group A XP human cells after transfection with mouse DNA fragments (Tanaka et al., 1989). A year later, human *XPA* was cloned, identifying a zinc finger-containing polypeptide 273 amino acid residues in length that abolishes the UV-sensitivity of XP-A patient cell lines (Tanaka et al., 1990). Loss of *XPA* function is viable and *XPA*-null mice display no developmental or pathological abnormalities (Nakane et al., 1995). Despite this, loss of *XPA* results in NER-deficiency and extreme sensitivity to UV radiation and UVB-induced skin carcinogenesis (Nakane et al., 1995, Tanaka et al., 1989, Tanaka et al., 1990). Functionally, XPA acts as a scaffold during both pathways of NER, interacting with RPA (Matsuda et al., 1995, He et al., 1995), ERCC1 (Li et al., 1994), TFIIH (Park et al., 1995, Nocentini et al., 1997), XPC-RAD23B (Bunick et al., 2006), DDB2 (Wakasugi et al., 2001, Wakasugi et al., 2009) and PCNA (Gilljam et al., 2012). XPA also binds to DNA in cooperation with RPA (Patrick and Turchi, 2002; Saijo, Takedachi and Tanaka, 2011), with the consensus now being that XPA acts in concert with RPA to organise the NER machinery into a pre-incision complex (Sugitani et al., 2016, Pulzová et al., 2020). This ensures positioning of the NER nucleases in a manner that results in accurate excision of the lesion (Schärer, 2013). Overall, XPA is fundamental to NER and loss of XPA protein function provides a clear assayable phenotype in extreme sensitivity to UV radiation.

Structurally, human XPA is a 31 kDa 273 amino acid residue polypeptide (Tanaka et al., 1990), that typically exists as a homodimer (Yang et al., 2002b). It is made up of three separate domains: a central globular domain and disordered N and C-termini (Pulzová et al., 2020). The central globular domain (AA 98-219) contains a C4 type zinc-finger motif comprising residues C105, C108, C126 and C129 (Kuraoka et al., 1996, Buchko et al., 1999, Ikegami et al., 1998, Morita et al., 1996). This domain, plus a C-terminal extension (AA 98-239), binds to Y shaped

ssDNA-double stranded (ds) DNA junctions (Sugitani et al., 2016). Overall, the DNA-binding domain of XPA is assigned to amino acid residues 98-239 (Sugitani et al., 2016). The interaction between XPA and RPA is mediated by the N-terminus of XPA, and the RPA32C subunit protein recruitment domain, as well as residues in the DNA-binding domain of XPA and RPA70 (Saijo et al., 1996, Li et al., 1995a, Mer et al., 2000, Topolska-Woś et al., 2020, Kim et al., 2022). Disruption of these interactions results in abrogated NER activity (Saijo et al., 2011, Topolska-Woś et al., 2020, Kim et al., 2022), highlighting its importance in NER. XPA interacts with XPC via amino acid residues 154-334 of XPC (Bunick et al., 2006). To date, no structure of the XPA-XPC complex has been obtained, so further details of this interaction are lacking. XPA binds to DDB2 via residues 185-225 (Wakasugi et al., 2009) and TFIIH via residues 226-273 (Park et al., 1995, Nocentini et al., 1997, Ziani et al., 2014). Lastly, XPA interacts with ERCC1 via amino acid residues 59-114 of XPA and amino acid residues 91-118 of ERCC1 (Li et al., 1994, Li et al., 1995a). Structural data of this complex is now available, identifying three highly conserved glycines (G72-74) important for ERCC1 binding (Tsodikov et al., 2007). Disruption to this binding results in abrogated NER activity (Li et al., 1995a, Saijo et al., 2004). Overall, multiple domains of XPA are crucial for organising multiple components of the NER pathway, highlighting its fundamental role in this pathway (Sugitani et al., 2016).

XPA undergoes several PTMs which can positively or negatively regulate NER function (Park and Kang, 2016). XPA is phosphorylated on S196 and S173 by ATR in a UV dose-dependent manner, with mutation of S196 to alanine resulting in decreased cellular survival in response to UV (Wu et al., 2006b). Phosphorylation of XPA on S196 is also important for XPA stability in response to UV, by preventing HERC2-mediated ubiquitination and subsequent degradation (Lee et al., 2014b). Furthermore, phosphorylation of S196 facilitates retention of XPA on chromatin, and its interaction with the RPA70 subunit (Lee et al., 2014b). Dephosphorylation

of this residue is performed by the WIP1 phosphatase (Nguyen et al., 2010), and WIP1 KO mice show faster repair of CPDs than WT mice, resulting in lower levels of UV-induced apoptosis (Nguyen et al., 2010). XPA is acetylated on K63 and K67 (Fan and Luo, 2010). Acetylation of these residues disrupts the interaction between XPA and the RPA32 subunit (Fan and Luo, 2010). *XPA*-deficient cells complemented with XPA K63Q K67Q - mimicking hyperacetylated XPA - display significantly higher UV-sensitivity (Fan and Luo, 2010). The deacetylation of XPA is performed by SIRT1, and downregulation of SIRT1 results in increased UV-sensitivity (Fan and Luo, 2010). Concomitantly, increased levels of SIRT1 result in increased levels of chromatin-bound XPA, facilitating the NER of UV and cisplatin-induced DNA lesions (Choi et al., 2015). As a result, cancers that display elevated levels of SIRT1, present resistance to platinum-based therapies, with SIRT1 inhibition overcoming this treatment barrier (Pulzová et al., 2020). XPA is also PARylated in response to UV radiation, facilitating its recruitment to DNA damage sites (King et al., 2012, Fischer et al., 2014). PARylation of XPA also promotes the interaction of XPA with PARP1, facilitating proficient NER (Fischer et al., 2014, King et al., 2012). In addition to the modifications discussed above, many more PTMs of XPA have been discovered via high-throughput approaches such as mass-spectrometry. This includes phosphorylation, ubiquitination, acetylation and SUMOylation of XPA (Akimov et al., 2018, Hendriks et al., 2014, Zhou et al., 2013). Despite their identification, knowledge of their functional role is absent due to a lack of direct functional assessment. Overall, PTM of XPA clearly plays a role in regulating its function. Identifying whether PTM sites with no known function play a role in regulating XPA function may offer further insight into the role of this protein in the NER pathway.

Mutations in *XPA* give rise to XP (DiGiovanna and Kraemer, 2012). Mutations in *XPA* are found throughout the gene, with the resulting phenotypes ranging from mild to severe, depending on the location and nature of the mutation (Sugitani et al., 2016). Mutations that

produce little to no functional XPA, give rise to the most severe XP phenotypes (Sugitani et al., 2016). Examples include severe truncations or disruptive mutations to the zinc-finger domain (Sugitani et al., 2016, States et al., 1998). In contrast, deletion of exon 1, coding for the ERCC1 and RPA32C-binding regions, only results in modest NER disruption (Miyamoto et al., 1992). Furthermore, deletion of exon 6, coding for the TFIIH-binding region, also results in modest NER disruption (Miyamoto et al., 1992). To this end, two patients with C-terminal truncation mutations of *XPA* demonstrate mild XP symptoms (Takahashi et al., 2010). Many XP patients demonstrate a severely increased risk of skin cancer (DiGiovanna and Kraemer, 2012). Despite this, clear evidence that somatic mutations in *XPA* drive carcinogenesis is lacking (Sugitani et al., 2016, Pulzová et al., 2020). Pan-analysis of 121 cancer genomic studies identifies 56 somatic mutations in *XPA* (Sugitani et al., 2016). Notably, any given *XPA* mutation appears in no more than three patient samples (Sugitani et al., 2016), suggesting that carcinogenesis is typically not driven by somatic mutations in *XPA*. Despite this, some single nucleotide polymorphisms (SNPs) in *XPA* are linked to cancer risk, and/or treatment outcomes and chemotherapy resistance (Pulzová et al., 2020). Most SNPs in *XPA* with links to carcinogenesis are found within the 5' UTR and intronic regions of *XPA* (Pulzová et al., 2020). Despite this, recent studies link the R228G variant to having a protective effect in lung cancer patients heterozygous for the c.23A>G mutation in the 5' UTR of *XPA* (Butkiewicz et al., 2000, Mellon et al., 2002). Indeed, R228G enhances survival of cells treated with benzo(a)pyrene diol epoxide (BPDE), a DNA-intercalating agent (Porter et al., 2005, Mellon et al., 2002). In addition to this, the variant V234L in XPA increases survival in response to BDPE (Mellon et al., 2002, Porter et al., 2005). However, to my knowledge, no association between this variant and cancer outcome has been established. Overall, this highlights that whilst rare, SNPs in *XPA* coding regions may have an impact on carcinogenesis and responses to chemotherapeutic treatments. To date, multiple *XPA* mutations observed in patients with XP-like symptoms are

still classified as VUS on ClinVar. Our current understanding of their contribution to the phenotype observed is still lacking, highlighting a requirement for the functional assessment of variants in XPA.

XPA orthologues are present in almost all species ranging from human to yeast, and demonstrate varying degrees of sequence divergence (Tanaka et al., 1989, Shimamoto et al., 1995, Bankmann et al., 1992, Galande et al., 2021). Indications of the ability of diverged *XPA* orthologues to complement human cells come from studies performed over 25 years ago. Mouse *XPA* restores protection against UV in human XP-A group cells (Tanaka et al., 1989). However, expression of the *Drosophila melanogaster* or *Hydra vulgaris XPA* orthologues in human cells only partially restores UV-protection (Shimamoto et al., 1995; Galande et al., 2021). To my knowledge, additional orthologue complementation of *XPA* in human cells has not been performed. Collectively, the results show that significantly diverged orthologues of *XPA* demonstrate at least partial ability to complement human cells.

Overall, *XPA* is fundamental to the NER pathway, and *XPA* deficiency gives rise to extreme sensitivity to UV radiation (Sugitani et al., 2016). Additionally, *XPA* is highly conserved throughout the animal kingdom and multiple *XPA* orthologues demonstrate at least partial ability to complement human cells (Tanaka et al., 1989, Shimamoto et al., 1995, Bankmann et al., 1992, Galande et al., 2021). To date, many variants in XPA remain unclassified on ClinVar and warrant further investigation (Landrum et al., 2014). Furthermore, XPA is known to undergo several PTMs, with the role of some of these alterations remaining unclear (Hornbeck et al., 2015). Thus, *XPA* is a suitable first candidate to assess the methodology of orthologue complementation to reveal functional amino acid variants.

4.2 Aims and objectives

In this chapter, the complementation ability of vertebrate and invertebrate orthologues of *XPA* will be assessed via IP for a known binding partner and UV survival assays. Subsequently, sequence alignments of complementing orthologues will be constructed and assessed as to whether they can identify functional VUS and non-essential PTM sites in XPA. To validate the methodology, variants identified in complementing orthologues will be introduced into the human protein. Furthermore, PTM sites present in complementing orthologues as non-modifiable residues will be changed to non-modifiable residues in the human protein. These will then be expressed in human cells and UV survival assays will be performed to confirm that sequence alignments of complementing orthologues can identify tolerated variants and non-essential PTM sites. Additionally, the data will be compared to existing publications on variant effects and public variant databases to assess overall concordance.

4.3 Results

4.3.1 Selection of vertebrate XPA orthologues for complementation of human cells

To assess the ability of XPA orthologues to complement human cells, orthologues from discrete taxonomic groups with varying sequence-divergence from human were selected. Orthologue sequences for *Homo sapiens* (human), *Mus musculus* (mouse), *Gallus gallus* (chicken), *Xenopus laevis* (frog) and *Danio rerio* (fish) XPA were obtained from https://www.ncbi.nlm.nih.gov/ with accession numbers displayed in **Table 3-1.** Multiple accession numbers exist for each orthologue, and thus the entries showing the closest lengths to human were selected for each species. Sequences from each protein of interest were then aligned using the T-Coffee algorithm (https://www.ebi.ac.uk/Tools/msa/tcoffee) (Notredame et al., 2000) and the conservation of identical or similar residues was indicated using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html) **Figure 4-1**. The percentage identity to human XPA was then calculated using the % identity matrix output option from the T-Coffee alignment **Table 4-1.**

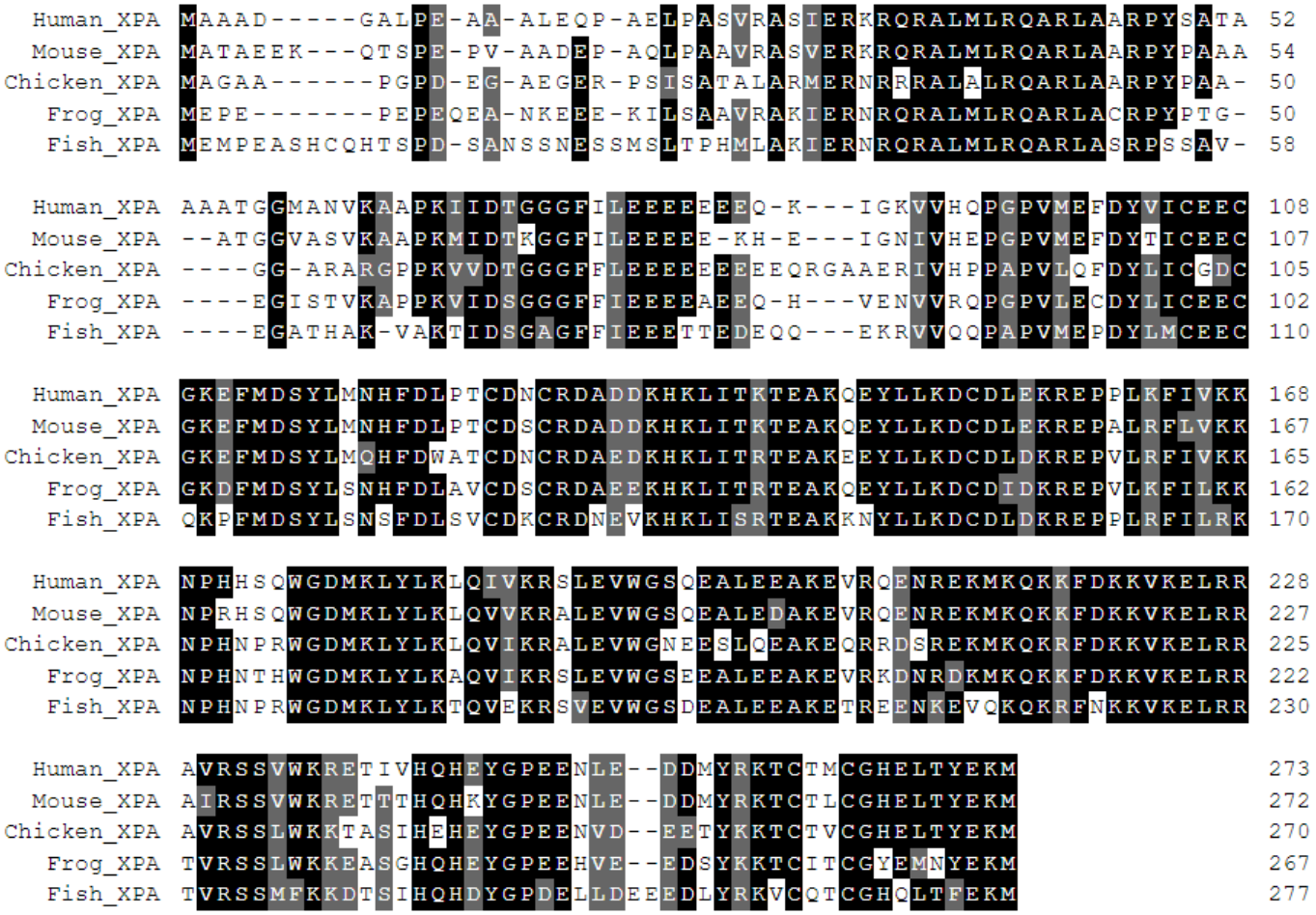


Figure 4-1: Sequence alignment of vertebrate XPA orthologues used in this study with conservation shading

Sequence alignment of vertebrate XPA orthologues used in this study. Sequences were aligned using the T-Coffee algorithm (Notredame et al., 2000) and conservation was scored using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html). Areas of black shading indicate 80% identical conservation of a residue, areas of gray shading indicate 80% conservation to at least an amino acid with similar biochemical properties. Areas of white shading indicate less than 80% conservation. Numbers indicate AA position.

XPA Orthologue	Accession Number	% Identity to Human Orthologue
Human	NP_000371.1	100
Mouse	NP_035858.2	86
Chicken	NP_990184.3	68
Frog	NP_001081354.1	71
Fish	XP_005159871.1	62

Table 4-1: Accession numbers and % identity to human orthologue for vertebrate XPA orthologues used

4.3.2 Vertebrate XPA orthologues can bind to XPC

The ability of an orthologue to fold, express and complement human cells may be inferred from its ability to interact with known binding partners in a human cell line. To test this, cDNAs encoding the human, mouse, chicken, frog and fish XPA orthologues were cloned into pEGFP-C1. These constructs were then expressed in 293FT cells and their ability to interact with XPC was assessed via immunoprecipitation. XPA binds to XPC as part of the NER process (Bunick et al., 2006). To this end, all vertebrate XPA orthologues could bind to human XPC with the same affinity **Figure 4-2**. This demonstrates that the vertebrate XPA orthologues used here can express, fold and bind to a known binding partner in human cells.

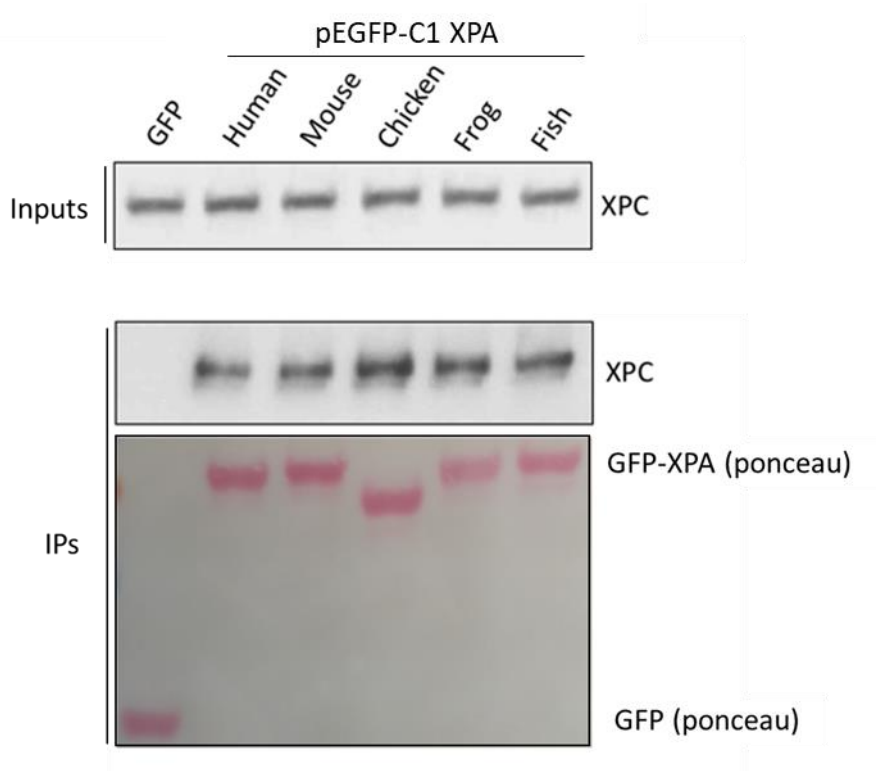


Figure 4-2: Vertebrate XPA orthologues can bind to human XPC

Western blot depicting GFP-pulldowns from 293FT cells transfected with GFP-tagged human, mouse, chicken, frog and fish XPA orthologues showing all can bind to human XPC. Ponceau staining demonstrates expression of GFP-tagged XPA orthologues and respective molecular weights. N=3. 1 repeat perform by Dr. Samuel E. Jones

4.3.3 Vertebrate XPA orthologues can restore protection against UV in human cells

XPA-deficiency renders cells extremely sensitive to UV radiation (Tanaka et al., 1990; Nakane et al., 1995) with complementation of *XPA* rescuing this defect (Tanaka et al., 1990). Thus, the ability of the vertebrate *XPA* orthologues to complement *XPA*-deficiency and restore protection against UV radiation was assessed. pEGFP-C1 constructs encoding vertebrate XPA orthologues were transfected into HAP1 Δ*XPA* cells **Figure 4-3**. GFP+ cells were then sorted via FACS **Figure 10-1** the following day, seeded as single cells and subsequently irradiated with different doses of UV. Cells were then left to grow for 7 days and then fixed, stained and counted to determine the number of surviving colonies. All vertebrate XPA orthologues could restore protection against UV to the same extent as WT cells, suggesting they are proficient in NER in response to UV radiation **Figure 4-3.**

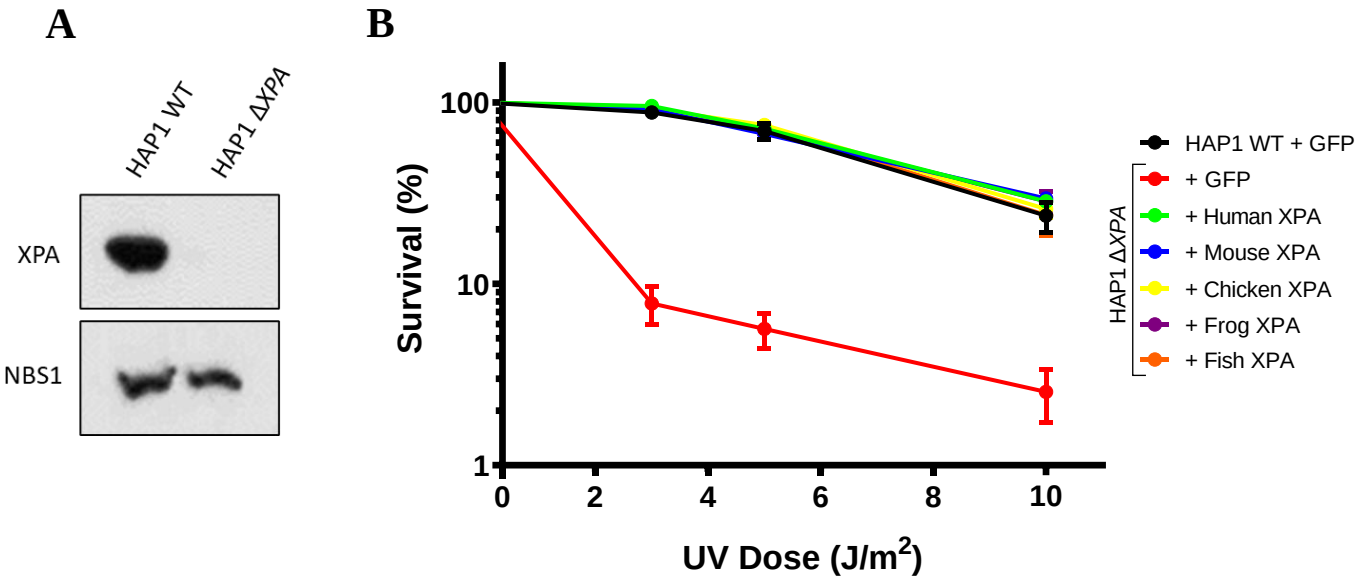


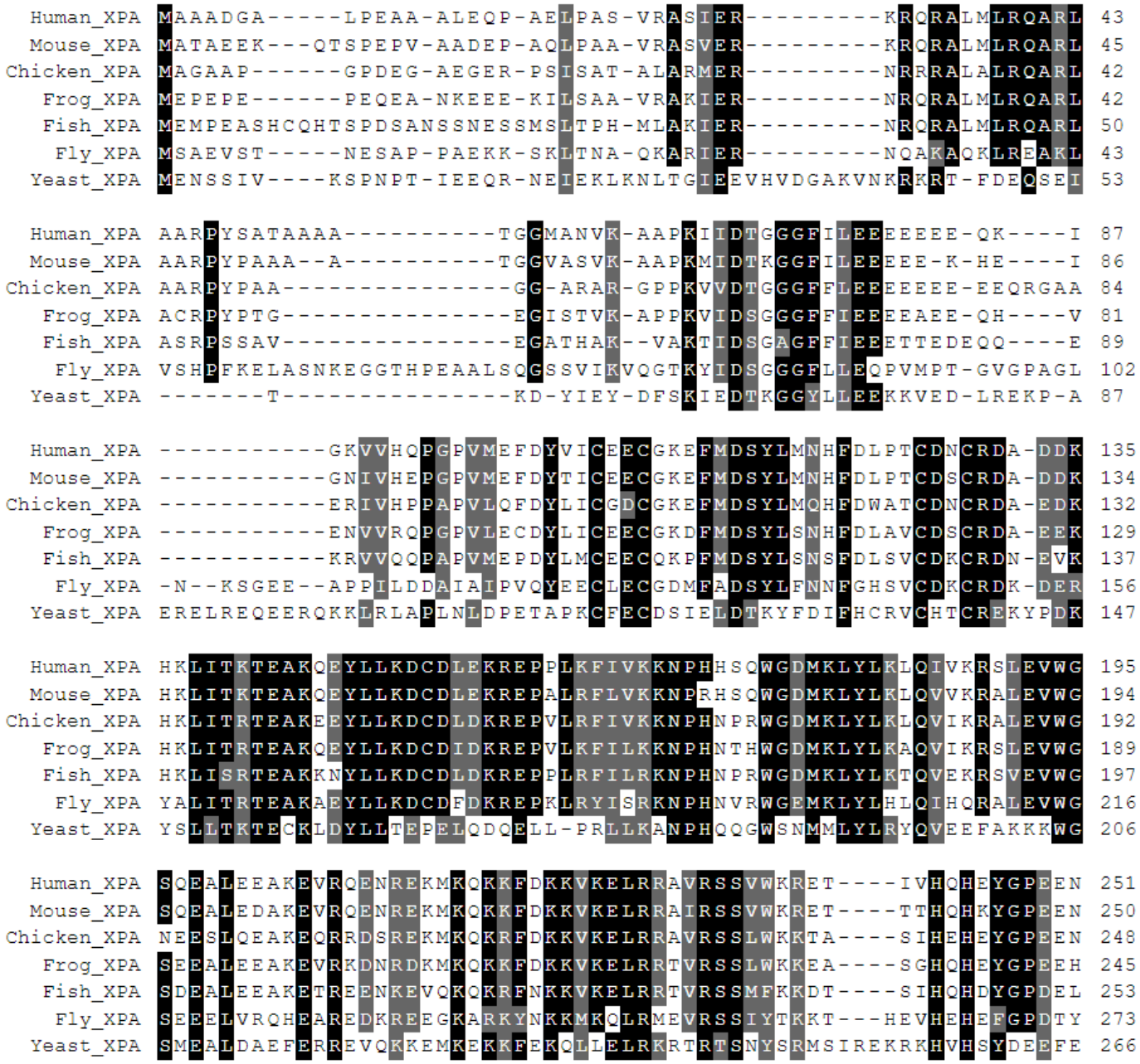
Figure 4-3: **Vertebrate XPA orthologues can restore protection against UV in human cells**

A: Western blot showing XPA protein levels in HAP1 WT and HAP1 Δ*XPA* cells with NBS1 used as a loading control. **B:** Survival curves demonstrating the colony survival % compared to untreated cells in response to increasing doses of UV radiation for HAP1 WT and HAP1 Δ*XPA* cells complemented with GFP-tagged XPA orthologues or GFP only controls. Data is plotted as the mean of N=3. Error bars indicate SEM.

4.3.4 Invertebrate XPA orthologues cannot complement human cells

Having demonstrated that vertebrate *XPA* orthologues were able to complement *XPA*-deficient cells, the ability of further diverged *XPA* orthologues was assessed. Obtaining complementation with a more diverged orthologue may reveal additional functional variation. To assess this, XPA orthologues from *Drosophila melanogaster* (fruit fly) and *Schizosaccharomyces pombe* (fission yeast) were obtained from https://www.ncbi.nlm.nih.gov/ with accession numbers displayed in **Table 3-1.** Multiple accession numbers existed for each orthologue and thus the entries showing the closest lengths to human were selected for each. Sequences from each protein of interest were then aligned using the T-Coffee algorithm (https://www.ebi.ac.uk/Tools/msa/tcoffee) (Notredame et al., 2000) and the conservation of identical or similar residues was indicated using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html) **Figure 4-4.** The percentage identity to human XPA was then calculated using the % identity matrix output option from the T-Coffee alignment **Table 4-2**. To assess complementation ability, the fly and yeast *XPA* orthologues were cloned into the pEGFP-C1 vector. Subsequently**,** these constructs were transfected into 293FT cells and their ability to interact with human XPC was tested via immunoprecipitation **Figure 4-5**. Both fly and yeast XPA could fold and express in human cells but both showed reduced affinity for human XPC **Figure 4-5**. The ability of these orthologues to restore protection against UV was then assessed. pEGFP-C1 constructs encoding fly and yeast XPA were transfected into HAP1 Δ*XPA* cells. GFP+ cells were then sorted via FACS **Figure 10-2** the following day, seeded as single cells, and subsequently irradiated with different doses of UV. Cells were then left to grow for 7 days and then fixed, stained and counted to determine the number of surviving colonies. Fly XPA demonstrated only partial ability to restore protection against UV compared to human and fish XPA, acting as fully complementing controls **Figure 4-5**. This results is consistent with results previously obtained for the ability

of fly XPA to only partially restore UV-protection (Shimamoto et al., 1995). In contrast, yeast XPA demonstrated almost a complete lack of ability to restore protection against UV radiation in human cells **Figure 4-5**. Collectively this shows that the complementation ability of *XPA* orthologues rapidly diminishes with increasing divergence from vertebrate orthologues.



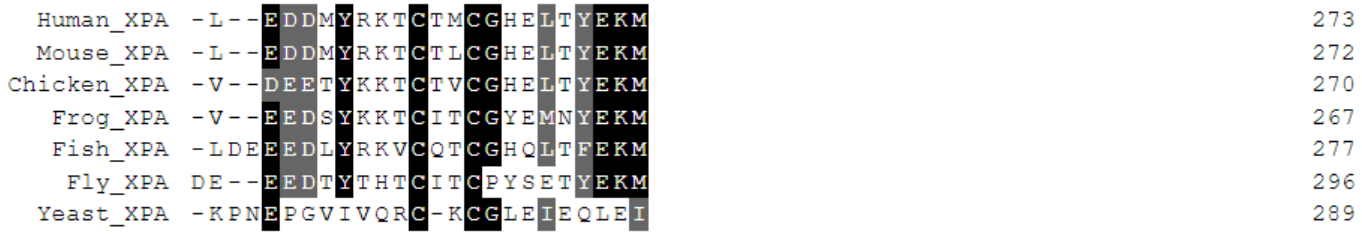


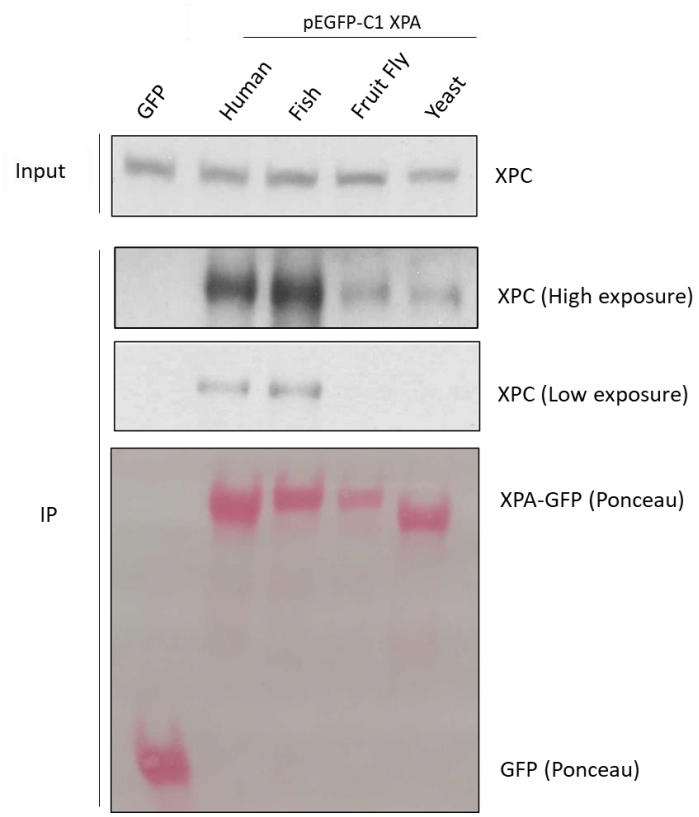
Figure 4-4: Sequence alignment of vertebrate and invertebrate XPA orthologues used in this study with conservation shading

Sequence alignment of vertebrate and invertebrate XPA orthologues used in this study. Sequences were aligned using the T-Coffee algorithm (Notredame et al., 2000) and conservation was scored using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html). Areas of black shading indicate 80% identical conservation of a residue, areas of grey shading indicate 80% conservation to at least an amino acid with similar biochemical properties. Areas of white shading indicate less than 80% conservation. Numbers indicate AA position.

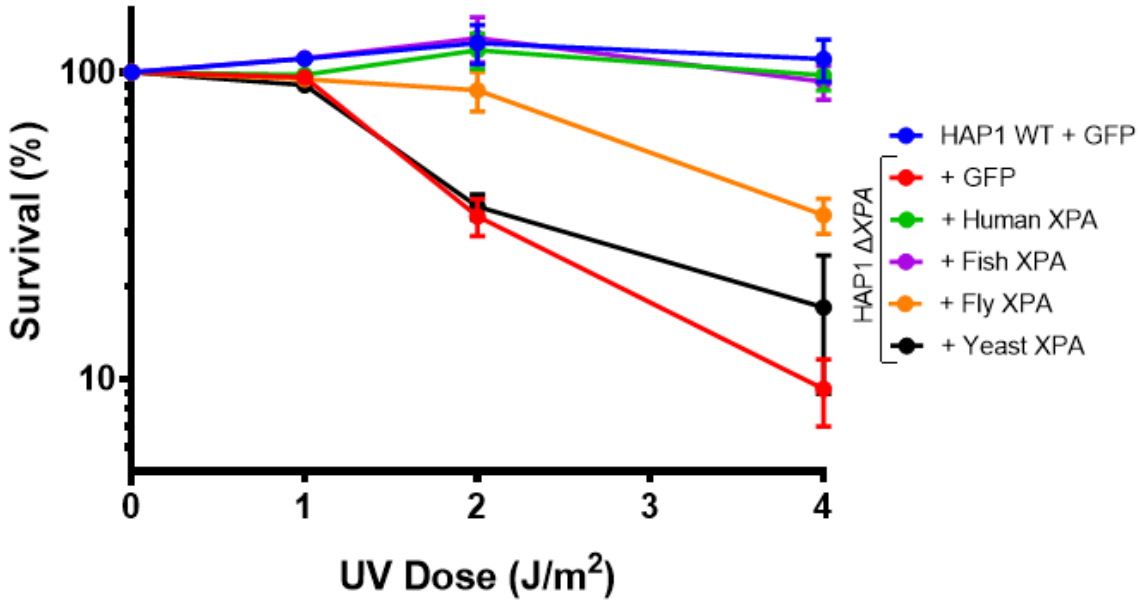
XPA Orthologue	Accession Number	% Identity to Human Orthologue
Human	NP_000371.1	100
Mouse	NP_035858.2	86
Chicken	NP_990184.3	68
Frog	NP_001081354.1	71
Fish	XP_005159871.1	62
Fly	NP_476866.1	45
Yeast	NP_595222.1	33

Table 4-2: Accession numbers and % identity to human orthologue for vertebrate and invertebrate XPA orthologues used

A


pEGFP-C1 XPA
GFP
Human
Fish
Fruit Fly
Yeast
Input
XPC
XPC (High exposure)
XPC (Low exposure)
IP
XPA-GFP (Ponceau)
GFP (Ponceau)

B


Survival (%)
100
10
0
1
2
3
4
UV Dose (J/m²)
HAP1 WT + GFP
HAP1 ΔXPA
+ GFP
+ Human XPA
+ Fish XPA
+ Fly XPA
+ Yeast XPA

C

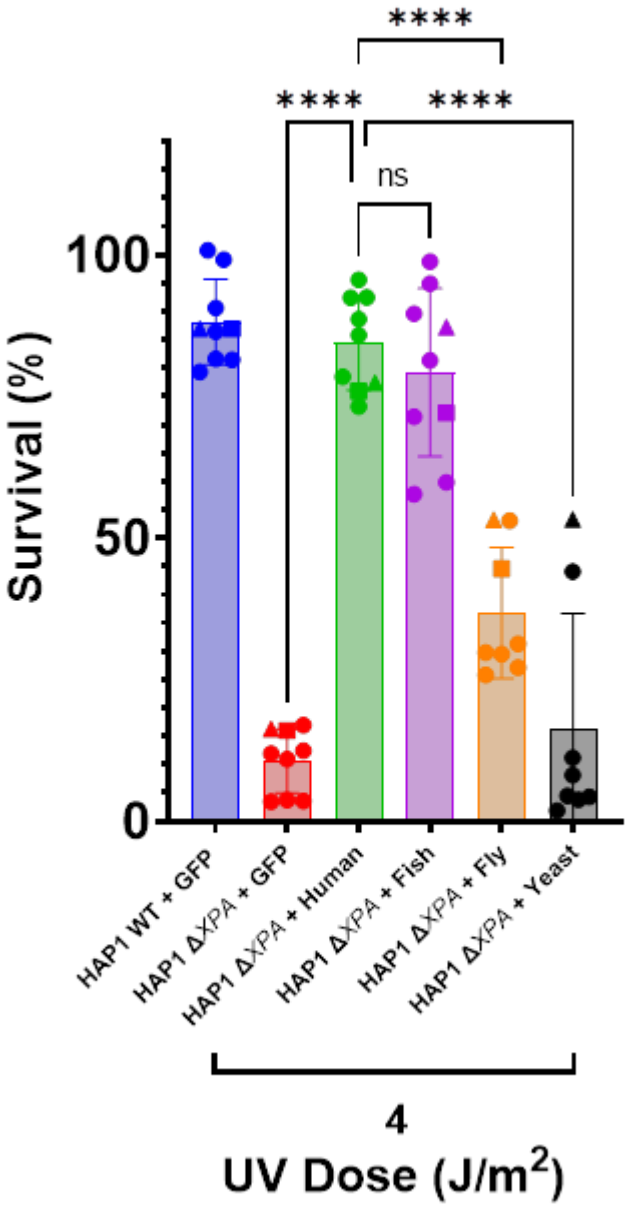


Figure 4-5: Invertebrate XPA orthologues cannot complement human cells

A: Western blot depicting GFP-pulldowns from 293FT cells transfected with GFP-tagged human, fish, fruit fly and yeast XPA orthologues with probing for XPC. Ponceau staining demonstrates expression of GFP-tagged XPA orthologues and respective molecular weights, N=2. **B**: Survival curves demonstrating the colony survival % compared to untreated cells in response to increasing doses of UV radiation for HAP1 WT and HAP1 Δ*XPA* cells complemented with GFP-tagged XPA orthologues or GFP only controls. Data is plotted as the mean of N=3. Error bars indicate SEM. **C:** One-way ANOVA was performed on all values of N=3 at 4 J/m^2 UV and included a Dunnett's multiple comparisons test comparing HAP1 Δ*XPA* + GFP-WT XPA to HAP1 Δ*XPA* + GFP, GFP-fish XPA, GFP-fly XPA and GFP-yeast XPA. Results from Dunnett's multiple comparisons test are displayed. ns = non-significant difference, **** = p <0.0001.

4.3.5 A sequence alignment of complementing XPA orthologues contains known benign variants and does not contain known pathogenic variants

The overall aim of the methodology is to use sequence alignments of complementing orthologues to identify functional amino acid variants tolerated by the human protein. As a proof of principle, I assessed whether known benign variants in the human protein were present at any of the corresponding positions in the fully complementing orthologues. The prominent source of benign variation is the gnomAD database, curated to catalogue variation identified in healthy individuals (Karczewski et al., 2020). Thus, the data for all missense mutations in *XPA* catalogued on the gnomAD database was downloaded. The dataset was then filtered to show *XPA* missense mutations present homozygously in at least one individual, as this eliminates any potential masking of the deleteriousness of an allele by the WT copy. I then searched the sequence alignment to determine whether any of these variants were present at the corresponding position in complementing XPA orthologues. Importantly, S190A is present in mouse and chicken XPA and L252V is present in chicken and frog XPA, **Table 4-3**. This reveals that sequence alignments of complementing orthologues contain known benign XPA variants. The ClinVar database also provides a source of benign XPA variants. Therefore, as performed for gnomAD variants, the sequence alignment was searched for benign XPA variants present at the corresponding positions in complementing orthologues. However, the sole XPA variant with a classification of benign, V166A, is not present in any of the orthologues **Table 4-4**. Evidently, identification of a known pathogenic human variant in a complementing orthologue would prevent the use of the alignment to identify tolerated variants. To assess whether this was the case, the sequence alignment was searched for XPA variants classified as pathogenic on ClinVar. The analysis confirmed that no pathogenic XPA variants are present in complementing orthologues **Table 4-5**.

rsIDs (gnomAD identifier)	Residue position	WT	Variant	Present in	Allele frequency	Homozygote count
rs10983315	97	V	I	None	0.058%	1
rs143374902	166	V	A	None	0.098%	1
rs555812588	190	S	A	Mouse, chicken	0.077%	2
rs3176750	252	L	V	Chicken, frog	0.35%	16
rs57519506	256	M	V	None	0.28%	12

Table 4-3: Results of searching the sequence alignment of complementing orthologues for homozygous gnomAD XPA variants

The sequence alignment of complementing XPA orthologues (human-fish) was searched for homozygous XPA variants present in the gnomAD database. Analysis revealed 2 homozygous variants, L252V which is present in chicken and frog XPA and M256V which is present in mouse and chicken XPA.

ClinVar Accession Number	Residue Position	WT	Variant	Present in?
NM_000380.4(XPA):c.497T>C (p.Val166Ala)	166	V	A	None

Table 4-4: Results of searching the sequence alignment of complementing orthologues for benign XPA variants on ClinVar

The sequence alignment of complementing XPA orthologues (human-fish) was searched for benign XPA missense mutations catalogued on ClinVar. The sole XPA missense mutation with a benign classification, V166A, was not present in any of the complementing XPA orthologues used.

ClinVar Accession Number	Residue Position	WT	Variant	Present in?
NM_000380.4(XPA):c.555G>C (p.Gln185His)	185	Q	H	None

Table 4-5: Results of searching the sequence alignment of complementing orthologues for pathogenic XPA variants on ClinVar

The sequence alignment of complementing XPA orthologues (human-fish) was searched for pathogenic XPA missense mutations from the ClinVar database. No pathogenic XPA variants were present in complementing orthologues.

4.3.6 A sequence alignment of complementing XPA orthologues reveals functional amino acid variants in the human protein

Having determined that known benign XPA variants exist in complementing XPA orthologues, the sequence alignment of complementing orthologues was used to determine the functionality of variants in XPA from the ClinVar database. The sequence alignment was searched for XPA variants catalogued on ClinVar without sole classifications of either benign or pathogenic. The analysis identified 4 variants present in complementing orthologues with classifications of either uncertain significance, conflicting interpretations of pathogenicity, or likely benign. Specifically, I68T and L191V were identified in fish XPA, K141R in chicken, frog and fish XPA and V234L in chicken and frog XPA. Furthermore, S190A was detected in mouse and chicken and L252V was identified in chicken and frog **Table 4-6**. However, both S190A and L252V were excluded from this analysis as they were previously identified in the gnomAD database as homozygous variants, as explained in section 4.3.5.

ClinVar Accession Number	Residue Position	WT	Variant	Variant present in	ClinVar Classification	Encoded across an exon-boundary	Published functional data
NM_000380.4(XPA):c.203T>C (p.Ile68Thr)	68	I	T	Fish	Uncertain significance	No	None
NM_000380.4(XPA):c.571C>G (p.Leu191Val)	191	L	V	Fish	Conflicting interpretations of pathogenicity	No	None
NM_000380.4(XPA):c.568T>G (p.Ser190Ala)	190	S	A	Mouse, chicken	Benign/likely benign	No	None
NM_000380.4(XPA):c.422A>G (p.Lys141Arg)	141	K	R	Chicken, frog, fish	Uncertain significance	No	None
NM_000380.4(XPA):c.700G>T (p.Val234Leu)	234	V	L	Chicken, frog	Likely benign	No	Restores protection against UV in an XPA-deficient cell line (Mellon et al., 2002)
NM_000380.4(XPA):c.754C>G (p.Leu252Val)	252	L	V	Chicken, frog	Benign/likely benign	No	None

Table 4-6: Results from searching the alignment of complementing XPA orthologues for ClinVar XPA variants not classified as either 'benign' or 'pathogenic'

The sequence alignment of complementing XPA orthologues (human-fish) was searched for XPA variants on the ClinVar database not classified as either 'benign' or 'pathogenic'. This revealed multiple variants present in complementing orthologues which require additional functional information.

As a proof of principle, I next determined whether the I68T, L191V, K141R and V234L variants were in fact tolerated by the human protein. I68T, L191V, K141R and V234L variants **Figure 4-6** were introduced into individual pEGFP-C1 human XPA constructs via site-directed mutagenesis. It is recommended by the ACMG/AMP guidelines for variant interpretation that known pathogenic variants should be included as controls in any functional assays (Richards et al., 2015, Brnich et al., 2019). However, the only pathogenic missense mutation in XPA on ClinVar is Q185H, which is known to affect the pre-mRNA splicing of XPA (Satokata et al.,

1992, States et al., 1998). Due to the nature of this method, variants which disrupt splicing cannot be studied as cDNA is used to complement human cells. To this end, the amino acids encoded by the last 3 bases of an exon and the 1st base of the subsequent exon position are marked in **Table 4-7**. Any variant encoded by bases must be excluded from the analysis, due to the alteration of bases encoding these residues potentially disrupting pre-mRNA splicing (Ma et al., 2015). Importantly, all variants tested here are not encoded across at such positions.

Exon number	**Amino acids**	**Amino acids encoded across exon boundaries**
Exon 1	1-58	G57, G58
Exon 2	59-95	P94, G95
Exon 3	96-130	C129, R130
Exon 4	131-185	Q185, I186
Exon 5	186-225	K224, E225
Exon 6	226-273	N/A

Table 4-7: Amino acid positions of XPA exons with amino acids encoded across exon boundaries

Table displaying the amino acid positions of the XPA exons and the amino acids encoded across the exon boundaries. Exon positions were obtained from CCDS https://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi?REQUEST=CCDS&DATA=CCDS6729.1.

To ascertain whether the variants present in complementing orthologues were functional in human XPA, the pEGFP-C1 constructs encoding XPA the variants were transfected into HAP1 Δ*XPA* cells. GFP+ cells were then sorted via FACS **Figure 10-3** the following day, seeded as single cells, and subsequently irradiated with different doses of UV. Cells were then left to grow for 7 days, and then fixed, stained and counted to determine the number of surviving colonies. Introduction of either I68T, L191V, K141R or V234L into human XPA did not affect its ability to restore protection against UV **Figure 4-6.** This demonstrates that sequence alignments of complementing orthologues identifies variants which can be tolerated by the human protein. Importantly, this enables the assessment of variants with uncertain clinical significance from databases such as ClinVar providing further functional information for their classification.

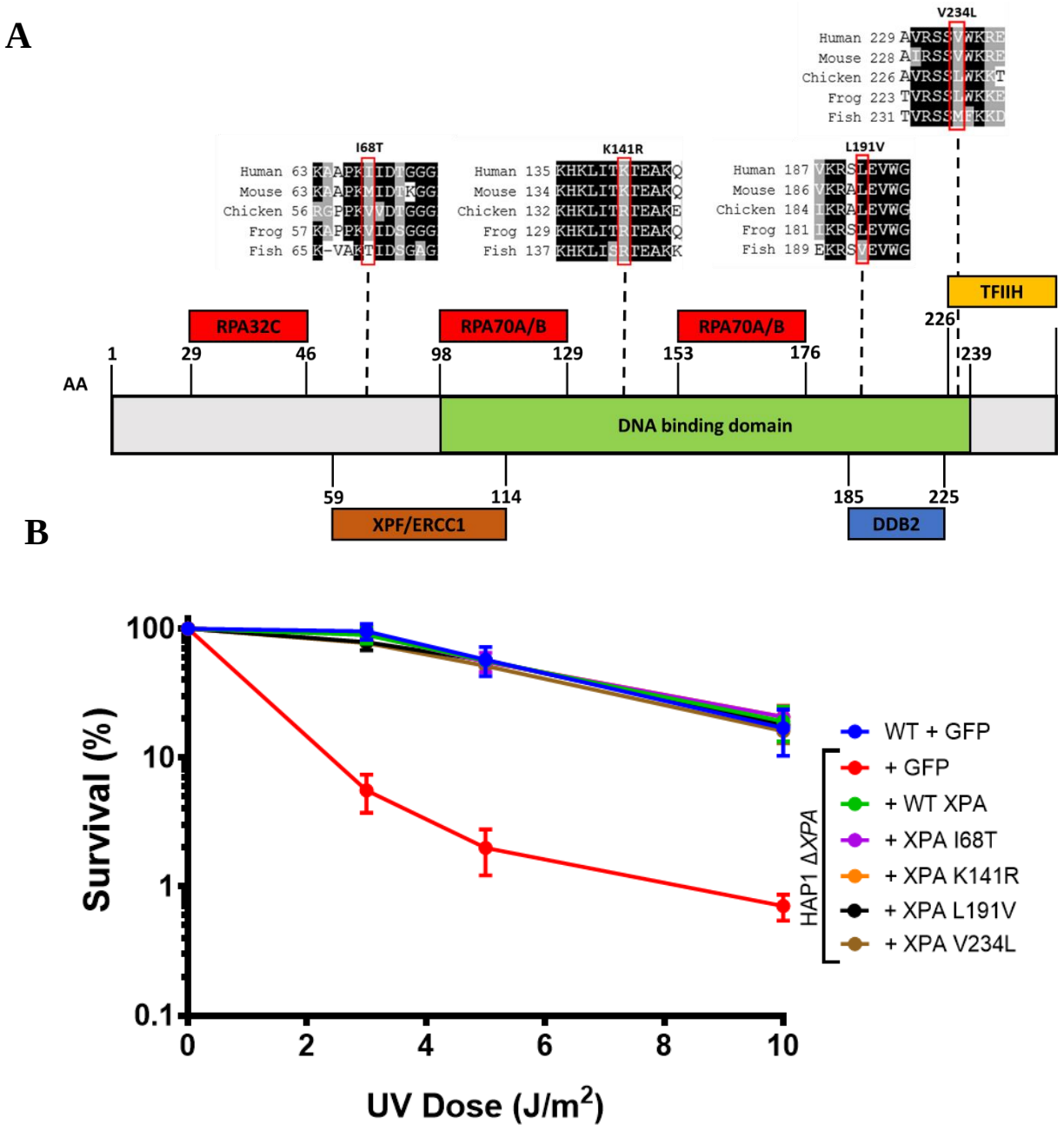


Figure 4-6: Variants present in complementing orthologues do not perturb XPA function in response to UV radiation

A: Schematic showing the positions of ClinVar variants introduced into XPA with sequence alignments, amino acid residues and binding domains displayed. **B**: Survival curves demonstrating the colony survival % compared to untreated cells in response to increasing doses of UV radiation for HAP1 WT and HAP1 Δ*XPA* cells complemented with GFP-tagged human XPA containing one of I68T, K141R, L191V or V234L or GFP only controls. Data is plotted as the mean of N=3. Error bars indicate SEM.

.

4.3.7 A sequence alignment of complementing XPA orthologues reveals non-essential PTM sites

Variants present in complementing orthologues could be tolerated by the human protein. Therefore, I hypothesised that sequence variation to non-modifiable residues in complementing orthologues at positions of PTMs in the human protein may indicate they are non-essential for XPA protein function. To determine this, the PTMs reported for XPA were downloaded from the PhosphoSitePlus database (Hornbeck et al., 2015). The sequence alignment of complementing orthologues was then searched for variation to non-modifiable residues at positions of PTMs in the human XPA protein. This revealed several PTM sites which may not be required for XPA function in human cells. Specifically, phosphorylation of S173, S196 and T239, SUMOylation of K86, K141 and K63 and ubiquitination of K141 **Table 4-8**.

Position	**Residue**	**Modification**	**Non-modifiable residue present in**	**Encoded across an exon boundary?**	**Reported function**
63	K	SUMOylation	Chicken (R)	No	None
86	K	SUMOylation	Mouse (E), Chicken (Q), Frog (H), Fish (Q)	No	None
141	K	SUMOylation, ubiquitination	Chicken (R), Frog (R), Fish (R)	No	None
173	S	Phosphorylation	Chicken (P), Fish (P)	No	None
196	S	Phosphorylation	Chicken (N)	No	Promotes survival after UV irradiation (Wu et al., 2006b). Prevents HERC-2 mediated ubiquitination (Lee et al., 2014b). Facilitates retention of XPA on chromatin (Lee et al., 2014b, Nguyen et al., 2010).
239	T	Phosphorylation	Chicken (A), Frog (A)	No	None

Table 4-8: Results from searching the sequence alignment of complementing XPA orthologues for variation at PTM sites in XPA reported on the PhosphoSitePlus database

The sequence alignment was searched for all existing PTMs catalogued for XPA on PhosphoSitePlus to identify whether non-modifiable residues are present in complementing orthologues at these positions.

I showed in **Figure 4-6** that mutation of K141 to an arginine, which cannot be ubiquitinated or SUMOylated (Callis, 2014, Wilkinson and Henley, 2010), does not impact cellular survival in response to UV radiation. Therefore, the focus shifted to phosphorylated residues, as phosphorylation is the primary way in which proteins of the DNA damage response are regulated (Blackford and Jackson, 2017). Furthermore, phosphorylation sites required for DDR protein function are often highly conserved (Blackford et al., 2015, Leimbacher et al., 2019). S196 is phosphorylated by ATR in response to UV-radiation (Wu et al., 2006b), promoting cellular survival (Wu et al., 2006b, Lee et al., 2014b, Nguyen et al., 2010). S173 is also phosphorylated by ATR in response to UV radiation, but has no effect on survival (Wu et al., 2006b). T239 was identified via antibody purification methodologies and therefore lacks direct functional testing. The methodology described here predicts that all 3 sites are not required for XPA function in response to UV radiation.

Having predicted that phosphorylation of S173, S196 and T239 is not required for XPA function in response to UV radiation, this was validated in a cellular model. Phosphorylation sites within DDR proteins are can experience functional redundancy with one another (Blackford and Jackson, 2017). To prevent this from masking whether or not S173, S196 and T239 are required for protein function, all 3 were mutated to alanine in a pEGFP-C1 construct encoding human XPA via site-directed mutagenesis to create a '3A' XPA mutant **Figure 4-7**. pEGFP-C1 constructs encoding '3A' XPA were then transfected into HAP1 Δ*XPA* cells. GFP+ cells were then sorted via FACS the following day **Figure 10-3**, seeded as single cells and subsequently irradiated with different doses of UV. Cells were then left to grow for 7 days and then fixed, stained and counted to determine the number of surviving colonies. Mutation of all 3 residues to alanine did not result in a survival defect in response to UV upon complementation

of HAP1 Δ*XPA* cells **Figure 4-7**. Overall, this shows that sequence alignments of complementing orthologues can identify non-essential PTM sites in the human XPA protein.

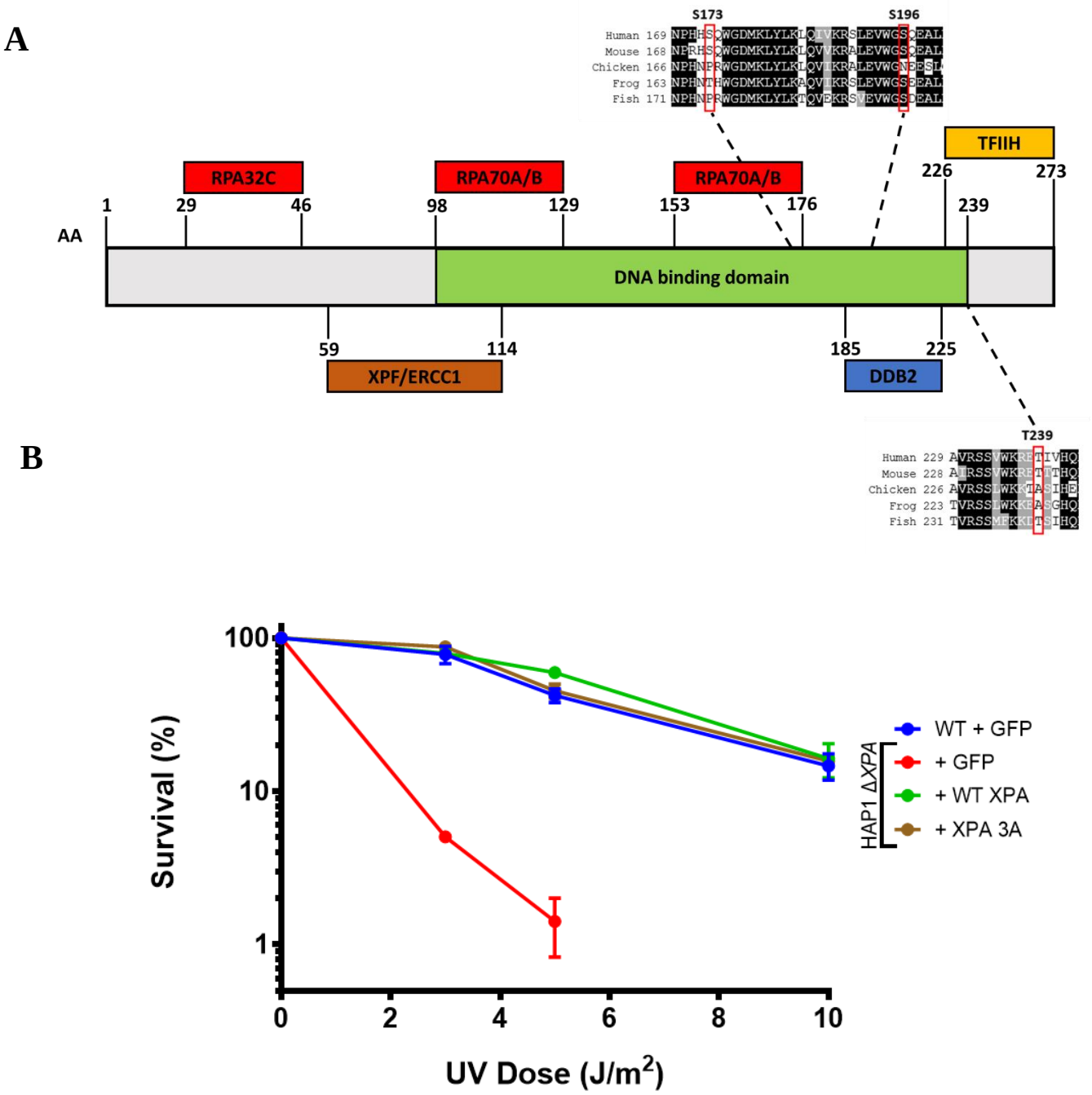


Figure 4-7: Variation to non-modifiable residues in complementing orthologues reveals non-essential PTM sites

A: Schematic showing the positions of reported PTM sites in XPA where alanine residues were introduced with sequence alignments, amino acid residues and binding domains displayed**. B:** Survival curves demonstrating the colony survival % compared to untreated cells in response to increasing doses of UV radiation for HAP1 WT and HAP1 Δ*XPA* cells complemented with GFP-tagged human XPA or GFP-tagged human XPA '3A' or GFP only controls. Data is plotted as the mean of N=3. Error bars indicate SEM.

4.3.8 Comparison to bioinformatic predictions of variant effects

Predictions of variant effects from functional assays are frequently compared to predictions from algorithms (Peterson et al., 2013). Such predictors use metrics such as evolutionary-conservation, amino acid physiochemical properties, and whether or not that amino acid exists within structured regions of the protein (Peterson et al., 2013). The number of algorithms used to predict variant effects are ever growing (Livesey and Marsh, 2022). However, algorithms that are commonly used (Peterson et al., 2013, Livesey and Marsh, 2022) include Align_GVGD (Tavtigian et al., 2006), SIFT (Ng and Henikoff, 2003) and PolyPhen-2 (Adzhubei et al., 2010). All variants tested in this chapter were predicted to be at least likely benign by all 3 algorithms **Table 4-9.** The concordance between functional data obtained here and *in silico* predictions should facilitate the reclassification of the variants tested to at least likely benign (Richards et al., 2015).

Variant	**Align_GVGD**	**SIFT Score**	**PolyPhen-2**
I68T	Class C0 (tolerated)	0.32 (tolerated)	Benign (0.003)
K141R	Class C0 (tolerated)	1 (tolerated)	Benign (0.009)
L191V	Class C0 (tolerated)	0.38 (tolerated)	Benign (0.22)
V234L	Class C0 (tolerated)	1 (tolerated)	Benign (0)
S173A	Class C0 (tolerated)	0.23 (tolerated)	Benign (0.015)
S196A	Class C25 (likely tolerated)	0.07 (tolerated)	Benign (0.021)
T239A	Class C0 (tolerated)	0.7 (tolerated)	Benign (0.003)

Table 4-9: Computational predictions for XPA variants tested

Computation predictions were obtained using the Align-GVGD, SIFT and PolyPhen-2 algorithms for the variants tested in this chapter. All scores are listed and their associated classifications provided.

4.4 Discussion

4.4.1 Summary of the work presented in this chapter

In this chapter I have demonstrated that XPA orthologues from mouse, chicken, frog and fish can complement human cells, whereas orthologues from fly and yeast were unable to do so. This was demonstrated via immunoprecipitation for a known binding partner (XPC) and the ability to restore protection against UV radiation in an XPA-deficient human cell line. Vertebrate XPA orthologues could all bind to human XPC to a similar extent **Figure 4-2** whereas invertebrate orthologues pulled down a reduced amount of human XPC **Figure 4-5**. Furthermore, vertebrate orthologues could all restore protection against UV in an XPA-deficient human cell line to the same extent **Figure 4-3** whereas invertebrate XPA orthologues only showed partial ability to do so **Figure 4-5**. It was then shown that the sequence alignment of complementing XPA orthologues contained known benign variants but did not contain known pathogenic variants **Table 4-3**, **Table 4-5**. Subsequently, VUS present at the corresponding position in a complementing orthologue **Table 4-6** were tolerated by the human protein, evidenced by not disrupting XPA's ability to restore protection against UV radiation **Figure 4-6**. Following this, PTM sites which had diverged to non-modifiable residues in complementing orthologues **Table 4-8** were shown to be non-essential for XPA function again via full restoration of UV protection **Figure 4-7**. In conclusion, I have shown that a sequence alignment of complementing orthologues can be sued to identify functional variants and non-essential PTM sites in the human protein. Overall, a sequence alignment of complementing XPA orthologues provides a reference for functional amino acid variants in the XPA protein

4.4.2 A discussion of the findings presented in this chapter

The ability of XPA orthologues to complement human cells has been studied previously. Mouse XPA was previously shown to complement human cells by restoring protection against UV-radiation (Tanaka et al., 1989), consistent with results here. Both fly and hydra XPA

orthologues demonstrate partial ability to restore protection against UV irradiation in human cell lines (Shimamoto et al., 1995, Galande et al., 2021). The findings presented in this chapter are consistent with previous finding regarding fly XPA (Shimamoto et al., 1995). Hydra XPA shows 48% similarity to the human protein used here (Galande et al., 2021), with fly showing 45% similarity. The most diverged orthologue shown to fully complement human cells in this chapter was fish XPA at 62% similarity to human. Based on the data shown here and existing publications, it is possible that full complementation of human cells could be achieved with an orthologue showing ≤ 62% but ≥ 49% similarity to the human protein, potentially revealing additional functional variants within human XPA..

The findings presented in figure show that both fly and yeast pull down reduced levels of human XPC **Figure 4-5**. XPA interacts with XPC via amino acid residues 154-334 of XPC (Bunick et al., 2006), but the residues in XPA required for this interaction are yet to be elucidated. Therefore it cannot be discussed as to which residues may have diverged causing the weakening of this interaction. However, ponceau staining of the GFP-XPA constructs presented in that figure appear to show reduced expression of the fly-XPA protein. Therefore, the reduced amount of human XPC pulled down by this orthologue may not be due to a reduced affinity per se, but rather the result of less fly XPA protein expressed compared to human XPA. This could be further clarified via quantitative analysis of the protein-protein interactions. One method which could be used is to quantify the immunoblots displayed in figure using image quantification software such as imageJ (Schneider et al., 2012, Burckhardt et al., 2021). Using this, the ratio of expressed bait protein compared to the levels of XPC pulled down normalized to the XPC present in the input lanes could be quantified (Burckhardt et al., 2021). However, analysis of protein-protein interactions in this way is limited by the semi-quantitative nature of immunoblotting (Janes, 2015, Mahmood and Yang, 2012). ECL-based signal amplification of HRP conjugated secondary antibodies (as performed here) is non-linear and thus is not

recommended for use in determining the concentration of protein present (Mahmood and Yang, 2012, Janes, 2015).

There are several methods which more quantitatively assess the extent of protein-protein interaction that could have been employed in addition to the IPs performed here. One example is the proximity ligation assay (PLA) (Bonagura et al., 2022), which relies on the binding of two complementary DNA strands conjugated to secondary antibodies bound to primary antibodies recognizing each of the two binding partners. Fluorescent oligonucleotides are added and recognize the intact DNA molecule providing a fluorescent foci upon protein-protein interaction. In this way, foci can be quantified to provide a measure of the affinity of the two interacting partners (Bonagura et al., 2022). However, this would rely on the assumption that antibodies raised to human XPA would also recognize XPA orthologues with similar affinity, or require raising antibodies specific to XPA orthologues. Alternatively, Y2H assays could be employed which are based on the fusion of a Gal4 transcription factor activating domain and binding domain, which are fused to each of your proteins of interest (Brückner et al., 2009). Upon binding of your proteins, transcriptional activation of a reporter gene occurs, typically LacZ, which can then be quantified via lysing yeast colonies and colorimetric analysis of the lysate at OD 420nm (Kondo et al., 2001). This would provide an estimation of the affinity for human XPC with XPA orthologues. However, it is dependent on the stability of each orthologue in the yeast system and thus the same expression defect may occur as seen in human cells. One way to remove the stability/expression problem would be to perform in vitro assays such as surface plasmon resonance (SPR). This assay measures the changes in reflective light intensity or resonance angle shift of photons shone at a thin metal surface layer. Your protein of interest is bound to the metal surface and interaction of the binding partner with this changes the angle of the light reflected (Nguyen et al., 2015) In this way, the association of your proteins of interest can be measured but also the dissociation via rinsing the surface with a buffer

(Nguyen et al., 2015) The rates of association and disassociation can be determined providing an accurate picture of the affinity of the interaction (Nguyen et al., 2015). However, this is done in a cell free system, and thus may miss contributing factors to the interaction *in vivo* (Nguyen et al., 2015). Overall, the lack of quantification of the interactions displayed in the IP is a clear limitation of this work, and prevents us gaining a full understanding of the level of complementation observed. Future work should address this limitation to enable a better understanding of the complementation achieved.

In addition to binding to XPC, the ability of XPA to restore protection against UV depends on its ability to bind to DNA (Saijo et al., 2011, Camenisch et al., 2006, Hilton et al., 2014). The residues required for this are K141, K168, K179, K221, K222, K224 and K236. Yeast XPA shows significant divergence at residues K168(A), K179(M), K222(Q) and K236(S) **Figure 4-4**, suggesting that its ability to bind to DNA in human cells may be perturbed. In contrast, fly XPA shows high conservation at all but one of these residues (K236(S)) **Figure 4-4**, suggesting that its ability to bind to DNA may be only slightly disrupted. Disruption to the interaction between ERCC1-XPF and XPA results in reduced protection against UV radiation (Li et al., 1995a). This interaction is mediated by the 'E-cluster' AA 78-84 and the 'G-cluster' AA 72-75 of XPA (Li et al., 1995a, Tsodikov et al., 2007). Deletion of the E-cluster only partially disrupts the ability of XPA to restore protection against UV, whereas deletion of the G-cluster abrogates this function (Li et al., 1995a). The G-cluster is well conserved throughout all orthologues used here **Figure 4-4**. In contrast, the E-cluster has significantly diverged in fly and yeast XPA **Figure 4-4.** This may underlie the partial complementation ability observed for fly XPA. Notably, this is also the case for hydra XPA, and is one of the reasons thought to underlie the partial complementation ability of this orthologue in previous work (Galande et al., 2021). XPA is a metalloprotein and contains a C4 zinc finger motif consisting of residues C105, C108, C126 and C129 (Morita et al., 1996). Disruption to any one of these residues

drastically reduces the survival of UV irradiated cells (Miyamoto et al., 1992). Notably, all 4 of these residues are conserved in all orthologues used here **Figure 4-4**, and therefore likely do not underlie the loss of complementation observed with fly and yeast XPA.

The XPA variants S190A and L252V which are present homozygously in ostensibly healthy individuals (Karczewski et al., 2020), were identified at the corresponding position in at least one complementing XPA orthologue **Table 4-3**. This provided the initial verification that the sequence alignment could identify known benign XPA variants in the human protein. However, ClinVar variants with sole classifications of benign were not present in any of the complementing orthologues. It must be noted that, to date, only 1 benign XPA variant (V166A) exists on ClinVar, providing a very small sample size on which to conduct this analysis. Notably, V166 is well conserved throughout complementing orthologues as valine in mouse and chicken, and the biochemically similar leucine in frog and fish **Figure 4-1**. This is also the case for the homozygous gnomAD variant V97I, which is a valine in all 5 complementing orthologues **Figure 4-1**. This demonstrates that positions of benign amino acid variants may exhibit high conservation in complementing orthologues, and therefore conservation does not invariably imply importance of any given amino acid residue.

Analysis of the ClinVar database revealed 4 variants present in complementing XPA orthologues which did not have sole classifications of 'benign' or 'pathogenic' **Table 4-6**. These were: I68T, K141R, L191V and V234L. I68T and K141R were identified and submitted by Mayo Clinic Laboratories as variants of uncertain significance, but the condition of the affected individual(s) was not specified. Here, I68T was identified in fish and K141R was identified in chicken, frog and fish, **Table 4-6** with both then shown to be tolerated by the human protein **Figure 4-6**.

I68T is an unconservative substitution from a non-polar hydrophobic isoleucine to a polar uncharged threonine. Conservative variation exists at other complementing orthologues at this position with the non-polar hydrophobic methionine in mouse and the non-polar hydrophobic valine in chicken and frog **Figure 4-1**. To my knowledge, no overt role for this residue in XPA protein function has been elucidated. The variation that exists at this residue in complementing orthologues shows it is tolerant to methionine, valine and threonine. K141R is a conservative substitution, from a positively charged lysine to a positively charged arginine. Arginine was the only variant present at this position in both complementing and non-complementing orthologues **Figure 4-4.** K141A results in reduced binding to RPA70 and *in vitro* dual incision activity using whole cell extracts (Saijo et al., 2011). Substitution of the positively charged K141 to negatively charged glutamic acid (K141E), results in moderate loss of XPA DNA binding activity and NER repair in human fibroblasts (Camenisch et al., 2006). The data presented here and in previous publications suggests that maintenance of a positive charge at this residue may be required for XPA function, and is at least tolerant to the presence of arginine. L191V has been identified in multiple individuals including some affected by XP. Invitae classified this variant as likely benign using their modification of the ACMG-AMP variant classification criteria termed Sherloc (Nykamp et al., 2017). The remaining submitters evaluated this variant as a VUS using multiple forms of classification criteria. The distinct classifications of ‘likely benign’ and ‘uncertain significance’ lead to its classification as a variant with ‘conflicting interpretations of pathogenicity’ (Richards et al., 2015). Here, it was identified in fish XPA, **Table 4-6** and then shown to be tolerated by the human protein **Figure 4-6**. This is a conservative substitution, with both valine and leucine being non-polar hydrophobic amino acid residues. Furthermore, this residue was well conserved throughout complementing orthologues, with mouse, chicken and frog also containing a valine at this position **Figure 4-1**. This variant was shown not to impact the DNA binding activity of XPA

(Sugitani et al., 2017), but this residue may play a role in XPA binding to XPE (Pradhan et al., 2018). Overall, it is tolerant to a valine at this position. Lastly, V234L has been identified by 3 separate submitters, 1 of which identified it in an XP patient. All 3 submitters classified it as 'likely benign', using separate variant-classification criteria. Here it was identified in chicken and frog XPA **Table 4-6** and then shown to be tolerated by the human protein **Figure 4-6**. This is a conservative substitution, with both valine and leucine being non-polar hydrophobic amino acid residues. Furthermore, a non-polar hydrophobic methionine is present at this position in fish XPA and valine is present in mouse XPA **Figure 4-1**. This variant was previously shown to display WT function in restoring protection against UV in an XPA-deficient cell line (Mellon et al., 2002), agreeing with the results shown here **Figure 4-6**. Interestingly, XPA V234L repairs BPDE-induced DNA damage better than WT XPA (Porter et al., 2005). The authors hypothesised that this may have altered XPAs affinity for PARP1 (Porter et al., 2005). Overall, V234L is tolerated and may even enhance XPA function in some cases. The results presented here show that methionine is also tolerated at this position.

The data presented in this chapter may be eligible for consideration as strong functional data (BS3) by the ACMG/AMP framework (Richards et al., 2015). The functional data shown here in combination with benign predictions from 3 bioinformatic algorithms (BP4) may enable the reclassification of these variants to at least 'likely benign' (Richards et al., 2015).

In addition to the above, sequence alignments of complementing XPA orthologues demonstrated the non-essentiality of reported PTMs in the human XPA protein. The positions of multiple PTMs of XPA showed variation to non-modifiable residues in complementing orthologues, **Table 4-8**. Introduction of non-modifiable residues at these positions (S173, S196 and T239) in the human XPA protein confirmed that these PTMs were not required to restore protection against UV radiation **Figure 4-7**. It was previously shown that S173 and S196 are phosphorylated by ATR in response to UV radiation (Wu et al., 2006b). S196A impairs both

cellular viability and colony formation in response to UV, whereas S173A has no effect (Wu et al., 2006b). The results here differ with previous findings for S196A, but agree with those for S173A (Wu et al., 2006b). ATR recognises an SQ or TQ consensus site for phosphorylation (Matsuoka et al., 2007). This motif was lost for both sites in chicken, frog and fish and introduction of S196A into the human protein had no effect on colony formation in response to UV **Figure 4-7**. The difference between the findings presented here and in previous work is challenging to narrow down. It cannot be attributed to different doses of UV, as similar doses were used here compared to the previous study (Wu et al., 2006b). However, the wavelength of UV radiation used was not specified (Wu et al., 2006b). As DNA shows the greatest absorption of UV at 260 nm (UV-C), UV radiation of different wavelengths induces UV associated DNA lesions at different efficiencies (Rastogi et al., 2010). In my work, I utilised UV-C radiation (254 nm, UV Stratalinker 2400, https://ipmb.sinica.edu.tw/microarray/index.files/Stratalinker%20UV%20Crosslinker.pdf) to irradiate *XPA*-deficient cells in **Figure 4-7**. It was previously shown that phosphorylation of S196 by ATR is involved in the repair of CPDs, but not 6-4PPs, induced by UV-C irradiation (Shell et al., 2009). However, more recent work shows that S196A affected the repair of both UV-C induced 6-4PPs and CPDs to the same extent (Lee et al., 2014b). Despite this, neither study determined the effect S196A has on cellular survival in response to UV radiation. Therefore, consensus is lacking on the importance of S196 phosphorylation in the repair of specific types of UV-C associated DNA lesion and subsequent cell survival. As a result, it is difficult to say that potential differences in the type and frequencies of UV associated DNA lesion underlie the difference in results (Wu et al., 2006b). Overall, the difference between findings presented here and in previous reports is challenging to narrow down. However, the lack of conservation of these consensus sites across complementing orthologues and the direct demonstration that mutation to alanine does not impact survival after UV lends weight to them

being non-essential for XPA function. In addition to S173 and S196, it was shown that introduction of alanine at T239 did not disrupt cellular survival in response to UV **Figure 4-7**. Phosphorylation of this residue was previously reported in a high-throughput screen utilising an anti phospho-tyrosine antibody and therefore lacked any functional assessment. Whilst not always directly functionally tested, non-modifiable residues in complementing orthologues were present at sites of ubiquitination, acetylation and SUMOylation in the human protein **Table 4-8**. SUMOylation of K63 is detectable via mass-spectrometry (Hendriks et al., 2014), but lacks direct functional assessment with no assigned role in XPA protein function. Based on the presence of a non-modifiable arginine at this position in chicken **Figure 4-1** and lack of existing functional data, it can be argued that modification of this residue is not required for XPA function. SUMOylation of K86 is also identified via mass-spectrometry (Lumpkin et al., 2017), but exists as non-modifiable glutamic acid in mouse, glutamine in chicken and fish and a histidine in frog **Figure 4-1**. Concomitantly, SUMOylation of this residue likely serves no role for XPA function. Lastly, both SUMOylation and ubiquitination of K141 has also been identified via mass-spectrometry screens (Akimov et al., 2018, Hendriks et al., 2014), but lack any assigned roles in XPA function. The introduction of a non-modifiable arginine at this position was shown to have no effect on XPA function in response to UV **Figure 4-7**. Overall, modification of this residue is likely not essential for XPAs role in repairing UV-induced DNA damage allowing cellular survival.

4.5 Future experiments based on the work presented in this chapter

Future experiments to be performed on this work could include quantitative analysis of the IP experiments performed in this chapter or quantitative assessment of the protein-protein interactions using methods described in section 4.4.2. This would accurately determine the extent of interaction between vertebrate and invertebrate XPA orthologues and human XPC. This may reveal whether fly XPA has a similar affinity for human XPC to vertebrate

orthologues, seeing as its expression is reduced. Furthermore, work could be done to identify why fly and perhaps yeast XPA are not complementing. One example is to perform the humanisation of Fly XPA by mutating it's E-cluster back to the human sequence. This may result in enabling this orthologue to fully complement the UV protection defect in XPA null human cells. This may reveal additional functional variation in human cells. Despite this, previous work has shown that fly XPA cannot fully complement human cells (Shimamoto et al., 1995) corroborating the work shown here.

Having demonstrated that sequence alignments of complementing orthologues can identify functional variants and non-essential PTM sites in the XPA protein, in the next chapter I will perform this analysis using XRCC4. This will ensure that this is not unique to XPA, permitting the expansion of the method to further proteins of interest.

5 Orthologue complementation of XRCC4

5.1 Introduction

XRCC4 was first identified using a mutant Chinese hamster ovary (CHO) cell line termed XR-1. XR-1 cells are extremely sensitive to IR in G1-phase, but do not display the same level of sensitivity during S-phase (Giaccia et al., 1985). XR-1 cells also display a severe V(D)J recombination defect (Pergola et al., 1993). Somatic cell hybrids of XR-1 cells and human fibroblasts with subsequent segregation analysis mapped the gene responsible for IR sensitivity to chromosome 5 (Giaccia et al., 1990). Subsequent complementation of XR-1 cells with human chromosome 5 ablated the sensitivity of these cells to IR (Giaccia et al., 1990). Following this, human *XRCC4* was cloned and encodes polypeptide of 336 amino acid residues with a molecular weight of 38 kDa and no obvious biochemical activity (Li et al., 1995b). Importantly, complementation of XR-1 cells with human *XRCC4* cDNA restores V(D)J recombination and resistance to IR (Li et al., 1995b). Subsequently, XRCC4 was found to be a nuclear protein, that typically forms a homodimer *in vivo* (Mizuta et al., 1997). Structurally, XRCC4 contains a globular N-terminal head-domain followed by a long C-terminal coiled-coil helical tail and typically exists as a dimer in equilibrium with a tetrameric form of the protein (Junop et al., 2000). Dimerisation of XRCC4 is mediated by the association of two XRCC4 head-domains, and the initial portions of the helical tails (Junop et al., 2000). The tetramer is formed by interaction of the tails of each XRCC4 dimer, forming a four-helix bundle (Junop et al., 2000). Single-particle electron microscopy of full-length XRCC4 reveals that the C-terminal region of XRCC4 forms a dimeric globular-domain, connected to the N-terminus by a coiled-coil (Recuero-Checa et al., 2009).

XRCC4 is fundamental to the NHEJ pathway of DSB repair **Figure 2-5**, evidenced by the extreme sensitivity of *XRCC4*-deficient cells to IR (Giaccia et al., 1990, Li et al., 1995b). Mechanistically, XRCC4 binds to LIG4, with this interaction stimulating ligation of DSB ends by LIG4 five-fold in a cell-free system (Critchlow et al., 1997, Grawunder et al., 1997). Furthermore, XRCC4 is essential for the stability of LIG4 *in vivo*, with *XRCC4*-deficient cells displaying almost undetectable levels of LIG4, which is restored upon complementation with *XRCC4* (Bryans et al., 1999). Additionally, XRCC4 promotes the adenylation of LIG4 *in vitro*, an important PTM in its role in end-joining (Modesti et al., 1999). The interaction between LIG4 and XRCC4 exhibits a 1:2 stoichiometry and is mediated by the linker-domain between the two BRCT-domains of LIG4 and residues 180-213 of XRCC4 (Grawunder et al., 1998, Sibanda et al., 2001, Wu et al., 2009) Tetramerisation of XRCC4 and its interaction with LIG4 are mutually exclusive due to significant overlap of the XRCC4 tetramerisation and LIG4 interaction interfaces (Modesti et al., 2003).

XRCC4 also binds to XLF (Tsai et al., 2007, Gu et al., 2007). XLF, promotes the re-adenylation of LIG4 after an aborted ligation reaction, priming it for its next catalytic activity (Buck et al., 2006, Ahnesorg et al., 2006, Riballo et al., 2009). XRCC4 stimulates this activity *in vitro*, with a preference towards mismatched and non-cohesive ends (Tsai et al., 2007, Gu et al., 2007). More recently, *in vitro* and structural work shows that XRCC4 mediates DNA end-synapsis during NHEJ. Specifically, transition from long-range to short-range complex formation is dependent on DNA-PKcs catalytic activity, XLF and the XRCC4-LIG4 interaction (Graham et al., 2016). More recent work demonstrates that the interaction between XLF and XRCC4 is also required for short-range complex formation (Graham et al., 2018a). This was validated by structural work showing that both long-range and short-range synaptic complexes at DSB ends contain XRCC4 (Chen et al., 2021). Some studies have suggested that XRCC4 dimers are able to interact with XLF dimers to form long superhelical filaments (Hammel et al., 2010, Ropars

et al., 2011). Such filaments were shown to bridge DNA ends in various *in vitro* assays (Andres et al., 2012, Brouwer et al., 2016). Furthermore, immunofluorescent super-resolution imaging of U2OS cells stained with anti-XLF and anti-XRCC4 antibodies reveals elongated nuclear foci, proposed to be XLF-XRCC4 filaments (Reid et al., 2015). However, short-range complex formation only requires a single dimer of XLF, and NHEJ is not dependent on XLF-XRCC4 filament formation *in vitro* (Graham et al., 2018a). Furthermore, structural studies of DNA end-synapsis did not reveal XLF-XRCC4 filament formation (Chen et al., 2021). Additionally, the interaction between XLF and XRCC4 is not required for stimulation of LIG4 activity, or indeed NHEJ, in some cell types (Roy et al., 2015). Structurally, XRCC4 and XLF dimers interact via their head-domains (Deshpande and Wilson, 2007, Andres et al., 2007, Malivert et al., 2010) mediated by residues K63, K65 and K99 of XRCC4 (Andres et al., 2007, Malivert et al., 2010, Ropars et al., 2011, Wu et al., 2011). In addition to XLF and LIG4, XRCC4 either alone or in conjunction with Ku binds to additional factors involved in end-processing and ligation (Davis et al., 2014), specifically, PNKP (Koch et al., 2004), APTX (Clements et al., 2004) and APLF (Iles et al., 2007). CK2 mediated phosphorylation of T233 permits the binding of PNKP to XRCC4 via its FHA-domain (Koch et al., 2004, Bernstein et al., 2005). Additionally, pT233 mediates the binding of both APTX and APLF via their FHA-domains to XRCC4 (Clements et al., 2004, Iles et al., 2007). Importantly, disruption of T233 results in moderately reduced survival in response to IR (Koch et al., 2004).

In addition to CK2-mediated phosphorylation of T233, XRCC4 experiences extensive PTM including additional phosphorylation but also ubiquitination and SUMOylation. Several modifications regulate the interaction of XRCC4 with its binding partners or its function in response to DSB formation. However, some PTMs have no discernible role, or their importance is debated. XRCC4 is extensively phosphorylated on its C-terminus by DNA-PKcs *in vitro* and *in vivo* (Critchlow et al., 1997, Modesti et al., 1999, Leber et al., 1998, Lee et al., 2002, Yu et

al., 2003, Lee et al., 2004). This phosphorylation occurs in response to IR, in a DNA-PKcs-dependent manner (Matsumoto et al., 2000). However, phosphorylation of XRCC4 on S260 and S318 by DNA-PKcs is not essential for survival in response to IR or V(D)J recombination (Yu et al., 2003, Lee et al., 2004), or for repair of DSBs in an *in vitro* end-joining assay (Lee et al., 2004). Furthermore, phosphorylation of serine 193, 259, 302, 313, 325 and 326, and threonine 321 by DNA-PKcs is also not essential for XRCC4 function in response to IR, or in V(D)J recombination (Yu et al., 2003). However, one study has shown that introduction of S260A results in a modest survival defect in response to IR (Moghani et al., 2018). Furthermore, phosphorylation of these sites, in combination with phosphorylation of XLF mediates the stability and DNA bridging capacity of XRCC4/XLF complexes (Normanno et al., 2017). Introduction of phospho-mimicking mutations in both XRCC4 and XLF in tandem, results in faster dissociation of XRCC4-XLF complex from DNA, perturbing DNA bridging and survival in response to radio-mimetic drugs, in a cell specific manner (Normanno et al., 2017). However, no effect was seen on repair of RAG and I-SceI-induced DSBs in 293T cells (Normanno et al., 2017). The authors argue that phosphorylation of XRCC4 and XLF facilitates their dissociation from DNA once repair is complete, but that functional redundancy between XRCC4 and XLF phosphorylation exists in this situation (Normanno et al., 2017). Furthermore, this phosphorylation was not required for survival in response to radio-mimetics in multiple cell lines (Normanno et al., 2017).

XRCC4 is also phosphorylated by CDK1 and PLK1 on S326 (Terasawa et al., 2014). Notably, the phosphorylation of this residue in M-phase prevents anaphase bridge formation (Terasawa et al., 2014). Cells expressing phosphorylation-defective XRCC4 (S326A) show more efficient M-phase DSB repair, accompanied by an increase in anaphase bridge formation (Terasawa et al., 2014). In addition to phosphorylation, XRCC4 is also ubiquitinated. In response to IR, FBXW7 is phosphorylated on S26 by ATM, which promotes its recruitment to DSB sites where

it ubiquitinates XRCC4 on K296 to facilitate association of XRCC4 and the Ku heterodimer (Zhang et al., 2016). Concomitantly, genetic or pharmacological inactivation of FBXW7 results in impaired NHEJ and sensitises cells to IR (Zhang et al., 2016). Notably, phosphorylation of S325 and S326 by DNA-PK promotes binding of XRCC4 to FBXW7 (Zhang et al., 2016). Mutation of these residues to alanine abrogates FBXW7 binding and ubiquitination of XRCC4 (Zhang et al., 2016). Importantly, complementation using a phospho-dead S325A, S326A mutant results in increased sensitivity to IR in a cell survival assay (Zhang et al., 2016), contradicting previous results (Yu et al., 2003). SUMOylation of XRCC4 on K210 occurs both *in vitro* and *in vivo,* and is important for the nuclear localisation of XRCC4 and DSB repair (Yurchenko et al., 2006). A SUMOylation defective XRCC4 mutant (K210R) remains in the cytoplasm, and does not restore protection against IR or V(D)J recombination proficiency (Yurchenko et al., 2006). In contrast, more recent work suggests that SUMOylation of K271, not K210, is important for XRCC4 nuclear localisation and DSB repair after IR (Koike et al., 2022, Koike et al., 2016, Fukuchi et al., 2015). In addition to the PTM sites described above, many more have been identified via high-throughput methodologies such as mass-spectrometry or proteomic screens, but lack *in vivo* functional assessment and any assigned role (Huttlin et al., 2010, Bian et al., 2014, Sharma et al., 2014, Sacco et al., 2016, Hornbeck et al., 2015, Akimov et al., 2018, Boeing et al., 2016, Kim et al., 2011). Overall, XRCC4 experiences extensive PTM, but the importance of the majority of these modifications is either debated, or currently unknown. Therefore, it presents itself as an ideal candidate for assessment with orthologue complementation, which may reveal non-essential PTM sites for XRCC4 protein function.

LOF mutations to *XRCC4* give rise to several phenotypes including radiosensitivity, dwarfism, microcephaly, endocrine dysfunction and encephalocardiomyopathy (Woodbine et al., 2014, Saito et al., 2016). Furthermore, *Xrcc4* knockout is embryonic lethal in mice, accompanied by

defective lymphogenesis and extensive cell death of newly generated postmitotic neuronal cells (Gao et al., 1998). Disease causing mutations in XRCC4 are rare (Saito et al., 2016). However, mutations in XRCC4 causing primordial dwarfism are known (Shaheen et al., 2014) as well as additional phenotypes such as endocrine dysfunction and nervous system abnormalities and radiosensitivity (Bee et al., 2015, Murray et al., 2015, Bruin et al., 2015, Rosin et al., 2015, Guo et al., 2015). Importantly, mutations in XRCC4 have been associated with cancer predisposition including breast and bladder (He et al., 2014, Long et al., 2013, Tseng et al., 2008, Mittal et al., 2012, Zhou et al., 2012a). To date, 40 variants for XRCC4 are present on ClinVar without sole classifications of either benign or pathogenic. This, in combination with the function of extensive numbers of PTMs of this protein remaining unknown or debated, make it a suitable candidate for assessment via orthologue complementation.

The NHEJ pathway is over 3 billion years old and extensively conserved (Demogines et al., 2010, Pears and Lakin, 2014) with orthologues of XRCC4 present in almost all species (Li et al., 1995b, Herrmann et al., 1998, Jeggo, 1998, Block and Lees-Miller, 2005, Koike et al., 2016, Koike et al., 2022). Indications of the orthologue complementation ability of XRCC4 are present in the literature. As mentioned previously, human *XRCC4* can restore V(D)J recombination and protection against IR in XR-1 cells derived from the Chinese hamster (Li et al., 1995b). Recently, complementation of CHO cells with feline *XRCC4* shows that feline XRCC4 could be recruited to DSB sites in CHO cells in a Ku dependent manner (Koike et al., 2022).

Overall, the NHEJ pathway and XRCC4 are widely conserved across evolution with orthologues present in almost all species, ranging from humans to yeast. Furthermore, XRCC4 deficiency presents a clear, assayable phenotype in sensitivity to IR. Additionally, XRCC4 is known to undergo extensive PTM with the importance of the majority these modifications debated, or unknown. Lastly, variants in XRCC4 are without classifications of benign or

pathogenic on ClinVar, and thus require further functional assessment. Collectively, XRCC4 presented itself as a suitable candidate for further validation of the orthologue complementation methodology.

5.2 Aims and objectives

In the previous chapter, I showed that orthologue complementation of XPA could reveal functional amino acid variants and non-essential PTM sites in the human protein. In this chapter, I seek to demonstrate that this is not specific to XPA and can be applied to additional proteins. In this chapter, the complementation ability of vertebrate orthologues of XRCC4 will be assessed via IP for a known binding partner and IR survival assays. Subsequently, sequence alignments of complementing orthologues will be constructed and analysed for their ability to identify functional VUS and non-essential PTM sites in XRCC4. To validate the methodology, PTM sites present in complementing orthologues as non-modifiable residues are changed to non-modifiable residues in the human protein. These will then expressed in human cells and IR survival assays will be performed to confirm that sequence alignments of complementing orthologues can identify non-essential PTM sites in an additional protein. Additionally, the data will be compared to existing publications on variant effects and public variant databases to assess overall concordance.

5.3 Results

5.3.1 Selection of vertebrate XRCC4 orthologues for complementation of human cells

To assess the ability of XRCC4 orthologues to complement human cells, orthologues from distinct taxonomic groups with varying sequence divergence from human were selected. Orthologue sequences for *Homo sapiens* (human) *Mus musculus* (mouse), *Gallus gallus* (chicken), *Xenopus tropicalis* (frog) and *Danio rerio* (fish) XRCC4 were obtained from https://www.ncbi.nlm.nih.gov/ with accession numbers displayed in **Table 3-1**. Multiple accession numbers existed for each orthologue, and thus the entries showing the closest lengths to human were selected for each species. Sequences from each protein of interest were then aligned using the T-Coffee algorithm (Notredame et al., 2000), and the conservation of identical or similar residues was indicated using Multiple Align Show **Figure 5-1**. The percentage identity to human XRCC4 was then calculated using the % identity matrix output option from the T-Coffee alignment **Table 5-1.**

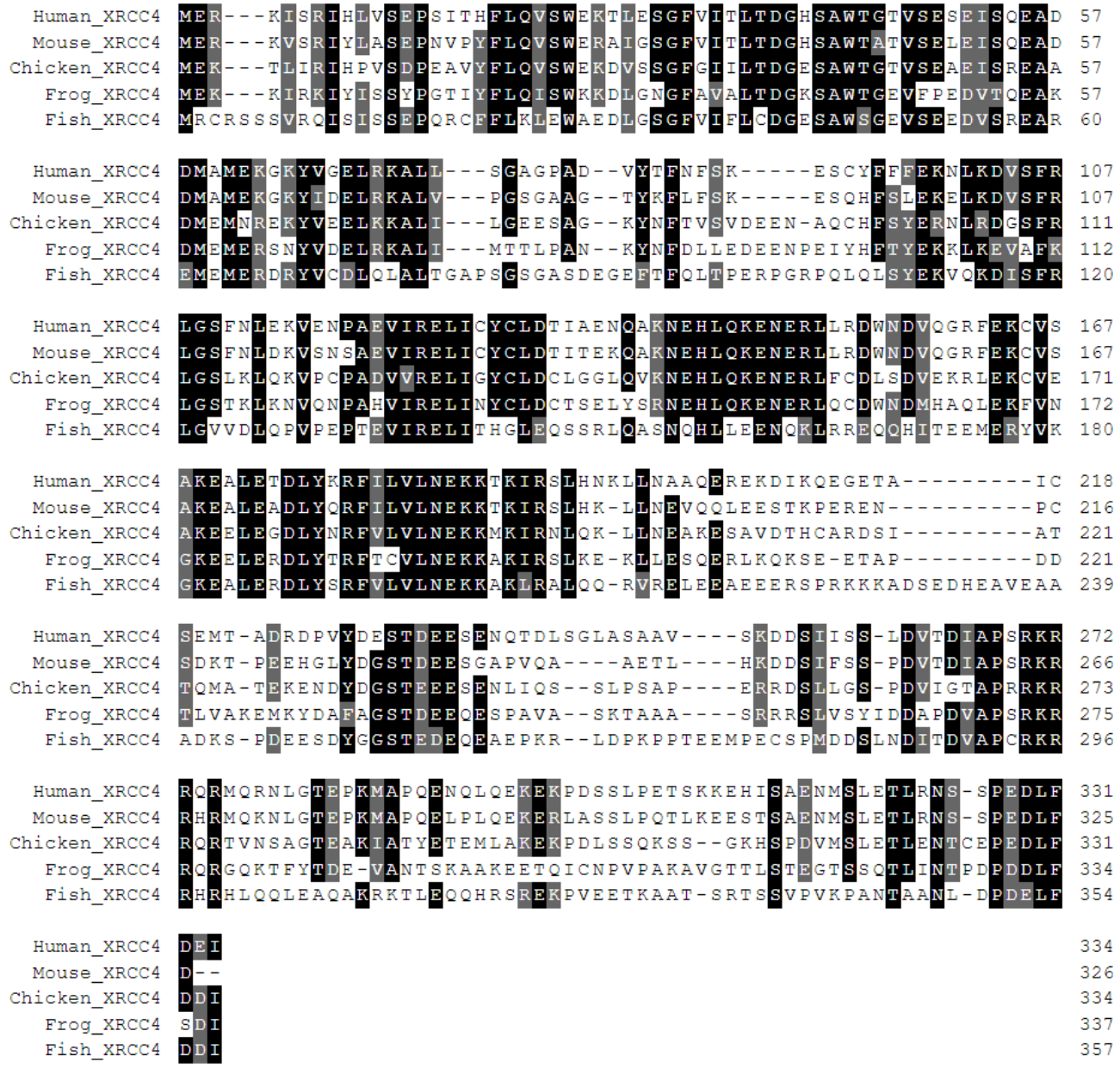


Figure 5-1: Sequence alignment of vertebrate XRCC4 orthologues used in this study with conservation shading

Sequence alignment of vertebrate XRCC4 orthologues used in this study. Sequences were aligned using the T-Coffee algorithm (Notredame et al., 2000) and conservation was scored using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html). Areas of black shading indicate 80% identical conservation of a residue, areas of gray shading indicate 80% conservation to at least an amino acid with similar biochemical properties. Areas of white shading indicate less than 80% conservation. Numbers indicate AA position.

XRCC4 Orthologue	Accession Number	% Identity to Human Orthologue
Human	NP_003392.1	100
Mouse	Q924T3.1	75
Chicken	XP_015136170.1	55
Frog	XP_012826245.1	47
Fish	NP_957080.1	39

Table 5-1: Accession numbers and % identity to human orthologue for vertebrate XRCC4 orthologues used

5.3.2 Vertebrate XRCC4 orthologues can all bind to LIG4

The primary role of XRCC4 is to bind and stabilise LIG4 to ensure that the ligation step of NHEJ can take place (Critchlow et al., 1997, Grawunder et al., 1997, Bryans et al., 1999, Davis et al., 2014). Furthermore, loss of LIG4 results in loss of XRCC4 nuclear import and cell lines from patients with hypomorphic mutations in LIG4 show reduced XRCC4 protein levels (Francis et al., 2014). Therefore, it was determined whether the XRCC4 orthologues selected could bind to LIG4 via immunoprecipitation. The selected XRCC4 orthologues were cloned into pEGFP-C1 and transfected into 293FT cells along with a pEGFP-C1 empty vector control. GFP-XRCC4 constructs were pulled down and their ability to bind to LIG4 was determined via western blotting. All orthologues of XRCC4 could interact with human LIG4 **Figure 5-2**.

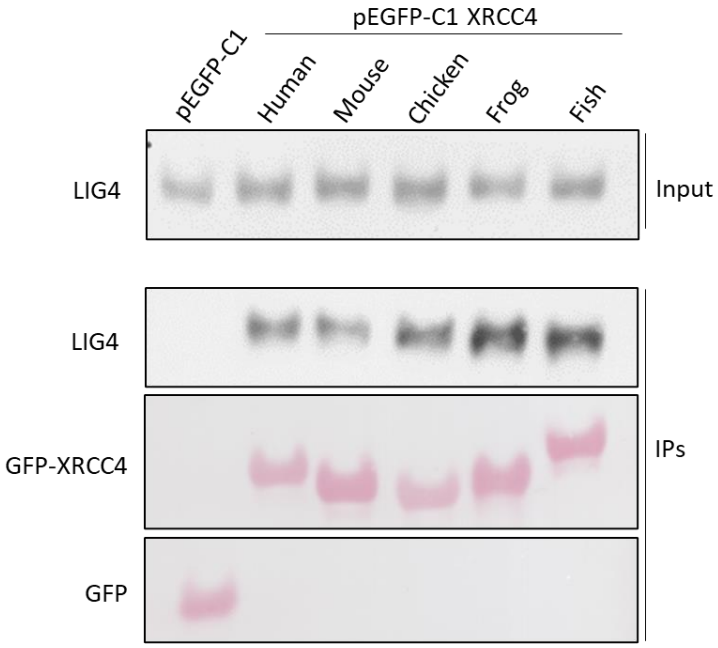
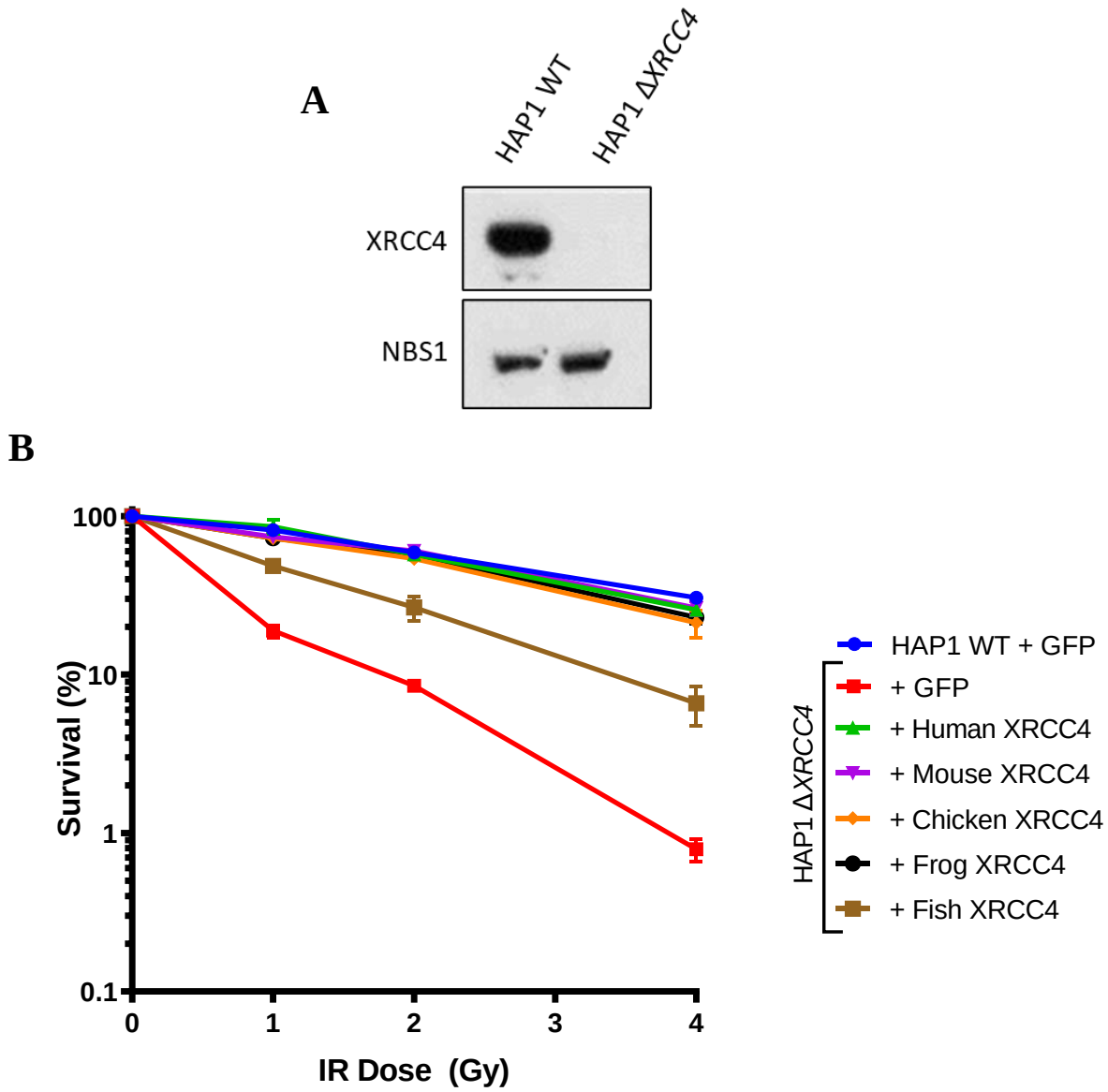


Figure 5-2: Vertebrate XRCC4 orthologues can all bind to LIG4

Western blot depicting GFP-pulldowns from 293FT cells transfected with either GFP only or GFP-tagged human, mouse, chicken, frog and fish XRCC4 orthologues with probing for LIG4. Ponceau staining demonstrates expression of GFP and GFP-tagged XRCC4 orthologues and their respective molecular weights. N=3.

5.3.3 Fish XRCC4 cannot restore protection against IR in human cells

Having demonstrated that the selected XRCC4 orthologues could bind to LIG4 in human cells, it was determined whether these orthologues could restore protection against IR in an XRCC4-deficient cell line, with loss of XRCC4 giving rise to extreme IR sensitivity (Li et al., 1995b). pEGFP-C1 constructs encoding the selected orthologues of XRCC4 were transfected into HAP1 Δ*XRCC4* cells **Figure 5-3**. GFP+ cells were then sorted via FACS **Figure 10-4** the following day, seeded as single cells and subsequently irradiated with different doses of IR. Cells were then left to grow for 7 days and then fixed, stained and counted to determine the number of surviving colonies. Mouse, chicken and frog XRCC4 could restore protection against IR to the same extent as human XRCC4 **Figure 5-3**. However, fish XRCC4 showed only a partial ability to restore protection against IR **Figure 5-3**.


A
HAP1 WT
HAP1 ΔXRCC4
XRCC4
NBS1
B
100
10
1
0.1
Survival (%)
0
1
2
3
4
IR Dose (Gy)
HAP1 WT + GFP
HAP1 ΔXRCC4
+ GFP
+ Human XRCC4
+ Mouse XRCC4
+ Chicken XRCC4
+ Frog XRCC4
+ Fish XRCC4

C

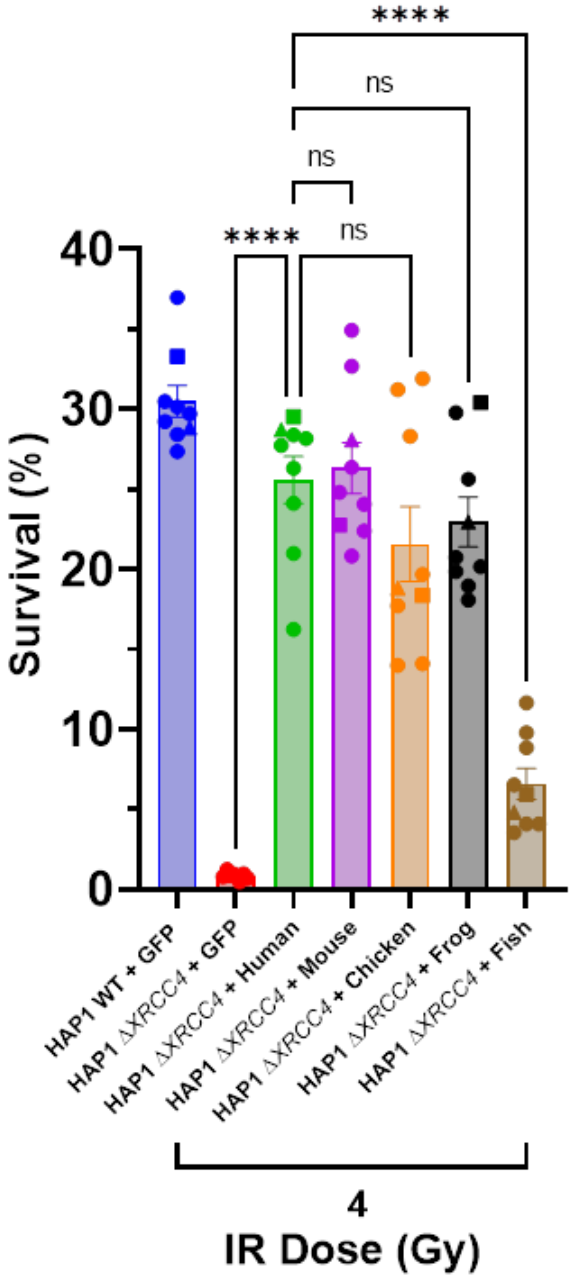


Figure 5-3: Fish XRCC4 cannot restore protection against IR in human cells

A: Western blot showing XRCC4 protein levels in HAP1 WT and HAP1 Δ*XRCC4* null cells with NBS1 used as a loading control. **B:** Survival curves demonstrating the colony survival % compared to untreated cells in response to increasing doses of IR for HAP1 WT and HAP1 Δ*XRCC4* cells complemented with GFP-tagged vertebrate XRCC4 orthologues or GFP only controls. Data is plotted as the mean of N=3. Error bars indicate SEM. **C:** One-way ANOVA was performed on all values of N=3 at 4 Gy IR and included a Dunnett's multiple comparisons test comparing HAP1 Δ*XRCC4* + GFP-WT XRCC4 to HAP1 Δ*XRCC4* + GFP, GFP-mouse XRCC4, GFP-chicken XRCC, GFP-frog XRCC4 and GFP-fish XRCC4. Results from Dunnett's multiple comparisons test are displayed. ns = non-significant difference, **** = p <0.0001.

5.3.4 A sequence alignment of complementing XRCC4 orthologues contains known benign variants but also a known pathogenic variant

Having identified orthologues of XRCC4 which complement human cells, an alignment of these orthologues was created **Figure 5-4**. As performed for XPA, I assessed whether known benign variants in the human protein were present at any of the corresponding positions in the fully complementing orthologues. The data for all missense mutations in XRCC4 catalogued on the gnomAD database was downloaded and then filtered to show gnomAD variants present homozygously in at least one individual. The sequence alignment of complementing orthologues was then searched to determine whether any of these variants were present at the corresponding position in complementing XRCC4 orthologues. Importantly, two such variants, I134T and N137K, were present in the mouse and frog XRCC4 orthologues respectively **Figure 5-4**, **Table 5-2**. This demonstrates that, similar to XPA, sequence alignments of complementing XRCC4 orthologues contain known benign variants which exist in the general population. The sequence alignment was then searched for all benign missense XRCC4 variants on ClinVar. This again identified the I134T and N137K variants as being present in frog and mouse XRCC4 respectively **Figure 5-4**, **Table 5-3**. Lastly, the alignment was searched for all pathogenic missense mutations in XRCC4 on ClinVar. Strikingly, it identified the pathogenic R161Q variant as being present at the corresponding position in frog XRCC4 **Figure 5-4**, **Table 5-4**. However, this variant is encoded across an exon boundary **Table 5-5** and is known to disrupt splicing of XRCC4 pre-mRNA (Rosin et al., 2015). As explained in the previous chapter, the methodology cannot allow for splice site disrupting mutations due to the cDNA basis of the technique.

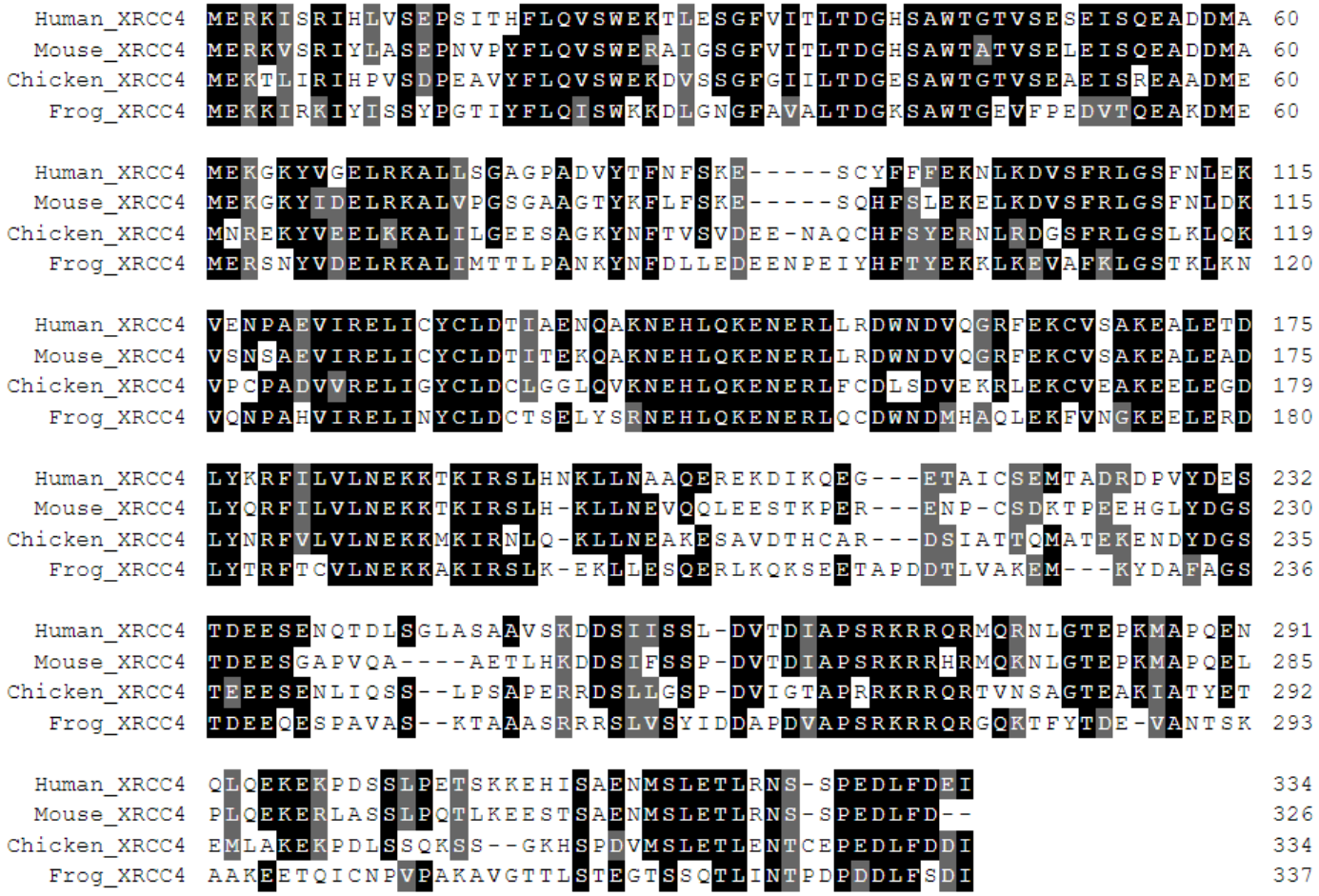


Figure 5-4: Sequence alignment of vertebrate XRCC4 orthologues which complement human cells

Sequence alignment of vertebrate XRCC4 orthologues which could restore protection against IR to WT levels. Sequences were aligned using the T-Coffee algorithm (Notredame et al., 2000) and conservation was scored using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html). Areas of black shading indicate at least 70% identical conservation of a residue, areas of grey shading indicate at least 70% conservation to at least an amino acid with similar biochemical properties. Areas of white shading indicate less than 70% conservation. Numbers indicate AA position.

rsIDs (gnomAD identifier)	Residue position	WT	Variant	Variant present in	Allele frequency	Homozygote count
rs28383138	12	S	C	None	0.25%	8
rs144653114	103	D	N	None	0.10%	3
rs28360135	134	I	T	Frog	2.52%	124
rs56334522	137	N	K	Mouse	0.17%	4
rs1482288279	189	T	I	None	0.0016%	1
rs3734091	247	A	S	None	2.24%	120
rs141304949	336	I	T	None	0.047%	1

Table 5-2: Results of searching the sequence alignment of complementing XRCC4 orthologues for homozygous XRCC4 variants on gnomAD

The sequence alignment of complementing XRCC4 orthologues (human-frog) was searched for homozygous XRCC4 variants present on the gnomAD database (Karczewski et al., 2020). Analysis revealed 2 homozygous variants, I134T and N137K which were present in frog and mouse XRCC4 respectively.

ClinVar Accession Number	Residue Position	WT	Variant	Variant present in
NM_003401.5(XRCC4):c.35C>G (p.Ser12Cys)	12	S	C	None
NM_003401.5(XRCC4):c.307G>A (p.Asp103Asn)	103	D	N	None
NM_003401.5(XRCC4):c.401T>C (p.Ile134Thr)	134	I	T	Frog
NM_003401.5(XRCC4):c.411T>G (p.Asn137Lys)	137	N	K	Mouse
NM_003401.5(XRCC4):c.424G>C (p.Glu142Gln)	142	E	Q	None
NM_003401.5(XRCC4):c.739G>T (p.Ala247Ser)	247	A	S	None
NM_003401.5(XRCC4):c.1001T>C (p.Ile334Thr)	334	I	T	None

Table 5-3: Results of searching the sequence alignment of complementing orthologues XRCC4 for benign XRCC4 variants on ClinVar

The sequence alignment of complementing XRCC4 orthologues (human-frog) was searched for the presence of benign XRCC4 variants from the ClinVar database. This identified the 134T and N137K variants as being present in frog and mouse XRCC4 respectively.

ClinVar Accession Number	Residue Position	WT	Variant	Variant present in
NM_003401.5(XRCC4):c.127T>C (p.Trp43Arg)	43	W	R	None
NM_003401.5(XRCC4):c.246T>G (p.Asp82Glu)	82	D	E	None
NM_003401.5(XRCC4):c.482G>A (p.Arg161Gln)	161	R	Q	Frog

Table 5-4: Results of searching the sequence alignment of complementing XRCC4 orthologues for pathogenic XRCC4 variants on ClinVar

The sequence alignment of complementing XRCC4 orthologues (human-frog) was searched for the presence of pathogenic XRCC4 missense mutations from the ClinVar database. This identified the R161Q variant as being present in frog XRCC4.

Exon number	Amino acids	Amino acids encoded across exon boundaries
Exon 1	1-47	T46, V47
Exon 2	48-105	S105, F106
Exon 3	106-161	G160, R161
Exon 4	162-213	E212, G213
Exon 5	214-249	S248, A249
Exon 6	250-298	E297, K298
Exon 7	298-334	N/A

Table 5-5: Amino acid positions of XRCC4 exons and the amino acids encoded across exon boundaries

A table displaying the amino acid positions of the XRCC4 exons and the amino acids encoded across the exon boundaries. Exon positions were obtained from CCDS. https://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi?REQUEST=CCDS&DATA=CCDS4058.1

5.3.5 A sequence alignment of complementing XRCC4 orthologues does not contain ClinVar XRCC4 VUS

As performed for XPA, the sequence alignment of complementing XRCC4 orthologues was searched to determine whether any variants in XRCC4 from the ClinVar database without sole classifications of benign or pathogenic were present at the corresponding positions in complementing orthologues. However, in contrast to XPA, none of the 40 XRCC4 variants which fall into this category were present at the corresponding positions in complementing orthologues.

5.3.6 A Sequence alignment of complementing XRCC4 orthologues reveals multiple non-essential PTM sites

As no ClinVar variants were identified for further assessment, the focus switched to PTMs of XRCC4. As performed for XPA, all PTMs reported for XRCC4 were downloaded from the PhosphoSitePlus database (Hornbeck et al., 2015). The sequence alignment of complementing orthologues was then searched for variation to non-modifiable residues at positions of PTMs in the human XRCC4 protein. This revealed several PTM sites which may not be required for XRCC4 function in human cells **Table 5-6**. Specifically, phosphorylation of S6, T17, S219, T222, Y229, T241, S259, T264 and S326, ubiquitination of K65. K146, K178 and K296 and SUMOylation of K210. As explained previously, phosphorylation of serine and threonine residues by the three DDR kinases ATM, ATR and DNA-PKcs is the primary way in which proteins of the DNA damage response are regulated (Blackford and Jackson, 2017). DNA-PKcs is the primary kinase for NHEJ (Blackford and Jackson, 2017). Furthermore, phosphorylation sites important for DDR protein function are often highly conserved (Blackford et al., 2015, Leimbacher et al., 2019). Therefore, serine and threonine phosphorylation sites in XRCC4 which showed variation to non-phosphorylatable residues were analysed. Phosphorylation of S259 and S326, targets of DNA-PKcs, are not required for XRCC4 function in response to IR (Yu et al., 2003) and were subsequently removed from the

analysis presented here. This left S6, T17, S219, T222, T241 and T264 for direct functional assessment.

Having predicted that phosphorylation of S6, T17, S219, T222, T241 and T264 is not required for XRCC4 function in response to IR, this was validated in a cellular model. S6A, T17A, S219A, T222A, T241A and T264A were all introduced into a pEGFP-C1 construct encoding human XRCC4 via site-directed mutagenesis to create a '6A' mutant **Figure 5-5**. The '6A' XRCC4 pEGFP-C1 construct was then transfected into HAP1 Δ*XRCC4* cells. GFP+ cells were then sorted via FACS **Figure 10-5** the following day, seeded as single cells and subsequently irradiated with different doses of IR. Cells were then left to grow for 7 days and then fixed, stained and counted to determine the number of surviving colonies. Mutation of all 6 residues to alanine did not result in a survival defect in response to IR upon complementation of HAP1 Δ*XRCC4* cells **Figure 5-5**. Collectively, the findings have demonstrated that sequence alignments of complementing orthologues can identify non-essential PTM sites in XRCC4.

Position	Residue	Modification	Non-modifiable residue present in	Reported function
6	S	Phosphorylation	Chicken (I), Frog (R)	None
17	T	Phosphorylation	Mouse (P), Chicken (V), Frog (I)	None
65	K	Ubiquitination	Frog (N)	None
102	K	Ubiquitination	Chicken (R)	None
178	K	Ubiquitination	Mouse (Q), Chicken (N)	Ubiquitinated in response to UV (Boeing et al., 2016).
210	K	SUMOylation	Chicken (H), Frog (S)	Proposed to be important for the nuclear localisation of XRCC4 (Yurchenko et al., 2006) but contradicted by more recent findings (Koike et al., 2022, Koike et al., 2016, Fukuchi et al., 2015)
222	T	Phosphorylation	Chicken (A), Frog (A)	None
229	Y	Phosphorylation	Frog (F)	None
241	T	Phosphorylation	Mouse (V), Chicken (I), Frog (A)	None
259	S	Phosphorylation	Chicken (G)	Proposed to be non-essential for cellular survival in response to IR (Yu et al., 2003)
264	T	Phosphorylation	Chicken (I), Frog (P)	None
296	K	Ubiquitination	Frog (E)	Promotes interaction of XRCC4 and Ku heterodimer at DSB sites (Zhang et al., 2016).
326	S	Phosphorylation	Chicken (E) Frog (D)	Proposed to be non-essential for survival after IR (Yu et al., 2003). Proposed to mediate binding of XRCC4 to FBXWY facilitating interaction of XRCC4 and Ku (Zhang et al., 2016)

Table 5-6: Results from searching the sequence alignment of complementing XRCC4 orthologues for variation at PTM sites in XRCC4 reported on the PhosphoSitePlus database

The sequence alignment was searched for all existing PTMs catalogued for XRCC4 on PhosphoSitePlus to identify whether non-modifiable residues are present in complementing orthologues at these positions. In the table, position, residue in human, associated PTM, residue in complementing orthologue and any associated function are provided.

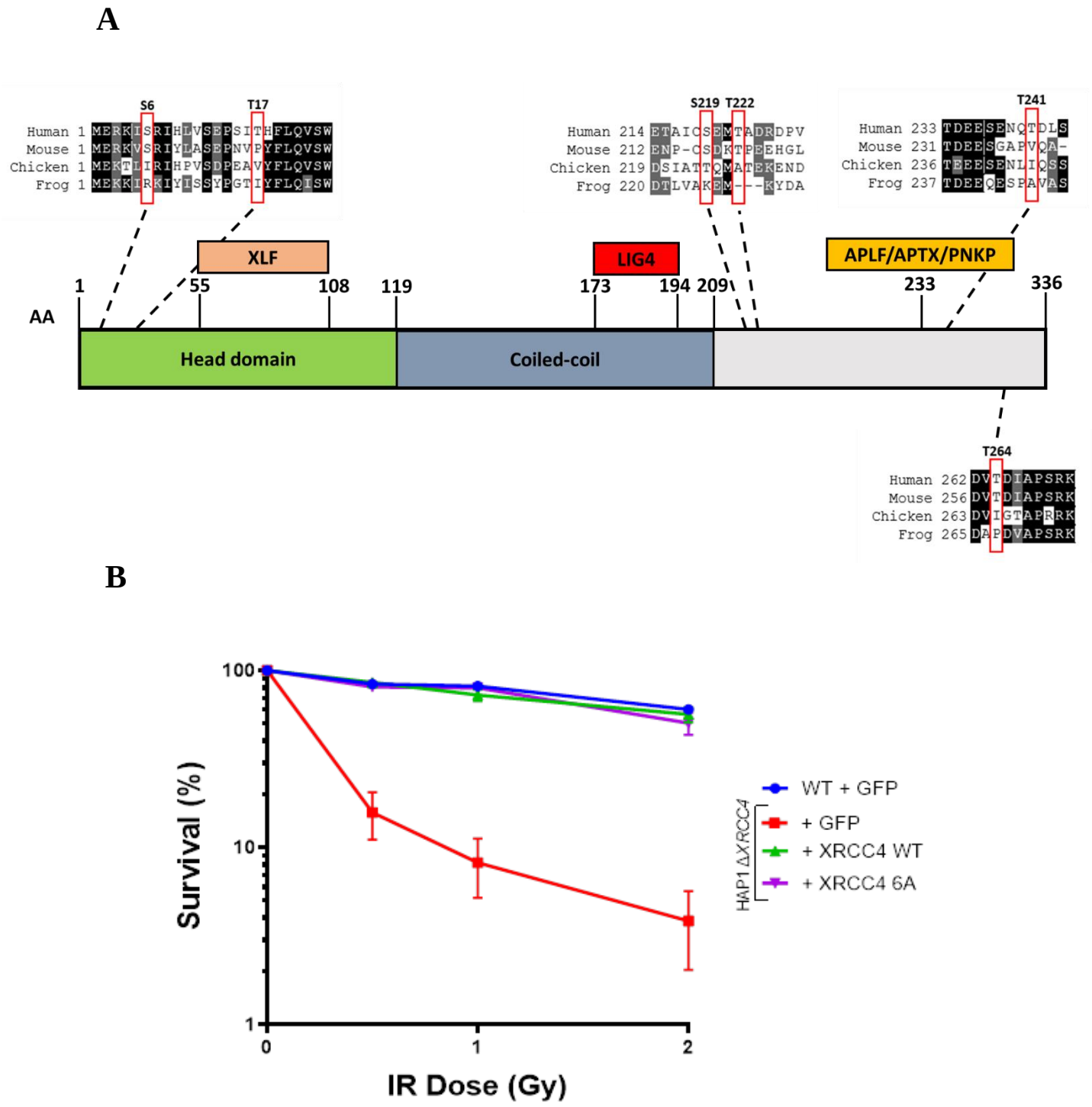


Figure 5-5: Sequence alignments of complementing XRCC4 orthologues reveal non-essential PTM sites in XRCC4

A: Schematic showing positions of reported PTM sites in XRCC4 where alanine residues were introduced with sequence alignments, amino acid residues and binding domains displayed**. B:** Survival curves demonstrating the colony survival % compared to untreated cells in response to increasing doses of IR, for HAP1 WT and HAP1 Δ*XRCC4* cells complemented with GFP-tagged human XRCC4 or GFP-tagged human XRCC4 '6A' or GFP only controls. Data is plotted as the mean of N=3. Error bars indicate SEM.

5.3.7 Comparison to bioinformatic predictions of variant effects

As performed for XPA, the variants tested in this chapter were analysed using the Align_GVGD, SIFT and PolyPhen-2 algorithms **Table 5-7.** The majority of variants were predicted to be benign by all 3 algorithms, showing high concordance with the results shown above **Table 5-7.** However, S6A was predicted to be 'possibly damaging' by PolyPhen-2 **Table 5-7.** This is a clear example of the discordance between *in silico* predictions and *in vivo* assessments (Peterson et al., 2013).

Variant	**Align_GVGD**	**SIFT Score**	**PolyPhen-2**
S6A	Class C0 (tolerated)	Tolerated (0.83)	Possibly damaging (0.83)
T17A	Class C0 (tolerated)	Tolerated (0.47)	Benign (0.00)
S219A	Class C0 (tolerated)	Tolerated (0.82)	Benign (0.16)
T222A	Class C0 (tolerated)	Tolerated (0.31)	Benign (0.03)
T241A	Class C0 (tolerated)	Tolerated (0.68)	Benign (0.00)
T264A	Class C0 (tolerated)	Tolerated (0.79)	Benign (0.001)

Table 5-7: Computational predictions for XRCC4 variants tested

Computation predictions were obtained using the Align-GVGD, SIFT and PolyPhen-2 algorithms for the variants tested in this chapter. All scores are listed and their associated classifications provided.

5.4 Discussion

5.4.1 A summary of the work presented in this chapter

In this chapter, the complementation ability of multiple orthologues of XRCC4 with increasing evolutionary distance to human was determined via LIG4 IP and IR survival assays. XRCC4 orthologues from mouse, chicken, frog and fish can bind to LIG4 **Figure 5-2**. In addition to this, XRCC4 orthologues from mouse, chicken and frog can restore protection against IR **Figure 5-3**. In contrast, whilst fish XRCC4 can bind to LIG4 **Figure 5-2**, it only partially restores protection against IR **Figure 5-3**. Similar to XPA, variants known to be functional in the human XRCC4 protein were found at the corresponding position in complementing XRCC4 orthologues **Figure 5-4**, **Table 5-2**, **Table 5-3**. However, one pathogenic ClinVar variant (R161Q) was detected at the corresponding position in the complementing frog XRCC4 orthologue **Figure 5-4**, **Table 5-4**. Importantly, this variant is known to disrupt pre-mRNA splicing (Rosin et al., 2015), something which this methodology cannot assess due to the use of cDNA, highlighting that all amino acids encoded across exon-boundaries (**Table 5-5**) must be excluded from variant analysis. In contrast to XPA, the sequence alignment of complementing XRCC4 orthologues did not contain variants without sole classifications of either benign or pathogenic on ClinVar. However, the sequence alignment of complementing orthologues reveals that non-modifiable residues were present at sites of multiple reported PTMs of the human protein **Figure 5-4**, **Table 5-6**. Importantly, mutation of these sites in the human protein to non-modifiable residues did not perturb XRCC4 protein function in response to IR **Figure 5-5**. Finally, the variants demonstrated to be functional in this chapter were also predicted to be tolerated by bioinformatic prediction algorithms in almost all cases **Table 5-7.** Overall, a sequence alignment of complementing XRCC4 orthologues provides a reference for functional amino acid variants in the XRCC4 protein. Importantly, this demonstrates that the

orthologue complementation methodology can be expanded to an additional protein, and is not exclusive to XPA.

5.4.2 A discussion of the work presented in this chapter

Prior to this study, some examples of the orthologue complementation ability of XRCC4 were reported. Human XRCC4 cDNA can complement the V(D)J recombination defect and IR sensitivity of the CHO derived XR-1 cell line (Li et al., 1995b). In addition to this, feline XRCC4 is recruited to DSB sites in a Ku dependent manner in CHO cells (Koike et al., 2022). In addition to cross-species complementation, the frequent use of the mouse and xenopus model organisms to study NHEJ and XRCC4 function (Graham et al., 2016, Graham et al., 2018a, Roch et al., 2021, Balmus et al., 2016) provides further evidence of the cross-species conservation of NHEJ protein function. To my knowledge, no existing studies have performed complementation of human cells *XRCC4* orthologues. However, the evidence presented here does fit with the previously observed cross-species complementation of *XRCC4*, and the frequent use of model organisms to study its function. Here the threshold of complementation has been set at the frog XRCC4 orthologue, with 47% similarity to human **Table 5-1.** As fish XRCC4 shows 39% similarity to the human protein, it remains possible that orthologues with > 39% similarity but ≤ 47% could complement human cells.

The mechanism behind partial complementation of human cells with fish XRCC4 is difficult to narrow down. Fish XRCC4 could still bind to LIG4 **Figure 5-2**. However, no quantitative analysis of the interaction was performed, thus it cannot be said if all orthologues bind to human XRCC4 with similar affinities. Methods described in section 4.4.2 could be used to better understand the interactions and overcome this limitation of the work. It remains possible that whilst an interaction is clearly present, fish XRCC4 may have a reduced affinity for human LIG4, resulting in reduced complementation ability. Fish XRCC4 also, partially restores

protection against IR **Figure 5-3**, demonstrating at least some functionality in human cells. Loss of PNKP or APTX binding via introduction of an alanine at T233 results in partial radiosensitivity (Koch et al., 2004, Clements et al., 2004). However, T233 and the consensus S/T-X-X-D/E motif required for CK2 phosphorylation (St-Denis et al., 2015) is present in fish XRCC4, suggesting that this does not underlie the partial radiosensitivity observed. Loss of XLF also causes mild radiosensitivity, due to functional redundancy with PAXX (Li et al., 2008, Balmus et al., 2016). The interaction between XLF and XRCC4 has been assigned to K63, K65 and K99 of XRCC4 (Andres et al., 2007, Malivert et al., 2010, Ropars et al., 2011, Wu et al., 2011). K63 and K65 are both arginine and K99 is a lysine at the corresponding positions in fish **Figure 5-1** Therefore, the same amino acid or one with highly similar biochemical properties is present at the corresponding positions of residues required for XLF binding. It must be noted that the importance of the interaction between XRCC4 and XLF in NHEJ is dependent on the model organism and/or cell line used, (Roy et al., 2015, Graham et al., 2018a) so it would be difficult to assign lack of XLF binding as the reason behind partial complementation. In addition to the above, the reported NLS for XRCC4, AA 270-275 (Li et al., 1995b), is conserved in fish **Figure 5-1** suggesting that this orthologue would localise to the nucleus of human cells to carry out NHEJ. XRCC4 is also known to contain 2 SUMO-interacting motifs (SIM's) of which SIM1 (AA 33-36) is thought to facilitate the interaction between SUMOylated DNA-PK complex members and XRCC4 (Gonzalez-Prieto et al., 2021). Disruption to this SIM leads to a 2-fold increase in sensitivity to bleocin (Gonzalez-Prieto et al., 2021). However, fish XRCC4 shows greater similarity of this motif to human than chicken and frog, with 3 out 4 residues identical to human and thus is likely not responsible for the partial LOF **Figure 5-1**. The pathogenic variant W43R causes mild radiosensitivity in human cell lines (Murray et al., 2015). Importantly, this variant and all other pathogenic variants in XRCC4 on ClinVar were not present at the corresponding position in the fish orthologue, and

therefore could not be assigned to the partial complementation observed **Table 5-4**. It remains possible that loss of one of the many reported PTM sites in XRCC4 may disrupt fish XRCC4 function in human cells. Three reported PTMs were only present as non-modifiable residues in fish XRCC4, specifically, ubiquitination of K146 (present as E) and K164 (present as R) as well as phosphorylation of S256 (present as P) **Figure 5-1**. All these PTMs have only been reported in high-throughput screens and still have no assigned functional role for XRCC4 (Kim et al., 2011, Akimov et al., 2018, Mertins et al., 2013). Despite this, it is possible that loss of one or several of these PTMs in the fish XRCC4 orthologue underlies the partial complementation observed due to perturbation of a currently unknown function and/or interaction. Notably, the importance of the majority of PTMs of XRCC4 is still unclear (see section 5.1). The assessment of a selection of these sites via orthologue complementation provided the main findings of this chapter.

Here, it was shown that phosphorylation of S6, T17, S219, T222, T241 and T264 of XRCC4 is not required for protection against IR **Figure 5-5**. All these PTMs were identified via high-throughput screening methodologies and thus lack assigned roles in XRCC4 protein function (Huttlin et al., 2010; Bian et al., 2014). This highlighted that the methodology can be a useful tool in the large-scale assessment of PTMs with no previously reported function. In addition to these sites, multiple other PTMs were identified as non-modifiable residues in complementing orthologues but were not directly analysed.

Phosphorylation of Y229, S259 and S326 has been reported previously but all were present as non-modifiable residues in complementing orthologues. Phosphorylation of Y229 was identified in high-throughput screens (Bian et al., 2014; Yu et al., 2011). Here it is present as a non-phosphorylatable phenylalanine at the corresponding position in frog XRCC4, showing it is not required for cellular survival in response to IR **Figure 5-3**, **Table 5-6**. Furthermore, S259 was originally identified via mass-spectrometry and shown not to be required for XRCC4

protein function in response to IR (Yu et al., 2003). S259 is present as a non-phosphorylatable glycine in chicken XRCC4, meaning phosphorylation of this residue is not required for cellular survival in response to IR, consistent with previous results (Yu et al., 2003) **Figure 5-4**, **Table 5-6**. Lastly, phosphorylation of S326 has been reported to play multiple roles in XRCC4 protein function, but its importance is heavily debated. Here, it is present as a non-phosphorylatable glutamic acid in chicken and aspartic acid in frog **Figure 5-4**, demonstrating it is not required for cellular survival in response to IR **Figure 5-3**, consistent with previous findings (Yu et al., 2003). However, it was previously shown that phosphorylation of S326 by CDK1 and PLK contributes to the inhibition of DSB repair in M-phase (Terasawa et al., 2014). Importantly, use of a XRCC4 S326A mutant increased the efficiency of M-phase DSB repair, resulting in increased anaphase bridge formation after etoposide treatment in mitosis (Terasawa et al., 2014). Despite this, XRCC4 S326A can still localise to mitotic chromosomes and bind to LIG4 and XLF (Terasawa et al., 2014). Therefore, the role of this phosphorylation in the mechanism of XRCC4 inactivation in mitosis remains unclear (Terasawa et al., 2014). Importantly, the authors did not assess the overall cellular survival of cells expressing a S326A mutant, preventing direct comparison to the results shown here. Furthermore, they utilised nocodazole to enrich the number of mitotic cells on which to induce DSB formation via etoposide. In the assays performed here, an asynchronous population of cells was irradiated. Therefore, it remains possible that cells expressing chicken and frog XRCC4 experience elevated levels of anaphase bridge formation in mitosis, but this either does not significantly affect overall survival, or the population of mitotic cells was too small at the time of irradiation to produce a notable survival defect. It must be noted that mitotic phosphorylation of 53BP1 and RNF8 by CDK1 and PLK1 also suppresses DSB repair in M-phase (Orthwein et al., 2014). Therefore, the lack of survival defect observed with chicken and frog XRCC4 might be due to this being the primary mechanism of M-phase DSB repair suppression, or that this mechanism acted

redundantly with CDK1 and PLK1-mediated phosphorylation of XRCC4 S326. In addition to the above, phosphorylation of S326 by DNA-PKcs has also been shown to mediate the binding of XRCC4 to FBXW7 (Zhang et al., 2016), which serves as the substrate recognition component of the SCF E3 ubiquitin ligase (Lan and Sun, 2021). Mutation of both this and S325 to alanine abrogates FBXW7 binding and subsequent ubiquitination of XRCC4 on K296, reducing association with Ku70/80 (Zhang et al., 2016). Expression of either a K296R mutant or a S325A + S326A mutant mildly increases sensitivity to IR in a cell survival assay, contradicting previous results (Yu et al., 2003) and the results presented here regarding the requirement of S326 phosphorylation. Notably, S325 is conserved in complementing orthologues to at least a threonine **Figure 5-4**, which DNA-PKcs is capable of phosphorylating (Blackford and Jackson, 2017). Therefore, it is possible that functional redundancy exists between these two sites, as they have never been assessed individually (Zhang et al., 2016). In addition to this, K296 is present as a glutamic acid in frog **Figure 5-1**, which cannot be ubiquitinated. Therefore, even if FBXW7 bound to the corresponding S325 in this orthologue, there is no corresponding lysine to be ubiquitinated. Overall, this suggests that ubiquitination of this residue is not essential for complete survival in response to IR in human cells, contradicting previous results (Zhang et al., 2016).

In addition to ubiquitination of K296, the ubiquitination of K65, K102 and K178 has also been reported (Akimov et al., 2018; Kim et al., 2011; Boeing et al., 2016; Mertins et al., 2013; Udeshi et al., 2013). Here, all were present as non-modifiable residues in at least one complementing orthologue **Figure 5-4**, **Table 5-6** demonstrating that they are not required for survival in response to IR **Figure 5-3**. As these modifications were all identified via high-throughput screens and thus have no assigned role, further discussion on their importance cannot be given at this time.

Lastly, XRCC4 is SUMOylated on K210, reported to facilitate its nuclear localisation and cellular survival in response to IR (Yurchenko et al., 2006). However, residues which cannot be SUMOylated are present at the corresponding positions in chicken (H) and frog (S) respectively **Figure 5-4**, **Table 5-6**, contradicting these results. In addition to the work presented here, the importance of this modification has been called into question by other findings. In direct contradiction, mutation of this SUMOylation site to an arginine has no effect on nuclear import or survival after IR (Fukuchi et al., 2015). However, mutation of K271 to arginine significantly disrupts nuclear import of XRCC4 and cellular survival after IR (Fukuchi et al., 2015). Importantly, both canine and feline XRCC4 lack the presence of a lysine at this position and both canine and feline XRCC4 can localise to the nucleus in response to IR (Koike et al., 2016, Koike et al., 2022). Notably, K271 is conserved in all complementing orthologues used here **Figure 5-4**. Overall, the findings presented here agree with the majority of work showing that SUMOylation of K210 is not required for cellular survival in response to IR.

5.5 Future work based on the work presented in this chapter

One future experiment to be performed on this work would be quantification of the interactions between XRCC4 orthologues and LIG4 using experiments described in section 4.4.2 to accurately determine the interactions between XRCC4 orthologues and human LIG4. As described in the discussion, it is challenging to propose mechanistic studies to identify exactly what is behind the loss of complementation with fish XRCC4. Further experiments that may reveal the disrupted mechanism responsible for loss of complementation include assessing the extent of nuclear localization of fish XRCC4 via immunofluorescence. Additionally, a human-fish hybrid XRCC4 orthologue could be generated fusing different domains of each orthologue to try and determine the region in fish XRCC4 responsible for loss of complementation. This may pave a way to understanding the mechanism behind loss of complementation.

In conclusion, I have shown that orthologue complementation is possible with XRCC4, showing that the methodology is not exclusive to XPA. Despite failing to assist in demonstrating the functionality of variants without benign or pathogenic classifications on ClinVar, it has offered further insight into the requirement of multiple PTMs in XRCC4. In the next chapter, the methodology will be applied to the MLH1 protein of the MMR pathway. This protein holds significant clinical importance, and therefore orthologue complementation should enable the classification of multiple VUS in this protein.

6 Orthologue complementation of MLH1

6.1 Introduction

mutL was first identified in prokaryotes, whereby disruptive mutations to *mutL* led to a 50-100 fold increase in mutation rate in *E.coli* (Rydberg, 1978, Glickman and Radman, 1980). Subsequently, mutL was shown to be essential for repair of mismatched bases *in vitro* (Lu et al., 1983). mutL is recruited to mutS complexes bound to mismatched bases in an ATP-dependent fashion (Grilley et al., 1989). The endonuclease activity of mutH is required for the excision step (Welsh et al., 1987), dependent on the mutS-based recruitment of mutL to the sites of mismatched bases (Au et al., 1992). Following the identification of *mutL* in prokaryotes, a eukaryotic homologue was identified. It was already established that MSI was a hallmark of *E.coli* MMR mutants (Levinson and Gutman, 1987). The finding that tumour cells also displayed this phenotype suggested they may also be defective in MMR (Ionov et al., 1993, Aaltonen et al., 1993). Subsequently, deficiency of MutLα - a heterodimer of the mutL homologues MLH1 and PMS2 - was shown to be the cause of MSI in a patient-derived colorectal cancer cell line (Li and Modrich, 1995). Prokaryotic mutL possesses ATPase activity, required for MMR (Ban and Yang, 1998, Spampinato and Modrich, 2000). The motif responsible for this activity is conserved in both human MLH1 and PMS2, with MLH1 showing weak ATPase activity (Räschle et al., 2002, Wu et al., 2015). The ATPase domain of MLH1 is contained within the N-terminus of the protein and mutation of this region reduces MMR activity (Spampinato and Modrich, 2000, Tomer et al., 2002, Räschle et al., 2002, Kadyrov et al., 2006, Wu et al., 2015). However, the ATPase activity of MLH1 is not essential for interaction with PMS2 and MutSα (Räschle et al., 2002).

As explained previously, mutH is the endonuclease responsible for MMR in prokaryotes (Welsh et al., 1987). However, no mutH homologue was identified in eukaryotes. In a

breakthrough paper, MutLα was shown to contain a latent endonuclease activity, dependent on interaction with PCNA and the presence of a mismatch, a preexisting strand break, MutSα, ATP and RFC (Kadyrov et al., 2006). A highly conserved $DQHA(X)_2E(X)_4E$ motif in the C-terminus of PMS2 provides the nuclease activity of MutLα (Kadyrov et al., 2006, Kadyrov et al., 2007, Oers et al., 2010). Substitutions to amino acids within this motif result in loss of MutLα endonuclease activity and abolish MMR in human, yeast and mouse (Kadyrov et al., 2006, Kadyrov et al., 2007, Oers et al., 2010). Crucially, the interaction between MLH1 and PMS2 is essential for stability of PMS2 (Trojan et al., 2002). Structurally, the interaction of MLH1 and PMS2 is mediated by the C-terminal regions of each protein (Pang et al., 1997, Guerrette et al., 1999, Mohd et al., 2006, Guarné et al., 2004). The interaction was initially assigned to AA 506 – 756 in MLH1 (Guerrette et al., 1999, Kondo et al., 2001). Further analysis revealed that an MLH1 construct lacking the last 8 amino acid residues of this region can still bind to PMS2 *in vitro* (Mohd et al., 2006). However, this MLH1 construct does not restore PMS2 stability or mediate cell death in response to 6-TG, but still permits nuclear import of the MLH1-PMS2 heterodimer (Wu et al., 2003, Brieger et al., 2005, Mohd et al., 2006). Importantly, the highly conserved extreme C-terminus of Mlh1 forms part of the endonuclease site of Pms1 and Mlh3 (Gueneau et al., 2013, Dai et al., 2021), perhaps explaining the lack of 6-TG sensitivity observed previously (Mohd et al., 2006).

To enable endonuclease activity, MutL family complexes display multiple characteristics which facilitate their interaction with PCNA. Mispair bound MutS family members undergo conformational changes and form ATP hydrolysis-independent sliding clamps (Gradia et al., 1997, Gradia et al., 1999, Mazur et al., 2006, Heinen et al., 2011). As these conformational changes permit MutLα binding (Mendillo et al., 2009), MutSα-MutLα form sliding clamp complexes and diffuse along the DNA eventually coming into contact with PCNA (Hombauer et al., 2011, Liu et al., 2016, Srivatsan et al., 2014, Lee et al., 2014a). Structurally, the N-

terminus of MLH1 mediates the interaction between MutLα and MutSα (Plotz et al., 2006, Plotz et al., 2003). Mutations to the connector domain of Msh2 in yeast disrupt the interaction with Mlh1-Pms1 (Mendillo et al., 2009). The structure of the MutS/MutL complex shows that MutS complexes undergo conformational changes to form the sliding clamp and promote MutL complex binding (Groothuizen et al., 2015, Fernandez-Leiro et al., 2021). This interaction is mediated by the N-terminus of MutL with the connector domain, and ATPase and core domains of the two MutS subunits respectively (Groothuizen et al., 2015, Fernandez-Leiro et al., 2021). Specifically, Mlh1 is recruited to Msh2-Msh6 by interaction with the connector domain of Msh2 and the ATPase/core domains of Msh6 (DuPrie et al., 2022).

MutL has also been proposed to trap MutS at the sites of mismatches. MutS complex sliding clamps are capable of re-binding to the mismatch after previous recognition in an ATP hydrolysis dependent manner (Hao et al., 2020). MutL can trap MutS onto DNA (Hao et al., 2020), contrasting to the freely diffusing MutS-MutL complex previously reported (Gorman et al., 2007). Collectively, mismatch re-recognition by MutS in combination with trapping by MutL means that MutS sliding clamps are constrained near the mispair, ensuring MMR occurs in close proximity (Hao et al., 2020). Given that PCNA is mobile, the endonuclease activity of MutL complexes could be activated by PCNA migration rather than MutS/MutL sliding clamp migration (Hao et al., 2020). It has also been argued that MutS proteins can trigger polymerisation of MutL complexes allowing contact with PCNA (Hombauer et al., 2011, Modrich, 1987). Concomitantly, *in vivo* fluorescence studies in yeast and *E. coli* demonstrate that MMR foci contain more MutL complexes than MutS (Hombauer et al., 2011, Elez et al., 2012). Despite this, Msh2-Msh6 mutants capable of recruiting Mlh1-Pms1 but unable to form sliding clamps, cannot support MMR *in vivo* (Graham et al., 2018b).

In addition to its multifaceted role in post-replicative MMR, MLH1 in complex with MLH3 (MutLγ) plays a fundamental role in crossover formation and double Holliday junction

resolution during meiotic recombination (Manhart and Alani, 2016). Loss of either *Mlh1* or *Mlh3* results in infertility in mice (Baker et al., 1996, Lipkin et al., 2002). Mutations to the ATPase domain of MLH1 result in defective crossing-over (Argueso et al., 2003). Furthermore, mutations to the MLH3 $DQHA(X)_2E(X)_4E$ endonuclease domain result in meiotic crossing-over defects similar to *Mlh3* loss (Nishant et al., 2008). Recombinant MutLγ heterodimers from both yeast and human show endonuclease activity and high affinity and preference for double-Holliday junctions (Ranjha et al., 2014). Furthermore, the human MLH1-MLH3 heterodimer forms a complex with MSH4-MSH5, EXO1, RFC and PCNA which activates the endonuclease activity of MLH1-MLH3 and preferentially cleaves plasmid DNA containing a double Holliday junction (Cannavo et al., 2020, Kulkarni et al., 2020). A recent paper identified seven SNPs amongst human *MLH1* and *MLH3*, which result in female subfertility and male infertility in mice (Singh et al., 2021).

Lastly, MLH1 also forms a complex with PMS1 to form the MutLβ complex (Räschle et al., 1999). However, PMS1 lacks the endonuclease domain present in both PMS2 and MLH3 and was not shown to participate in MMR *in vitro* (Räschle et al., 1999). However, *Pms1* null mice did exhibit a very small increase in MSI, suggesting that this complex may play a minor role in MMR (Prolla et al., 1998). Deletion of *mlh2* in yeast (the homologue of PMS1) did not increase mutation rates (Campbell et al., 2014) but was proposed to act as an accessory factor to enhance MutLα activity (Campbell et al., 2014). Furthermore, Pms1-null mice are fertile (Prolla et al., 1998), suggesting no gross defect in meiotic recombination in contrast to Mlh1 or Mlh3 null mice (Baker et al., 1996, Lipkin et al., 2002).

MLH1 is post-translationally modified and roles for these modifications are beginning to emerge. MLH1 is acetylated on K33, K241, K361 and K377, with deacetylation of these

residues by HDAC6 blocking MutLα-MutSα complex formation and conferring 6-TG resistance (Zhang et al., 2019b). It has been claimed that Activation of MLH1 in response to doxorubicin requires the presence of BRCA1, dependent on the ATM/ATR-mediated phosphorylation of MLH1 on S406 (Romeo et al., 2011). However, the authors provided no evidence of defects of MLH1 activity, i.e. survival in response to 6-TG or endonuclease activity of MutL complexes. Furthermore, doxorubicin does not induce mismatches, rather DSBs via topoisomerase II inhibition (Thorn et al., 2011). Importantly, loss of BRCA1 is not known to give rise to an MSI phenotype (Sokol et al., 2022, Looij et al., 2001). Therefore, the role of this modification in MLH1's MMR activity remains uncertain and requires further clarification. Additional PTMs of MLH1 have been reported via high-throughput methodologies but lack direct functional assessment, requiring further study *in vivo* (Santamaria et al., 2011, Mertins et al., 2013, Cptac et al., 2016, Akimov et al., 2018, Hendriks et al., 2014, Lumpkin et al., 2017, Huttlin et al., 2010).

As explained in section 2.2.5, heterozygous germline LOF mutations to MLH1 is diagnostic of LS (Papadopoulos et al., 1994, Bronner et al., 1994, Hemminki et al., 1994). LOF of MLH1 arises via multiple mechanisms including missense mutations, nonsense mutations, genomic rearrangements and promoter hypermethylation (Lynch et al., 2015). Importantly, LOF mutations to either *MLH1* or *MSH2* represent approximately 80-90% of LS cases (Sobocińska et al., 2020). Patients with LS arising from loss of MLH1 function typically display colorectal cancers, with a mean age of onset of 43-46 years (Lynch et al., 2015) but variation in the age of onset has been observed in families harbouring the same founder mutation in MLH1 (Peltomäki et al., 2001). MLH1 associated LS also gives rise to endometrial cancer, as well as cancer of the ovaries, stomach, kidneys, bladder and other tissues (Baretti and Le, 2018, Lynch et al., 2015, Møller et al., 2018). A recent study of over 3000 LS patients harboring pathological mutations in one of *MLH1*, *MSH2* or *MSH6* has provided fundamental insights into the cancer

predisposition associated with LOF mutations in *MLH1* (Møller et al., 2018). Specifically, LOF *MLH1* mutations conferred a 75% risk of developing any cancer by age 75, compared to 80% for *MSH2* and 60% for *MSH6* (Møller et al., 2018). LOF of *MLH1* was most likely to give rise to CRC at age 75 (50% risk) compared to MSH2 (40%) and MSH6 (15%) (Møller et al., 2018). As frequent screening for CRC in LS families decreases overall mortality by 65% (Järvinen* et al., 2000, Stoffel and Chittenden, 2010), clarity regarding whether or not mutations to *MLH1* (and other MMR genes) are LOF is of paramount importance to both diagnosis and subsequent management of LS. To this end, significant efforts have been made to catalogue (see section 2.3.3) and classify (see sections 2.3.4 and 2.4) the effects of mutations on MLH1 protein function. However, of the 1,468 missense mutations in MLH1 catalogued on ClinVar to date, 1,198 (81.6%) remain classified as VUS. This highlights a significant need for the further study of the effect of MLH1 variants *in vivo* via orthologue complementation.

The MMR system is widely conserved throughout evolution, primarily highlighted by its initial identification in prokaryotes, followed by eukaryotes (Modrich, 2016). A systematic phylogenetic analysis of the MMR genes throughout evolution shows that *mutS* and *mutL* genes were transferred from bacteria to eukaryotes via horizontal gene transfer (Lin et al., 2007). Furthermore, *mutS* and *mutL* genes have undergone long-term co-evolution (Lin et al., 2007) MutS and MutL and its homologues are present in all organisms ranging from bacteria to humans (Modrich, 2016). However, MutH is specific to prokaryotes, as prokaryotic MutL possesses no inherent endonuclease activity and therefore requires the endonuclease activity of MutH. Another key difference centers around strand-discrimination. In prokaryotes, strand-discrimination occurs due to the transient unmethylated state of adenine in newly replicated DNA (Lu et al., 1983). In eukaryotes, this is mediated by mirror image loading of PCNA on the leading and lagging strands (Pluciennik et al., 2010). The widespread conservation of the MMR system has enabled the use of model organisms, including yeast, to study the effect of

variants in these proteins. However, it must be mentioned that humanisation of yeast Mlh1 proved non-functional in some cases (Ellison et al., 2001), and human MLH1 has a dominant-negative effect on yeast MMR (Shimodaira et al., 1998). Despite this, the widespread conservation of MLH1 suggests that it is a good candidate for orthologue complementation of human cells.

In summary, MLH1 is fundamental to the MMR pathway with LOF mutations in MLH1 giving rise to MSI and LS. To date, a vast number of MLH1 variants remain as VUS on ClinVar, requiring further assessment *in vivo*. Clarification of their effect on MLH1 protein function would facilitate diagnosis and subsequent preventative screenings and treatments, improving overall patient mortality. LOF of MLH1 provides a clear assayable phenotype in loss of MMR function assayable via sensitivity to 6-TG, but also restoration of PMS2 stability. Furthermore, MLH1 experiences several PTMs with the importance of these modifications debated or still unknown. Crucially, MLH1 is highly conserved throughout evolution. Overall, MLH1 presents itself as an ideal candidate for orthologue complementation.

6.2 Aims and objectives

In the previous chapters, I demonstrated that orthologue complementation of XPA and XRCC4 could identify functional amino acid variants in the human protein as well as identifying non-essential PTM sites. Whilst both of these proteins are mutated in disease, the associated diseases are extremely rare and biallelic LOF is required for presentation of the phenotype. In this chapter, the focus shifts to orthologue complementation of MLH1, a highly clinically relevant protein in which heterozygous inheritance of a LOF MLH1 mutation is diagnostic of LS. This will ensure that the methodology can be applied to proteins with high clinical relevance. Firstly, the complementation ability of MLH1 orthologues will be assessed via multiple assays including IPs for known binding partners PMS2, MSH6 and EXO1, restoration of PMS2 stability and restoration of 6-TG sensitivity. Subsequently, sequence alignments of complementing orthologues will be constructed and searched for known benign and pathogenic variants available on public databases of variant effects. Following this, the sequence alignment will be searched for VUS in MLH1. A fraction of these will then be introduced into the human protein and functionally validated via restoration of 6-TG sensitivity and PMS2 stability in a human cell line. Additionally, the findings will be compared to bioinformatic predictions of variant effects and existing publications/variant databases.

6.3 Results

6.3.1 Selection of vertebrate MLH1 orthologues

As performed for XPA and XRCC4, NCBI databases were searched for orthologues of MLH1 from discrete taxonomic groups. Length of the orthologue was the primary criterion for selection with orthologues showing similar lengths to the human protein chosen for complementation. Orthologues were selected from *Mus musculus* (mouse), *Gallus gallus* (chicken) and *Xenopus laevis* (frog). Sequences from each protein of interest were then aligned using the T-Coffee algorithm (https://www.ebi.ac.uk/Tools/msa/tcoffee) (Notredame et al., 2000) and the conservation of identical or similar residues was highlighted using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html) **Figure 6-1**. The percentage identity to human MLH1 was then calculated using the % identity matrix output option from the T-Coffee alignment. Accession numbers and % identity to the human protein can be seen in **Table 6-1**.

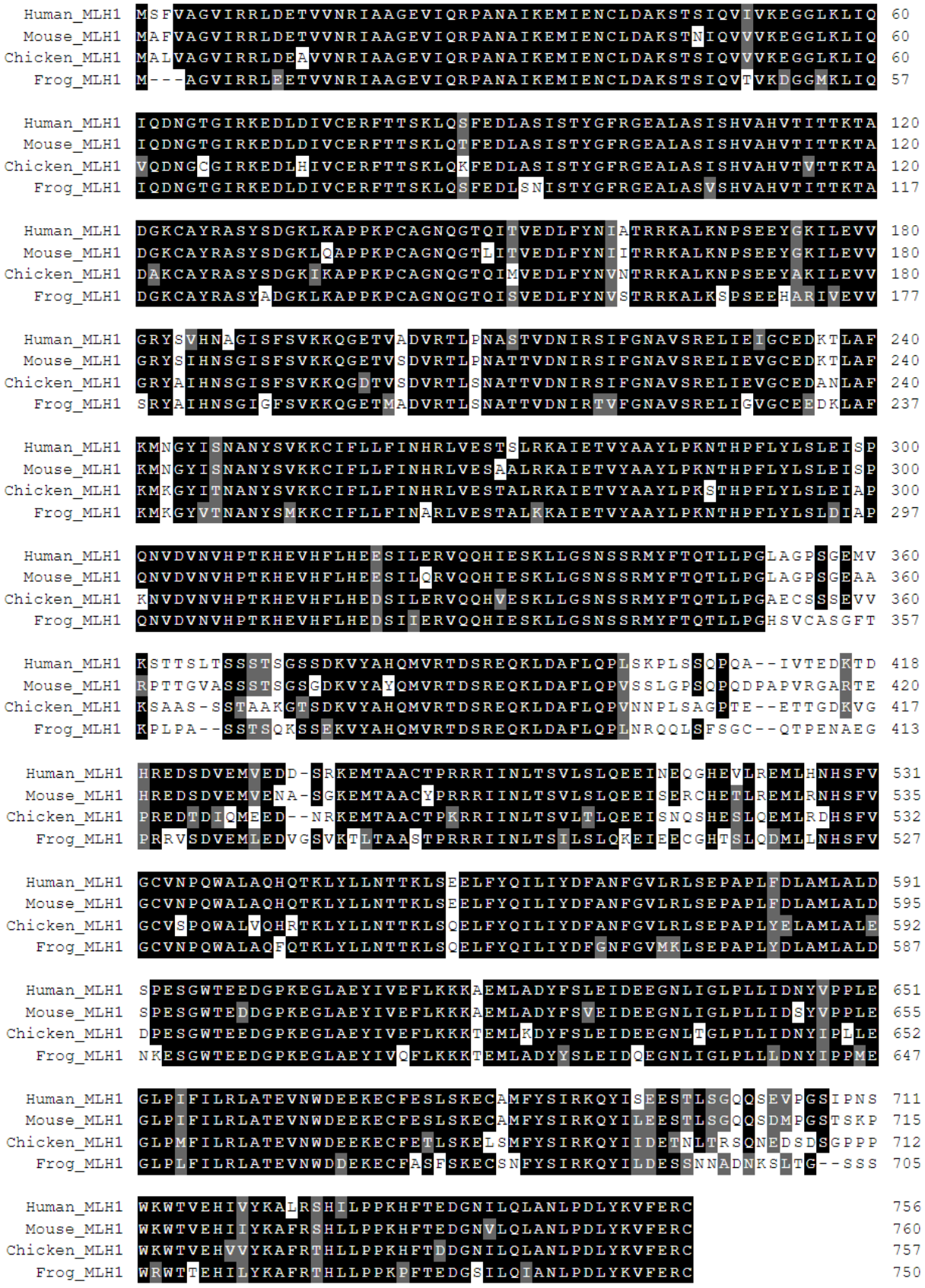


Figure 6-1: Sequence alignment of vertebrate MLH1 orthologues used in this study with conservation shading

Sequence alignment of vertebrate MLH1 orthologues used in this study. Sequences were aligned using the T-Coffee algorithm (Notredame et al., 2000) and conservation was scored using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html). Areas of black shading indicate 70% identical conservation of a residue, areas of gray shading indicate 70% conservation to at least an amino acid with similar biochemical properties. Areas of white shading indicate less than 70% conservation. Numbers indicate AA position.

MLH1 Orthologue	NCBI Reference Sequence	% Identity to Human Orthologue
Human	NP_000240.1	100%
Mouse	NP_081086.2	89
Chicken	XP_418828.1	80
Frog	NP_001090545.1	76

Table 6-1: Accession numbers and % identity to human orthologue for vertebrate MLH1 orthologues used

6.3.2 MLH1 orthologues display variable affinity for known binding partners in human cells

The ability of an orthologue to fold, express and complement human cells may be inferred from its ability to interact with known binding partners in a human cell line. To test this, GFP-tagged human, mouse, chicken and frog MLH1 orthologues were cloned into pEGFP-C1. These constructs were then expressed in 293FT cells and their ability to interact with the known binding partners MSH6 (Plotz et al., 2002, Räschle et al., 2002), EXO1 (Schmutte et al., 2001) and PMS2 (Li and Modrich, 1995) was assessed via immunoprecipitation. Human, mouse and chicken MLH1 retained the ability to interact with MSH6, EXO1 and PMS2 whereas frog MLH1 retained MSH6 and PMS2 binding, but showed reduced affinity for EXO1 **Figure 6-2.**

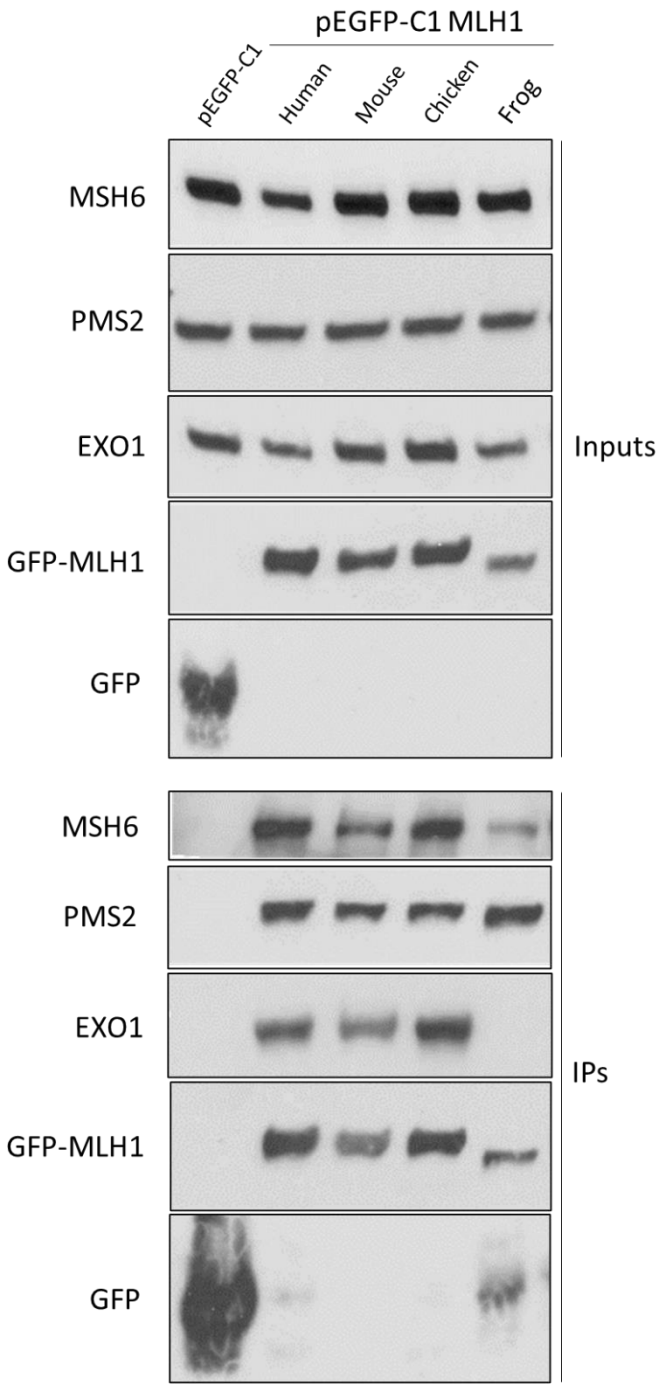


Figure 6-2: MLH1 orthologues show varying affinity for known binding partners in human cells

Western blot depicting GFP-pulldowns from 293FT cells transfected with either GFP or GFP-tagged human, mouse, chicken and frog MLH1 orthologues with probing for MSH6 (N=1), PMS2 (N=2) and EXO1 (N=2) and GFP (N=2).

6.3.3 MLH1 orthologues can restore PMS2 stability in 293T cells

Having observed the ability of MLH1 orthologues to express and bind to human PMS2 via IP, it was determined whether MLH1 orthologues could restore PMS2 stability in a PMS2-deficient human cell line. 293T cells are deficient in MLH1 due to promoter hypermethylation (Trojan et al., 2002). Complementation of 293T with MLH1 and subsequent restoration of PMS2 stability has been used as a readout for MLH1 function in previous studies (Trojan et al., 2002, Raevaara et al., 2005). Complementation of 293T cells with pEGFP-C1 constructs encoding human, mouse, chicken and frog MLH1 restored PMS2 stability to the same extent **Figure 6-3**. However, differences in expression level between the orthologues was observed, with mouse and frog showing reduced expression levels compared to human and chicken **Figure 6-3**. Evidently, this did not have an impact on the ability of all orthologues to restore PMS2 levels, corroborating the interaction observed between these orthologues and human PMS2 in the immunoprecipitation **Figure 6-2**.

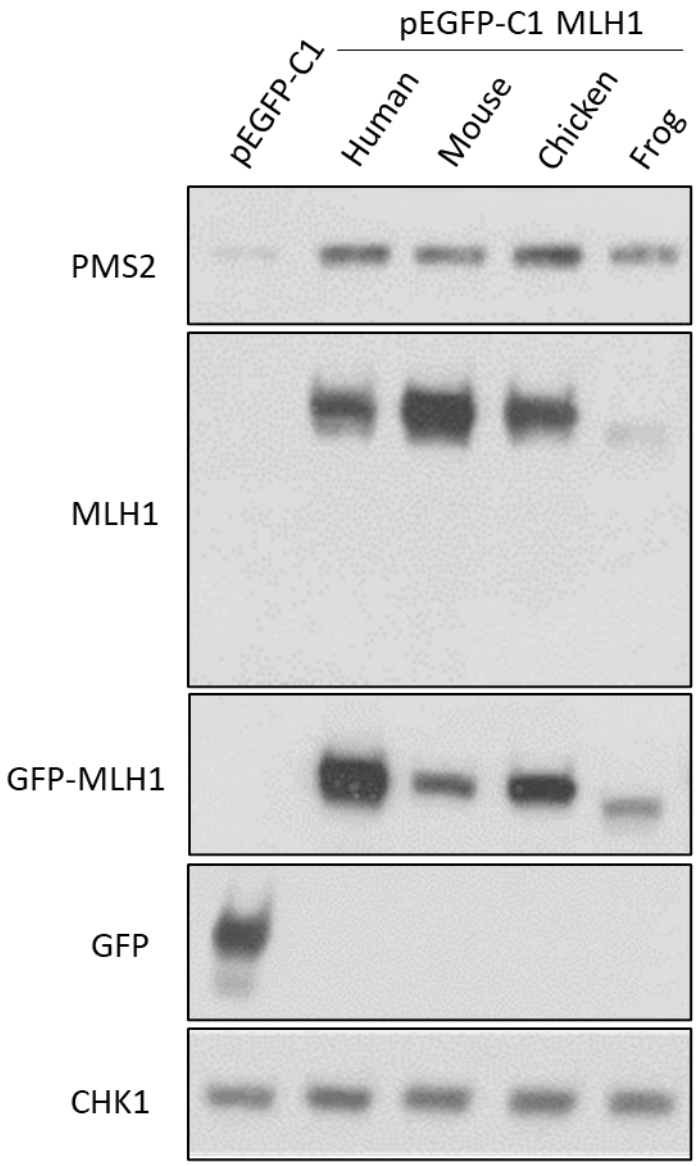


Figure 6-3: MLH1 orthologues can restore PMS2 stability in 293T cells

Western blot for PMS2, MLH1 and GFP of lysates from 293T cells transfected with either pEGFP-C1 or pEGFP-C1 constructs encoding human, mouse, chicken or frog MLH1. CHK1 is used as a loading control. N=1.

6.3.4 Frog MLH1 cannot restore sensitivity to 6-TG but restores PMS2 expression in HAP1 ΔMLH1 cells

MMR-deficiency leads to resistance to alkylating agents such as MNNG (Karran and Marinus, 1982) and antimetabolites such as 6-TG (Yamane et al., 2004, Yan et al., 2004). These compounds generate mismatches which are detected by the MMR pathway, resulting in the activation of the ATR/CHK1 axis and subsequent cell cycle arrest and apoptosis (Duckett et al., 1996, Cejka et al., 2003, Jiricny, 2006, Yamane et al., 2004, Yan et al., 2004). Because of this, sensitivity to such agents has been utilised in multiple studies as a readout for MLH1 function and the impact of MLH1 variants on protein function (Blasi et al., 2006, Houlleberghs et al., 2020, Rath et al., 2022). Therefore, I assessed the ability of MLH1 orthologues to restore sensitivity to 6-TG in MLH1-deficient human cells. MLH1 orthologues were tagged on their N-terminus with GFP and cloned into the pLenti-PGK vector and lentiviruses were produced. HAP1 Δ*MLH1* cells were then transduced with these lentiviruses and GFP+ cells were sorted via FACS **Figure 10-6** and seeded as single cells. Cells were then treated with varying doses of 6-TG for 24 hours. Colony formation was then assessed 1 week after FACS. Mouse and chicken MLH1 orthologues could restore sensitivity to 6-TG to the same extent as human MLH1 **Figure 6-4**. In contrast, frog MLH1 displayed an intermediate phenotype, showing only partial restoration of 6-TG sensitivity in this assay **Figure 6-4.** Expression levels of each orthologue differed in this system **Figure 6-5**, but PMS2 levels were restored by all orthologues tested **Figure 6-5**, consistent with results displayed in **Figure 6-3.**

#

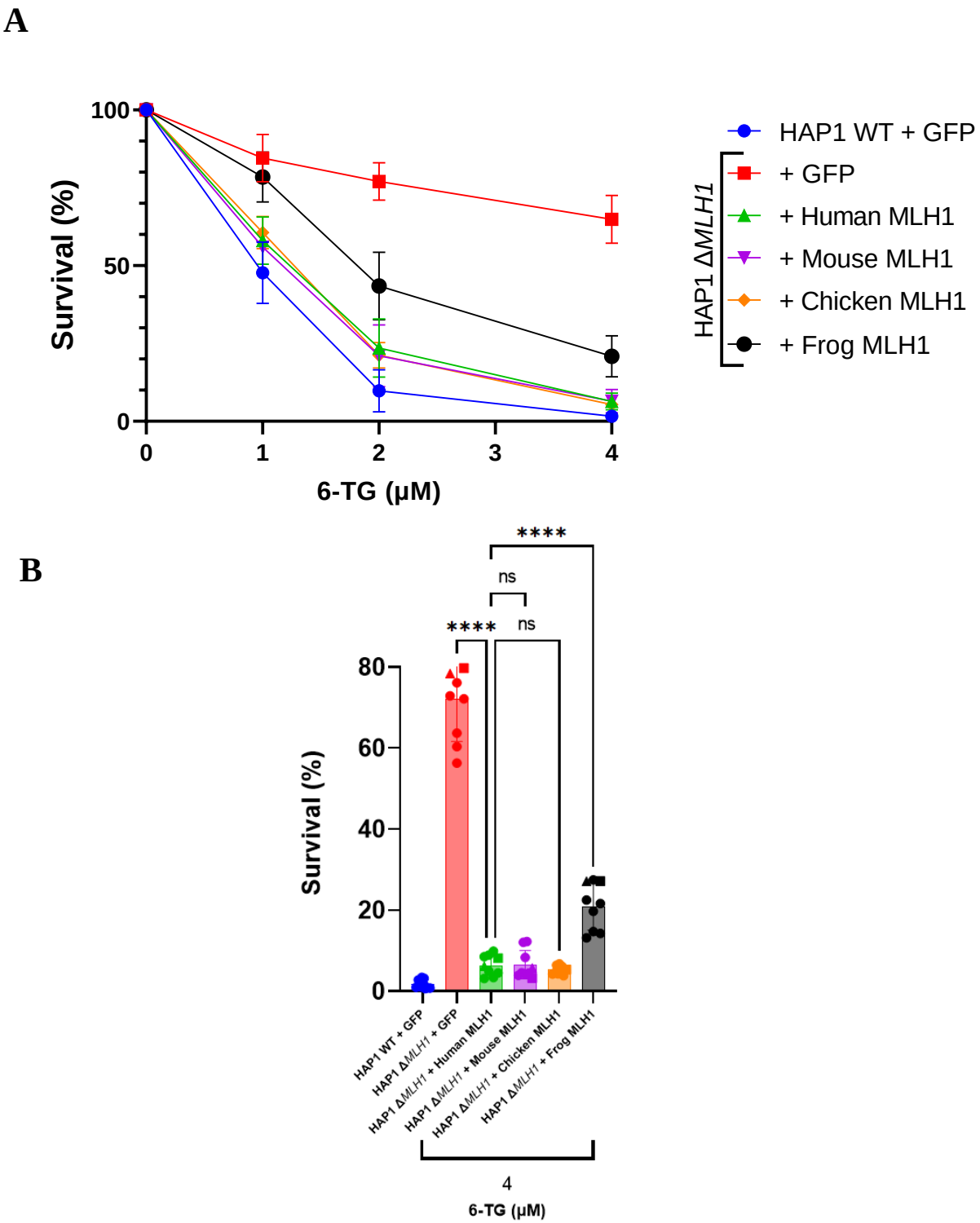


Figure 6-4: Frog MLH1 cannot restore sensitivity to 6-TG in HAP1 Δ*MLH1* cells

A: Survival curves demonstrating the survival % compared to untreated cells in response to increasing doses of 6-TG for HAP1 WT and HAP1 Δ*MLH1* cells complemented with lentivirus encoding either GFP or GFP-tagged MLH1 orthologues. Data is plotted as the mean of N=3. Error bars indicate SEM. **B**: One-way ANOVA was performed on all values of N=3 at 4 μM 6-TG and included a Dunnett's multiple comparisons test comparing HAP1 Δ*MLH1* + GFP-human MLH1 to HAP1 Δ*MLH1* + GFP, GFP-mouse, GFP-chicken and GFP-frog MLH1. Results from Dunnett's multiple comparisons test are displayed. ns = non-significant difference, **** = p <0.0001.

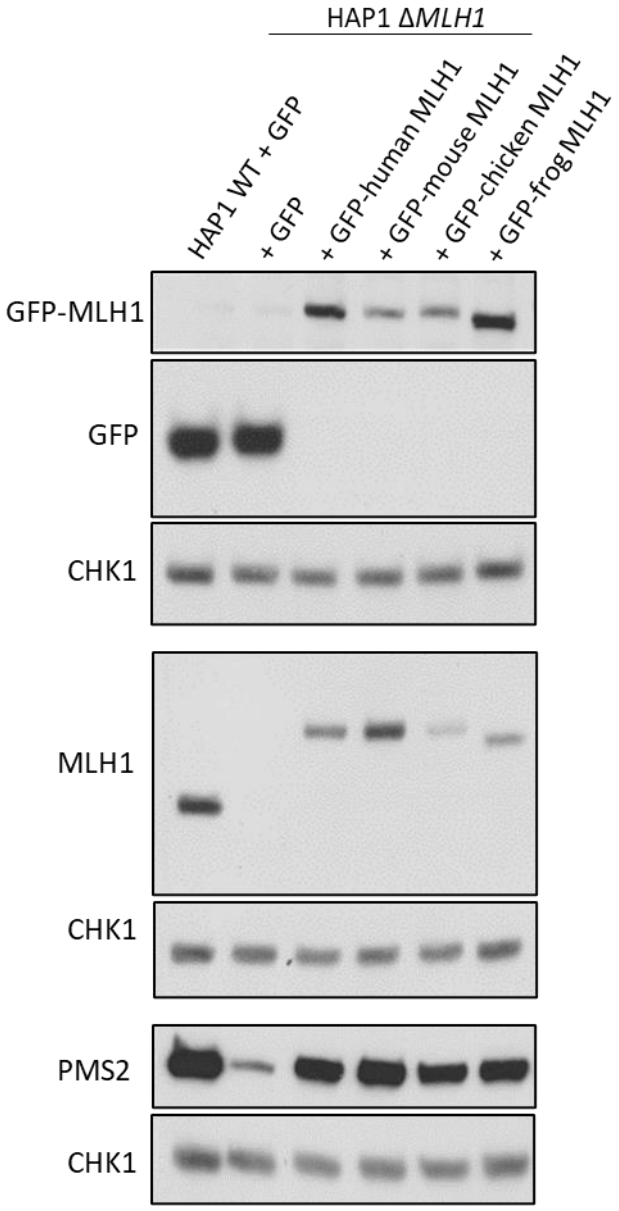


Figure 6-5: Vertebrate GFP-MLH1 orthologues restore PMS2 stability in HAP1 Δ*MLH1* cells

Western blot for GFP, PMS2 and MLH1 of lysates from HAP1 WT or HAP1 Δ*MLH1* cells constitutively expressing GFP-tagged MLH1 orthologues or GFP only. CHK1 was used as a loading control. N=2.

6.3.5 Frog MLH1 cannot restore sensitivity to 6-TG or PMS2 stability in HCT116 cells

Having assessed the ability of MLH1 orthologues to complement the HAP1 Δ*MLH1* cell line, an additional cell line was used to further validate the method. The HCT116 cell line was chosen due to it being derived from a malignant colonic carcinoma (Brattain et al., 1981). Furthermore, HCT116 has been utilised in multiple previous studies to assess the complementation ability of MLH1 (Trojan et al., 2002, Takahashi et al., 2007, Bouvet et al., 2019). Similar to HAP1 Δ*MLH1*, mouse and chicken MLH1 could restore sensitivity to 6-TG to the same extent as human, whereas frog MLH1 displayed only partial restoration **Figure 6-6**. PMS2 levels were restored to the same extent by human, mouse and chicken MLH1 in HCT116 **Figure 6-7**. In contrast to HAP1 Δ*MLH1*, frog MLH1 could not restore PMS2 stability to the same extent in HCT116 **Figure 6-7**. The discrepancy in this result may derive from the difference in expression level of frog MLH1 in HAP1 Δ*MLH1* **Figure 6-5** and HCT116 **Figure 6-7**. For all assays with both HAP1 Δ*MLH1* and HCT116, GFP + cells were selected via FACS **Figure 10-6**, **Figure 10-7** and maintained in selective growth conditions therefore the reasons underlying the difference in expression is challenging to narrow down.

In conclusion, human, mouse and chicken MLH1 were able to interact with known binding partners EXO1, PMS2 and MSH6 as well as restoring 6-TG sensitivity and PMS2 stability in two separate MLH1-deficient cell lines. In contrast, frog MLH1 was unable to bind to EXO1, failed to restore 6-TG sensitivity in either MLH1-deficient cell line and showed variable ability to restore PMS2 stability. These results are summarised in **Table 6-2**. Based on these results, an alignment of human, mouse and chicken MLH1 was constructed **Figure 6-8** for use in identifying functional amino acid variants in the human protein.

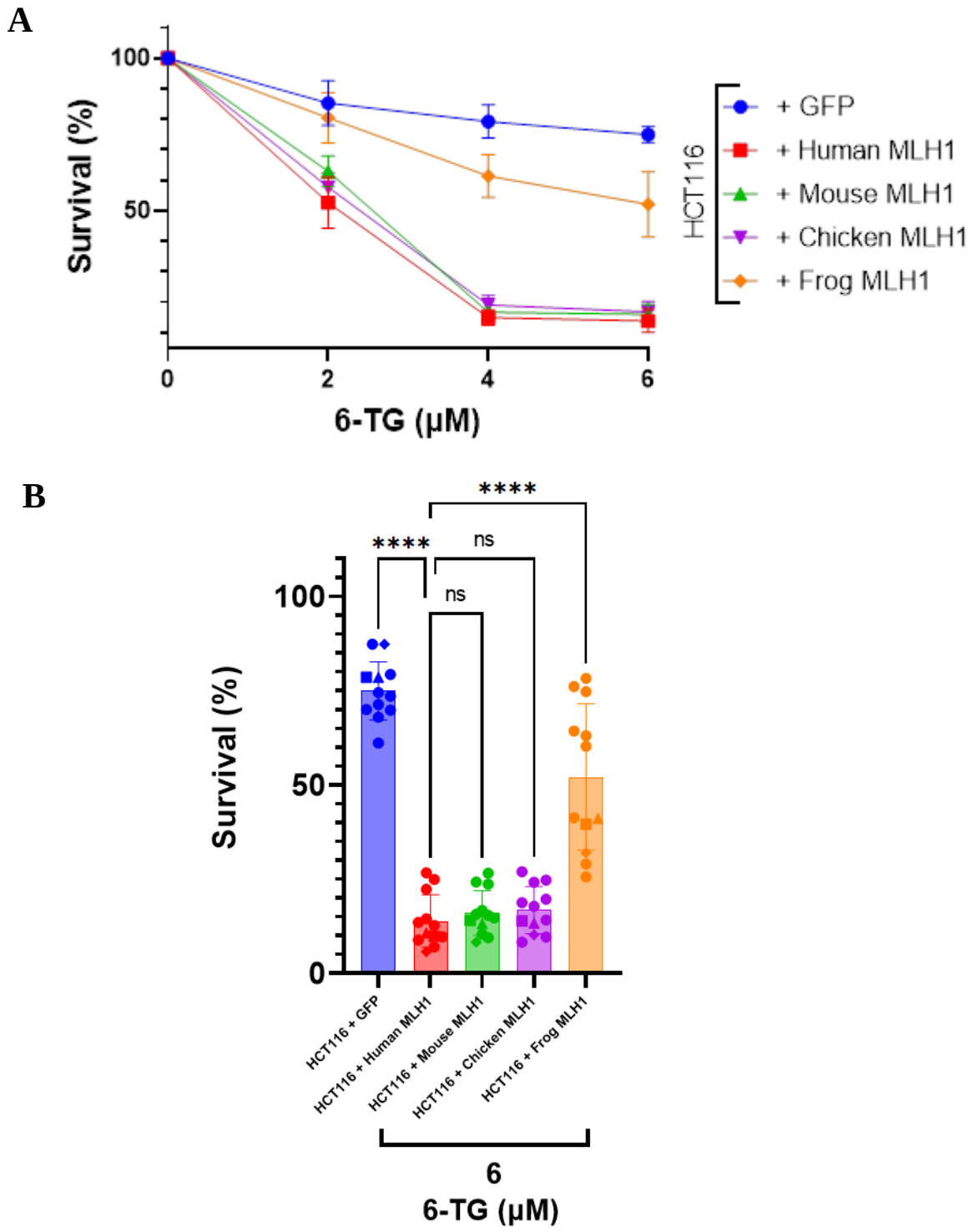


Figure 6-6: Frog MLH1 cannot restore sensitivity to 6-TG in HCT116 cells

A: Survival curves demonstrating the colony survival % compared to untreated cells in response to increasing doses of 6-TG for HCT116 cells complemented with GFP-tagged MLH1 orthologues or GFP only controls. Data is plotted as the mean of N=4. Error bars indicate SEM. **B:** One-way ANOVA was performed on all values of N=4 at 4 μM 6-TG and included a Dunnett's multiple comparisons test comparing HCT116 + GFP-human MLH1 to HCT116 + GFP-mouse, chicken and frog MLH1. Results from Dunnett's multiple comparisons test are displayed. ns = non-significant difference, **** = p <0.0001.

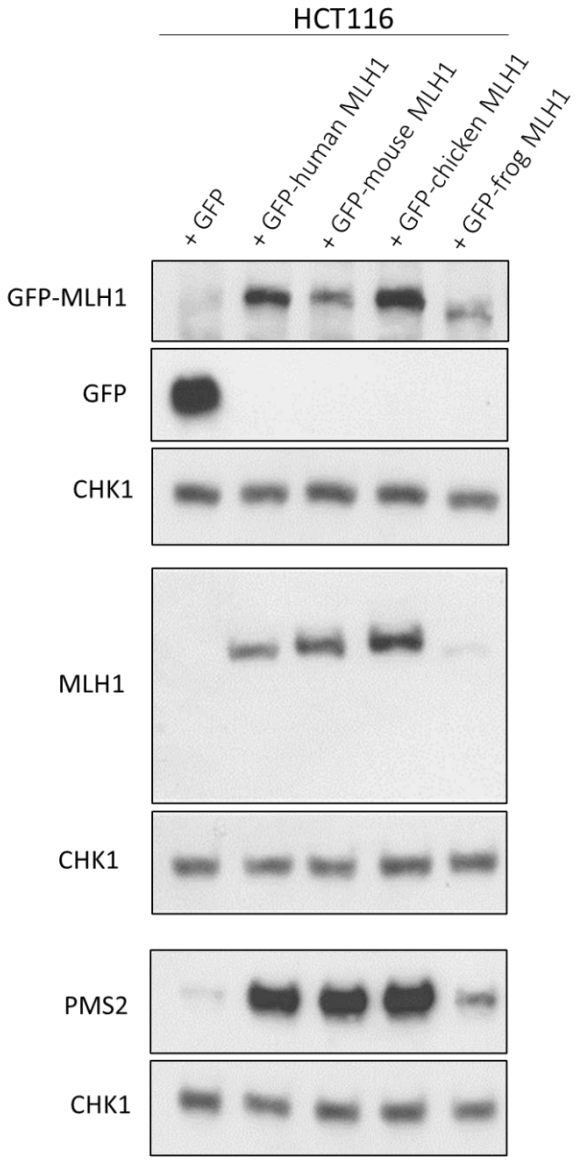


Figure 6-7: Frog MLH1 partially restores PMS2 expression in HCT116 cells

Western blot for GFP, MLH1 and PMS2 of lysates from HCT116 cells constitutively expressing GFP-tagged MLH1 orthologues or GFP only for PMS2. CHK1 was used as a loading control. N=2

MLH1 orthologue	MSH6 interaction	PMS2 interaction	EXO1 interaction	Restoration of PMS2 stability	Restoration of 6-TG sensitivity	Complementation?
Human	Yes	Yes	Yes	Yes	Yes	Yes
Mouse	Yes	Yes	Yes	Yes	Yes	Yes
Chicken	Yes	Yes	Yes	Yes	Yes	Yes
Frog	Yes	Yes	Reduced	Partial	Partial	No

Table 6-2: Summary of the complementation ability of MLH1 orthologues in human cells



Figure 6-8: **Sequence alignment of vertebrate MLH1 orthologues able to complement human cells with conservation shading**

Sequence alignment of vertebrate MLH1 orthologues used in this study. Sequences were aligned using the T-Coffee algorithm (Notredame et al., 2000) and conservation was scored using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html). Areas of black shading indicate 70% identical conservation of a residue, areas of gray shading indicate 60% conservation to at least an amino acid with similar biochemical properties. Areas of white shading indicate less than 60% conservation. Numbers indicate AA position.

6.3.6 A sequence alignment of complementing MLH1 orthologues does not contain known benign or pathogenic MLH1 variants

As performed for XPA and XRCC4, I sought to determine whether the sequence alignment of complementing MLH1 orthologues contained known benign variants. Firstly, the alignment was searched for homozygous MLH1 variants present in the gnomAD database. This revealed that the alignment did not contain any such variant in any of the complementing orthologues **Table 6-3.** The analysis was then expanded to all gnomAD variants with an allele frequency of >0.01% in a control group of >160 healthy individuals. The InSiGHT criteria for MMR variant classification states that variants present at this frequency in an ostensibly healthy population is sufficient evidence to at least contribute to a class 1 (benign) classification (Thompson et al., 2014). Analysis revealed that none of the MLH1 alleles in gnomAD present at frequencies > 0.01% were present in any of the complementing orthologues **Table 6-4.** The analysis was then expanded to all gnomAD variants, regardless of allele frequency. This revealed multiple variants identified heterozygously in ostensibly healthy individuals which were present in complementing MLH1 orthologues **Table 6-5**. Of all the variants identified, none exceeded an allele frequency of 0.01% and therefore do not fulfill the benign criteria (Thompson et al., 2014), remaining classified as VUS on ClinVar (where submitted) **Table 6-5**. However, their presence in complementing orthologues demonstrates their benignity. Lastly, the sequence alignment was searched for the presence of benign variants for MLH1 catalogued on ClinVar. This did not reveal any known benign variants as being present in complementing orthologues **Table 6-6**. Having failed to identify any known benign variants, it was determined whether any pathogenic variants exist within complementing orthologues outside of the exon boundaries of MLH1 **Table 6-8**. The sequence alignment of complementing orthologues was compared to all pathogenic MLH1 variants catalogued on ClinVar with expert

panel review status. This showed that none of the pathogenic variants were present in any of the complementing orthologues **Table 6-7**.

rsIDs (gnomAD identifier)	Residue position	WT	Variant	Residue in mouse	Residue in chicken	Allele frequency	Homozygote count	ClinVar Classification
rs2308317	213	V	M	V	V	0.27%	12	Benign
rs1799977	219	I	V	I	I	23.24%	9245	Benign
rs63750948	247	S	A	S	T	0.0036%	1	Uncertain significance
rs63751597	264	H	N	H	H	0.0008%	1	N/A
rs63751049	326	V	A	V	V	0.046%	1	Benign
rs63750447	384	V	D	V	V	0.26%	7	Benign
rs56185292	577	S	L	S	S	0.006%	1	Conflicting interpretations of pathogenicity
rs35001569	618	K	E	K	K	0.34%	4	Benign
rs63750449	618	K	T	K	K	0.34%	4	Benign
rs35831931	716	V	M	V	V	0.13%	1	Benign
rs2020873	718	H	Y	H	H	0.73%	86	Benign

Table 6-3: Results from searching the sequence alignment of complementing MLH1 orthologues for homozygous MLH1 variants present in gnomAD.

The sequence alignment of complementing MLH1 orthologues (human-chicken) was searched for homozygous MLH1 variants present on the gnomAD database (Karczewski et al., 2020). No homozygous MLH1 variants were present in complementing orthologues.

rsIDs (gnomAD identifier)	Residue position	WT	Variant	Residue in mouse	Residue in chicken	Allele frequency	ClinVar Classification
rs730881746	7	V	F	V	V	0.017%	Uncertain significance
rs41295280	22	G	A	G	G	0.013%	Benign
rs28930073	132	D	H	D	D	0.022%	Benign
rs2308317	213	V	M	V	V	0.27%	Benign
rs4986984	217	R	C	R	R	0.34%	Conflicting interpretations
rs1799977	219	I	V	I	I	23.24%	Benign
rs63750650	268	E	G	E	E	0.024%	Benign
rs63751049	326	V	A	V	V	0.046%	Benign
rs63750447	384	V	D	V	V	0.26%	Benign
rs41294980	406	S	N	S	A	0.09%	Benign
rs63750365	441	A	T	A	A	0.40%	Benign
rs63750527	454	G	R	E	P	0.013%	Conflicting interpretations
rs202038499	460	E	A	Q	G	0.015%	Conflicting interpretations
rs63751612	578	E	G	E	E	0.012%	Benign
rs63751684	581	P	L	P	P	0.017%	Likely benign
rs63750876	603	P	R	P	P	0.016%	Likely benign
rs41295284	607	L	H	L	L	0.019%	Likely benign
rs35001569	618	K	E	K	K	0.34%	Benign
rs63750449	618	K	T	K	K	0.34%	Benign
rs63750109	647	V	M	V	I	0.011%	Uncertain significance
rs55907433	655	I	V	I	M	0.12%	Likely benign
rs63750702	689	Q	R	Q	Q	0.025%	Benign
rs63750114	701	Q	K	Q	Q	0.049%	Likely benign
rs35831931	716	V	M	V	V	0.13%	Benign
rs2020873	718	H	Y	H	H	0.73%	Benign
rs566928243	725	R	H	R	R	0.017%	Conflicting interpretations

Table 6-4: Results of searching the sequence alignment of complementing MLH1 orthologues for gnomAD variants with allele frequencies > 0.01%

The sequence alignment of complementing MLH1 orthologues (human-chicken) was searched for MLH1 variants present on the gnomAD database present at frequencies > 0.01% (Karczewski et al., 2020). No such MLH1 variants were present in complementing orthologues.

rsIDs (gnomAD identifier)	Residue position	WT	Variant	Present in (orthologue)	Allele frequency	ClinVar classification
rs1448308275	2	S	A	Mouse, chicken	0.0032%	Uncertain significance
rs779759678	3	F	L	Chicken	0.0011%	Conflicting interpretations
rs371667663	115	I	V	Chicken	0.0008%	Uncertain significance
rs776969475	151	T	M	Chicken	0.0016%	Uncertain significance
rs63750012	185	V	I	Mouse, chicken	0.0004%	Uncertain significance
rs587781509	208	P	S	Chicken	0.0008%	Uncertain significance
rs371302926	270	T	A	Mouse	0.0036%	Conflicting interpretations
rs756347993	355	P	S	Chicken	0.0004%	Uncertain significance
rs1376611280	363	T	A	Chicken	0.0004%	N/A
rs1258614019	367	T	A	Mouse	0.0004%	N/A
rs762426947	371	T	A	Chicken	0.0004%	Uncertain significance
rs1060500694	375	S	G	Mouse	0.0004%	Uncertain significance
rs63750557	381	H	Y	Mouse	0.0004%	Uncertain significance
rs587779951	401	S	N	Chicken	0.0004%	Uncertain significance
rs754898711	418	D	G	Chicken	0.0004%	Uncertain significance
rs370687064	423	R	T	Chicken	0.0007%	Uncertain significance
rs763189331	444	N	S	Mouse	0.0004%	Conflicting interpretations
rs772245091	514	N	S	Mouse, chicken	0.0008%	Uncertain significance
rs63751630	522	R	Q	Chicken	0.0032%	Conflicting interpretations
rs1304802474	526	H	R	Mouse, chicken	0.0064%	Conflicting interpretations
rs757728402	541	A	V	Chicken	0.0004%	Uncertain significance
rs878853783	584	D	E	Chicken	0.0004%	Uncertain significance
rs781637991	675	S	T	Chicken	0.002%	Uncertain significance
rs749100096	710	N	K	Mouse	0.0004%	Uncertain significance

rs281864940	719	I	V	Chicken	0.0004%	Conflicting interpretations

Table 6-5: Results of searching the sequence alignment of complementing MLH1 orthologues for all MLH1 gnomAD variants

The sequence alignment of complementing MLH1 orthologues (human-chicken) was searched for MLH1 variants present on the gnomAD database, regardless of allele frequency (Karczewski et al., 2020).

ClinVar Accession Number	Residue Position	WT	Variant	Residue in mouse	Residue in chicken
NM_000249.4(MLH1):c.65G>C (p.Gly22Ala)	22	G	A	G	G
NM_000249.4(MLH1):c.394G>C (p.Asp132His)	132	D	H	D	D
NM_000249.4(MLH1):c.637G>A (p.Val213Met)	213	V	M	V	V
NM_000249.4(MLH1):c.655A>C (p.Ile219Leu)	219	I	L	I	I
NM_000249.4(MLH1):c.655A>G (p.Ile219Val)	219	I	V	I	I
NM_000249.4(MLH1):c.803A>G (p.Glu268Gly)	268	E	G	E	E
NM_000249.4(MLH1):c.977T>C (p.Val326Ala)	326	V	A	V	V
NM_000249.4(MLH1):c.1040C>A (p.Thr347Asn)	347	T	N	T	T
NM_000249.4(MLH1):c.1151T>A (p.Val384Asp)	384	V	D	V	V
NM_000249.4(MLH1):c.1217G>A (p.Ser406Asn)	406	S	N	S	A
NM_000249.4(MLH1):c.1321G>A (p.Ala441Thr)	441	A	T	A	A
NM_000249.4(MLH1):c.1652A>G (p.Asn551Ser)	551	N	S	N	N
NM_000249.4(MLH1):c.1733A>G (p.Glu578Gly)	578	E	G	E	E
NM_000249.4(MLH1):c.1799A>G (p.Glu600Gly)	600	E	G	E	E
NM_000249.4(MLH1):c.1852A>G (p.Lys618Glu)	618	K	E	K	K
NM_000249.4(MLH1):c.1852_1853delinsGC (p.Lys618Ala)	618	K	A	K	K
NM_000249.4(MLH1):c.1853A>C (p.Lys618Thr)	618	K	T	K	K
NM_000249.4(MLH1):c.2066A>G (p.Gln689Arg)	689	Q	R	Q	Q

NM_000249.4(MLH1):c.2074T>C (p.Ser692Pro)	692	S	P	L	I
NM_000249.4(MLH1):c.2146G>A (p.Val716Met)	716	V	M	V	V
NM_000249.4(MLH1):c.2152C>T (p.His718Tyr)	718	H	Y	H	H

Table 6-6: Results of searching the sequence alignment of complementing MLH1 orthologues for benign ClinVar MLH1 variants

The sequence alignment of complementing MLH1 orthologues (human-chicken) was searched for all benign MLH1 variants present on the ClinVar database. No known benign variants were found in complementing orthologues.

ClinVar Accession Number	Residue Position	WT	Variant	Residue in mouse	Residue in chicken	Encoded across an exon boundary?
NM_000249.4(MLH1):c.1A>T (p.Met1Leu)	1	M	L	M	M	No
NM_000249.4(MLH1):c.1A>G (p.Met1Val)	1	M	V	M	M	No
NM_000249.4(MLH1):c.2T>G (p.Met1Arg)	1	M	R	M	M	No
NM_000249.4(MLH1):c.2T>C (p.Met1Thr)	1	M	T	M	M	No
NM_000249.4(MLH1):c.2T>A (p.Met1Lys)	1	M	K	M	M	No
NM_000249.4(MLH1):c.3G>A (p.Met1Ile)	1	M	I	M	M	No
NM_000249.4(MLH1):c.55A>T (p.Ile19Phe)	19	I	F	I	I	No
NM_000249.4(MLH1):c.62C>T (p.Ala21Val)	21	A	V	A	A	No
NM_000249.4(MLH1):c.62C>A (p.Ala21Glu)	21	A	E	A	A	No
NM_000249.4(MLH1):c.83C>T (p.Pro28Leu)	28	P	L	P	P	No
NM_000249.4(MLH1):c.86C>G (p.Ala29Gly)	29	A	G	A	A	No
NM_000249.4(MLH1):c.104T>G (p.Met35Arg)	35	M	R	M	M	No
NM_000249.4(MLH1):c.104_105delinsAC (p.Met35Asn)	35	M	N	M	M	No
NM_000249.4(MLH1):c.112A>C (p.Asn38His)	38	N	H	N	N	Yes
NM_000249.4(MLH1):c.113A>C (p.Asn38Thr)	38	N	T	N	N	Yes
NM_000249.4(MLH1):c.113A>G (p.Asn38Ser)	38	N	S	N	N	Yes
NM_000249.4(MLH1):c.114C>G (p.Asn38Lys)	38	N	K	N	N	Yes
NM_000249.4(MLH1):c.121G>C (p.Asp41His)	41	D	H	D	D	No
NM_000249.4(MLH1):c.122A>G (p.Asp41Gly)	41	D	G	D	D	No
NM_000249.4(MLH1):c.131_132delinsTT (p.Ser44Phe)	44	S	F	S	S	No
NM_000249.4(MLH1):c.131C>A (p.Ser44Tyr)	44	S	Y	S	S	No
NM_000249.4(MLH1):c.131C>T (p.Ser44Phe)	44	S	F	S	S	No
NM_000249.4(MLH1):c.146T>A (p.Val49Glu)	49	V	E	V	V	No
NM_000249.4(MLH1):c.187G>A (p.Asp63Asn)	63	D	N	D	D	No
NM_000249.4(MLH1):c.189C>A (p.Asp63Glu)	63	D	E	D	D	No
NM_000249.4(MLH1):c.199G>C (p.Gly67Arg)	67	G	R	G	G	No
NM_000249.4(MLH1):c.199G>A (p.Gly67Arg)	67	G	R	G	G	No
NM_000249.4(MLH1):c.199G>T (p.Gly67Trp)	67	G	W	G	G	No

NM_000249.4(MLH1):c.200G>A (p.Gly67Glu)	67	G	E	G	G	No
NM_000249.4(MLH1):c.229T>C (p.Cys77Arg)	77	C	R	C	C	No
NM_000249.4(MLH1):c.230G>A (p.Cys77Tyr)	77	C	Y	C	C	No
NM_000249.4(MLH1):c.245C>T (p.Thr82Ile)	82	T	I	T	T	No
NM_000249.4(MLH1):c.299G>C (p.Arg100Pro)	100	R	P	R	R	No
NM_000249.4(MLH1):c.306G>C (p.Glu102Asp)	102	E	D	E	E	Yes
NM_000249.4(MLH1):c.320T>A (p.Ile107Lys)	107	I	K	I	I	No
NM_000249.4(MLH1):c.320T>G (p.Ile107Arg)	107	I	R	I	I	No
NM_000249.4(MLH1):c.350C>T (p.Thr117Met)	117	T	M	T	T	No
NM_000249.4(MLH1):c.380G>A (p.Arg127Lys)	127	R	K	R	R	Yes
NM_000249.4(MLH1):c.382G>C (p.Ala128Pro)	128	A	P	A	A	No
NM_000249.4(MLH1):c.464T>G (p.Leu155Arg)	155	L	R	L	L	No
NM_000249.4(MLH1):c.544A>G (p.Arg182Gly)	182	R	G	R	R	Yes
NM_000249.4(MLH1):c.545G>A (p.Arg182Lys)	182	R	K	R	R	Yes
NM_000249.4(MLH1):c.554T>G (p.Val185Gly)	185	V	G	I	I	No
NM_000249.4(MLH1):c.677G>A (p.Arg226Gln)	226	R	Q	R	R	Yes
NM_000249.4(MLH1):c.731G>A (p.Gly244Asp)	244	G	D	G	G	No
NM_000249.4(MLH1):c.739T>C (p.Ser247Pro)	247	S	P	S	T	No
NM_000249.4(MLH1):c.779T>G (p.Leu260Arg)	260	L	R	L	L	No
NM_000249.4(MLH1):c.793_794delinsGT (p.Arg265Val)	265	R	V	R	R	No
NM_000249.4(MLH1):c.793C>A (p.Arg265Ser)	265	R	S	R	R	No
NM_000249.4(MLH1):c.793C>T (p.Arg265Cys)	265	R	C	R	R	No
NM_000249.4(MLH1):c.827T>G (p.Ile276Arg)	276	I	R	I	I	No
NM_000249.4(MLH1):c.842C>T (p.Ala281Val)	281	A	V	A	A	No
NM_000249.4(MLH1):c.883A>G (p.Ser295Gly)	295	S	G	S	S	Yes
NM_000249.4(MLH1):c.883A>C (p.Ser295Arg)	295	S	R	S	S	Yes
NM_000249.4(MLH1):c.884G>A (p.Ser295Asn)	295	S	N	S	S	Yes

NM_000249.4(MLH1):c.918T>A (p.Asn306Lys)	306	N	K	N	N	No
NM_000249.4(MLH1):c.986A>C (p.His329Pro)	329	H	P	H	H	No
NM_000249.4(MLH1):c.1037A>G (p.Gln346Arg)	346	Q	R	Q	Q	Yes
NM_000249.4(MLH1):c.1038G>T (p.Gln346His)	346	Q	H	Q	Q	Yes
NM_000249.4(MLH1):c.1038G>C (p.Gln346His)	346	Q	H	Q	Q	Yes
NM_000249.4(MLH1):c.1649T>C (p.Leu550Pro)	550	L	P	L	L	No
NM_000249.4(MLH1):c.1667G>A (p.Ser556Asn)	556	S	N	S	S	Yes
NM_000249.4(MLH1):c.1667G>T (p.Ser556Ile)	556	S	I	S	S	Yes
NM_000249.4(MLH1):c.1745T>C (p.Leu582Pro)	582	L	P	L	L	No
NM_000249.4(MLH1):c.1865T>C (p.Leu622Pro)	622	L	P	L	L	No
NM_000249.4(MLH1):c.1865T>A (p.Leu622His)	622	L	H	L	L	No
NM_000249.4(MLH1):c.1942C>T (p.Pro648Ser)	648	P	S	P	P	No
NM_000249.4(MLH1):c.1961C>T (p.Pro654Leu)	654	R	L	R	R	No
NM_000249.4(MLH1):c.1976G>T (p.Arg659Leu)	659	R	L	R	R	No
NM_000249.4(MLH1):c.1976G>C (p.Arg659Pro)	659	R	P	R	R	No
NM_000249.4(MLH1):c.1989G>C (p.Glu663Asp)	663	E	D	E	E	Yes
NM_000249.4(MLH1):c.1989G>T (p.Glu663Asp)	663	E	D	E	E	Yes
NM_000249.4(MLH1):c.2041G>A (p.Ala681Thr)	681	A	T	A	S	No
NM_000249.4(MLH1):c.2059C>T (p.Arg687Trp)	687	R	W	R	R	No
NM_000249.4(MLH1):c.2103G>C (p.Gln701His)	701	Q	H	Q	Q	Yes
NM_000249.4(MLH1):c.2246T>C (p.Leu749Pro)	749	L	P	L	L	No
NM_000249.4(MLH1):c.2263A>G (p.Arg755Gly)	755	R	G	R	R	No

Table 6-7: Results of searching the sequence alignment of complementing MLH1 orthologues for pathogenic MLH1 variants present on the ClinVar database.

The sequence alignment of complementing MLH1 orthologues (human-chicken) was searched for all pathogenic MLH1 variants present on the ClinVar database. No known pathogenic variants were found in complementing orthologues.

Exon number	Amino acids	Amino acid(s) encoded across exon boundary
Exon 1	1-39	N38, C39
Exon 2	40-69	R69, K70
Exon 3	70-102	E102, A103
Exon 4	103-127	Y126, R127
Exon 5	128-151	T151, V152
Exon 6	152-182	G181, R182
Exon 7	183-196	K196, Q197
Exon 8	197-226	S225, R226
Exon 9	227-264	N263, H264
Exon 10	265-295	L294, S295
Exon 11	296-346	Q346, T347
Exon 12	347-470	P469, R470
Exon 13	471-520	E519, V520
Exon 14	521-556	L555, S556
Exon 15	557-577	S577, E578
Exon 16	578-632	E632, E633
Exon 17	633-663	E663, V664
Exon 18	664-701	Q701, S702
Exon 19	702-756	N/A

Table 6-8: Amino acid positions of MLH1 exons and the amino acids encoded across exon boundaries

A table displaying the amino acid positions of the MLH1 exons and the amino acids encoded across the exon boundaries. Exon positions were obtained from CCDS. https://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi?REQUEST=CCDS&DATA=CCDS2663.1

6.3.7 Selection of variants for assessment based on sequence alignments and database searches

As explained previously, the number of missense mutations in MLH1 catalogued as VUS on ClinVar stands at 81.6%. As previously demonstrated for XPA, the alignment of complementing orthologues should enable the identification of variants tolerated by the human protein. To this end, the ClinVar database was searched for variants present in the alignment of human, mouse and chicken MLH1 **Table 6-9**. In total the alignment identified 59 variants that were not classified as either benign or pathogenic and were present in either one or both of mouse and chicken MLH1, outside of exon boundaries **Table 6-9**. Scores from *in silico* analysis of the impact of these variants was then obtained using the MAPP (Stone and Sidow, 2005) and PolyPhen2 (Adzhubei et al., 2010) algorithms in line with recommendations by the InSiGHT consortium (Thompson et al., 2013), https://www.insight-group.org/content/uploads/2021/11/DRAFT_Nov_2021_TEMPLATE_SVI.ACMG_Specifications_InSiGHT_MMR_V1.pdf) from (http://priors.hci.utah.edu/PRIORS/BRCA/indexMLH1.php). Scores of >0.68 & indicate pathogenicity whereas a score of <0.11 is indicative of benignity. All variants identified in complementing orthologues demonstrates scores of <0.11 showing 100% concordance with such predictions.

To further validate the methodology, a selection of variants identified in complementing orthologues were introduced into the human protein to demonstrate their functionality. The N-terminus of MLH1 is responsible for its ATPase activity and MutSα binding capacity (Wu et al., 2015, Plotz et al., 2003, DuPrie et al., 2022). Previous studies have shown that several pathogenic missense mutations exist within these highly conserved regions (Wu et al., 2015, Thompson et al., 2014, Hardt et al., 2011, Ellison et al., 2004). Thus, VUS within this region were selected for validation, specifically, S46N, V185I and P208S. In line with

recommendations by the InSIGHT and ClinVar (ClinGen) committees, the pathogenic T82I and V185G variants were also introduced as controls (Thompson et al., 2014, Richards et al., 2015). Both T82I and V185G are classified as pathogenic by the InSiGHT expert panel (Thompson et al., 2014), demonstrate MAPP + PolyPhen2 scores of >0.11 and the pathogenic mechanism of these mutations is not attributed to pre-mRNA splicing defects (Raevaara et al., 2005, Shimodaira et al., 1998, Takahashi et al., 2007, Trojan et al., 2002, Kondo et al., 2003, Hinrichsen et al., 2013, Bouvet et al., 2019).

In addition to the N-terminus, the C-terminus of MLH1 is also highly conserved and is responsible for MLH1 binding to MutL homologues such as PMS2 (Plotz et al., 2006, Dai et al., 2021, Gueneau et al., 2013). Concomitantly, multiple studies have demonstrated the pathogenicity of variants in this region (Guerrette et al., 1999, Blasi et al., 2006, Takahashi et al., 2007, Kosinski et al., 2010, Thompson et al., 2014). Here, the A619T VUS was selected for further validation due to its presence in chicken MLH1 **Table 6-9**. Extensive variation has been reported at the A619 residue with A619P, A619V, A619D, A619T and A619S all submitted to ClinVar. A619P is reported as 'likely pathogenic' whereas the remaining variants are classified as VUS. Resolution of one variant at this residue would demonstrate that benign variation can exist at this position, dependent on the variant present.

Notably, variation was observed at residues where pathogenic variants are known to occur **Table 6-7.** One example, mentioned previously, is the MLH1 VUS V185, with V185I present in mouse and chicken MLH1 **Table 6-9** but V185G classified as pathogenic with expert panel review status. Biochemically, valine is highly similar to isoleucine with both containing hydrophobic side chains differing by only one CH_3 group. In contrast, glycine does not contain a hydrophobic side chain and is thus biochemically different to valine. Therefore, it is likely that a hydrophobic residue is required at this position. In addition to V185I, 2 additional variants are present at positions of known pathogenic variation, namely S247T and A681S

Table 6-7. S247P is the pathogenic residue at S247 and is known to disrupt protein localisation and MMR activity (Raevaara et al., 2005, Kansikas et al., 2011, Borràs et al., 2012). There is a marked biochemical difference between serine and proline. Proline's side chain comprises a pyrrolidine loop, with its distinctive cyclic structure giving proline high conformational rigidity in comparison to other amino acids (Gray et al., 2017). This results in proline restricting the conformation of the polypeptide. Furthermore, proline does not contain the amide necessary for hydrogen bonding. Because of this, proline has been shown to be the most frequently deleterious amino acid variant (Gray et al., 2017). In contrast, threonine is highly similar to serine, with the only difference being that the hydroxyl group is attached to the penultimate rather than terminal carbon atom. Notably, S247A is present homozygously in one individual catalogued on the gnomAD database, suggesting additional benign variation can exist at this position.

A681T is the pathogenic amino acid variant present at A681. Here, A681S was identified in chicken MLH1 **Table 6-7** and both serine and threonine show highly similar biochemical properties. A681T is classified as pathogenic by the InSiGHT expert panel (Thompson et al., 2014) and has been shown to disrupt binding to PMS2 and MMR activity in some studies (Takahashi et al., 2007, Shimodaira et al., 1998, Hardt et al., 2011, Bouvet et al., 2019, Guerrette et al., 1999, Kondo et al., 2003, Barnetson et al., 2008, Wanat et al., 2007). However some studies have demonstrated proficient MMR activity for this variant (Raevaara et al., 2005, Hinrichsen et al., 2013). Therefore, assessment of A681T and A681S in the human MLH1 protein would do two things. Firstly, it would demonstrate the acute sensitivity of the method showing that amino acids with highly similar biochemical properties can have marked differences in their effect on protein function. Secondly, it would provide further evidence regarding the effect of A681T on MLH1 protein function. Therefore, both A681S and A681T were introduced into the human MLH1 protein for further assessment.

ClinVar Accession number	Residue position	WT	Variant	Variant present in	ClinVar classification	Encoded across an exon boundary?	MAPP + PolyPhen-2 score	Published functional data?
NM_000249.4(MLH1):c.4T>G (p.Ser2Ala)	2	S	A	Mouse, Chicken	Uncertain significance	No	0.00 (benign)	Identified in 1 CRC patient with loss of MLH1 and PMS2 proteins (Jiang et al., 2019)
NM_000249.4(MLH1):c.7T>C (p.Phe3Leu)	3	F	L	Chicken	Conflicting interpretations of pathogenicity	No	0.00 (benign)	None
NM_000249.4(MLH1):c.40A>G (p.Thr14Ala)	14	T	A	Chicken	Uncertain significance	No	0.01 (benign)	None
NM_000249.4(MLH1):c.137G>A (p.Ser46Asn)	46	S	N	Mouse	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.148A>G (p.Ile50Val)	50	I	V	Mouse, Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.181A>G (p.Ile61Val)	61	I	V	Chicken	Uncertain significance	No	0.01 (benign)	None
NM_000249.4(MLH1):c.343A>G (p.Ile115Val)	115	I	V	Chicken	Uncertain significance	No	0.02 (benign)	None
NM_000249.4(MLH1):c.452C>T (p.Thr151Met)	151	T	M	Chicken	Uncertain significance	Yes	0.03 (benign)	None
NM_000249.4(MLH1):c.475A>G (p.Ile159Val)	159	I	V	Chicken	Uncertain significance	No	0.00 (benign)	None

NM_000249.4(MLH1):c.553G>A (p.Val185Ile)	185	V	I	Mouse, Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.622C>T (p.Pro208Ser)	208	P	S	Chicken	Uncertain significance	No	0.00 (benign)	Identified in 1 patient with T-cell ALL (Richter-Pechańska et al., 2017) and in breast cancer patients and controls (Consortium, 2021)
NM_000249.4(MLH1):c.808A>G (p.Thr270Ala)	270	T	A	Mouse	Conflicting interpretations of pathogenicity	No	0.05 (benign)	None
NM_000249.4(MLH1):c.960G>C (p.Glu320Asp)	320	E	D	Chicken	Conflicting interpretations of pathogenicity	No	0.00 (benign)	No effect on splicing (Tournier et al., 2008)
NM_000249.4(MLH1):c.988A>G (p.Ile330Val)	330	I	V	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1063C>T (p.Pro355Ser)	355	P	S	Chicken	Uncertain significance	No	0.00 (benign)	None

NM_000249.4(MLH1):c.1075A>G (p.Met359Val)	359	M	V	Chicken	Likely benign	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1087A>G (p.Thr363Ala)	363	T	A	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1090A>G (p.Thr364Ala)	364	T	A	Chicken	Uncertain significance	No	0.00 (benign)	Identified in MSI-H breast cancer in which MLH1 expression was lost (Curia et al., 1999). Identified in LS patients but considered non-pathogenic (Lellis et al., 2013)
NM_000249.4(MLH1):c.1096C>G (p.Leu366Val)	366	L	V	Mouse	Uncertain significance	No	0.00 (benign)	None

NM_000249.4(MLH1):c.1111A>G (p.Thr371Ala)	371	T	A	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1123A>G (p.Ser375Gly)	375	S	G	Mouse	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1141C>T (p.His381Tyr)	381	H	Y	Mouse	Uncertain significance	No	0.01 (benign)	Reported in an individual with early onset gastric cancer (Bacani et al., 2005). Does not impact MLH1 stability or PMS2 binding in a Y2H assay (Wang et al., 2012)
NM_000249.4(MLH1):c.1202G>A (p.Ser401Asn)	401	S	N	Chicken	Uncertain significance	No	0.00 (benign)	Identified in individuals with colorectal cancer (Pearlman et al., 2019,

								Yurgelun et al., 2017)
NM_000249.4(MLH1):c.1238C>G (p.Thr413Arg)	413	T	R	Mouse	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1253A>G (p.Asp418Gly)	418	D	G	Chicken	Uncertain significance	No	0.00 (benign)	Observed in an individual with CRC (Yurgelun et al., 2017)
NM_000249.4(MLH1):c.1268G>C (p.Arg423Thr)	423	R	T	Chicken	Uncertain significance	No	0.00 (benign)	WT expression levels when expressed in HEK293, and restored sensitivity to 5-FU and MMS in HCT116 to the same extent as WT (Arora et al., 2017)

NM_000249.4(MLH1):c.1271C>T (p.Ala424Val)	424	A	V	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1274G>C (p.Arg425Thr)	425	R	T	Mouse	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1277A>C (p.Gln426Pro)	426	Q	P	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1286A>C (p.Glu429Ala)	429	E	A	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1319T>C (p.Val440Ala)	440	V	A	Mouse	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1325C>A (p.Ala442Asp)	442	A	D	Chicken	Uncertain significance	No	0.001 (benign)	None
NM_000249.4(MLH1):c.1331A>G (p.Asn444Ser)	444	N	S	Mouse	Conflicting interpretations of pathogenicity	No	0.00 (benign)	Reported in one individual with colorectal cancer displaying MSI and loss of MLH1 but classified as a VUS (Rodríguez–Soler et al., 2013)

NM_000249.4(MLH1):c.1345G>A (p.Gly449Arg)	449	G	R	Mouse	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1351A>G (p.Thr451Ala)	451	T	A	Chicken	Uncertain significance	No	0.01 (benign)	None
NM_000249.4(MLH1):c.1378G>C (p.Glu460Gln)	460	E	Q	Mouse	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1403A>G (p.Asn468Ser)	468	N	S	Mouse	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1541A>G (p.Asn514Ser)	514	N	S	Mouse, chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1547A>G (p.Gln516Arg)	516	Q	R	Mouse	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1565G>A (p.Arg522Gln)	522	R	Q	Chicken	Conflicting interpretations of pathogenicity	No	0.00 (benign)	Heterozygous in patients with CRC and breast cancer (Niessen et al., 2006, Yurgelun et al., 2015) but also identified in healthy controls (Consortium, 2021). Colorectal cancer patients did

								not exhibit loss of MLH1 protein expression (Niessen et al., 2006)
NM_000249.4(MLH1):c.1577A>G (p.His526Arg)	526	H	R	Mouse, chicken	Conflicting interpretations of pathogenicity	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1579A>G (p.Asn527Asp)	527	N	D	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1622C>T (p.Ala541Val)	541	A	V	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1752C>G (p.Asp584Glu)	584	D	E	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1855G>A (p.Ala619Thr)	619	A	T	Chicken	Uncertain significance	No	0.01 (benign)	None
NM_000249.4(MLH1):c.1910T>C (p.Ile637Thr)	637	I	T	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1934A>G (p.Asn645Ser)	645	N	S	Mouse	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.1946C>T (p.Pro649Leu)	649	P	L	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.2024G>C (p.Ser675Thr)	675	S	T	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.2087C>A (p.Thr696Asn)	696	T	N	Chicken	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.2110G>A (p.Val704Met)	704	V	M	Mouse	Uncertain significance	No	0.00 (benign)	None

NM_000249.4(MLH1):c.2123T>C (p.Ile708Thr)	708	I	T	Mouse	Conflicting interpretations of pathogenicity	No	0.00 (benign)	None
NM_000249.4(MLH1):c.2130C>G (p.Asn710Lys)	710	N	K	Mouse	Uncertain significance	No	0.00 (benign)	None
NM_000249.4(MLH1):c.2131T>C (p.Ser711Pro)	711	S	P	Mouse, chicken	Conflicting interpretations of pathogenicity	No	0.00 (benign)	None
NM_000249.4(MLH1):c.2155A>G (p.Ile719Val)	719	I	V	Chicken	Conflicting interpretations of pathogenicity	No	0.00 (benign)	None

Table 6-9: **Results of searching the sequence alignment of complementing MLH1 orthologues for ClinVar variants without 'benign' or 'pathogenic' classifications.**

The sequence alignment of human, mouse and chicken MLH1 was searched for ClinVar variants in MLH1 without classifications of 'benign' or 'pathogenic'. 59 variants were identified as being present in mouse and/or chicken MLH1. These are displayed with accession number, residue position and ClinVar classification and if the variant is encoded across an exon boundary.

6.3.8 MLH1 constructs containing variants present in complementing MLH1 orthologues can restore 6-TG sensitivity

To further prove that variants present in the sequence alignment of complementing orthologues are tolerated by the human protein, I functionally tested the variants described above. The variants selected were introduced into the N-terminally GFP-tagged human MLH1 protein **Figure 6-9** via site-directed mutagenesis and cloned into pLenti-PGK. Lentiviruses were produced and HAP1 Δ*MLH1* cells were then transduced with these lentiviruses and GFP+ cells were sorted via FACS and seeded as single cells **Figure 10-8**. Cells were then treated with varying doses of 6-TG for 24 hours. Colony formation was then assessed 1 week after FACS. MLH1 constructs containing any of S46N, V185I, P208S, A619T and A681S were all able to restore 6-TG sensitivity to the same extent as WT MLH1, **Figure 6-10**, **Figure 6-11.** In contrast, MLH1 constructs containing the pathogenic controls T82I and V185G showed no restoration of 6-TG sensitivity, **Figure 6-10**, **Figure 6-11.** Importantly, A681T resulted in a mild, but statistically significant reduction in the ability of MLH1 to restore 6-TG sensitivity, **Figure 6-11.** This validates that variants in complementing orthologues can be tolerated by the human protein in a functional *in vivo* MMR assay.

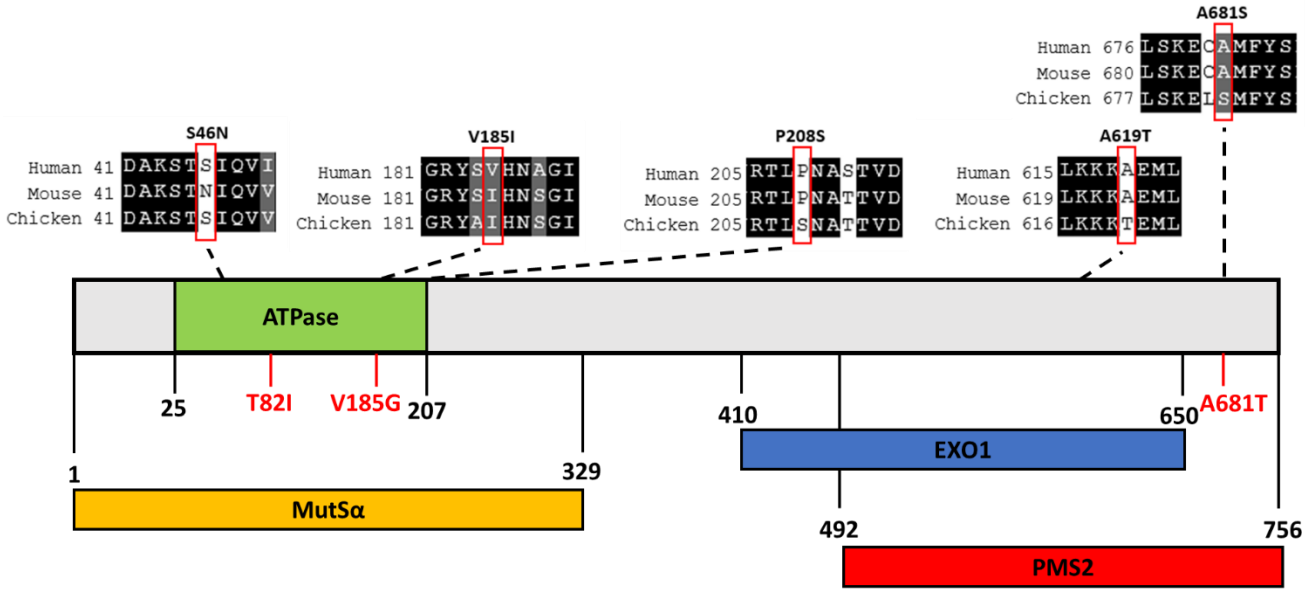


Figure 6-9: Schematic showing variants introduced into human MLH1

Schematic showing the variants introduced into MLH1 with the corresponding sequence alignment also displayed. The binding domains of MLH1 and the associated binding partner are also indicated with amino acid residue numbers displayed. Pathogenic variants introduced into the human protein as controls are indicated in red.

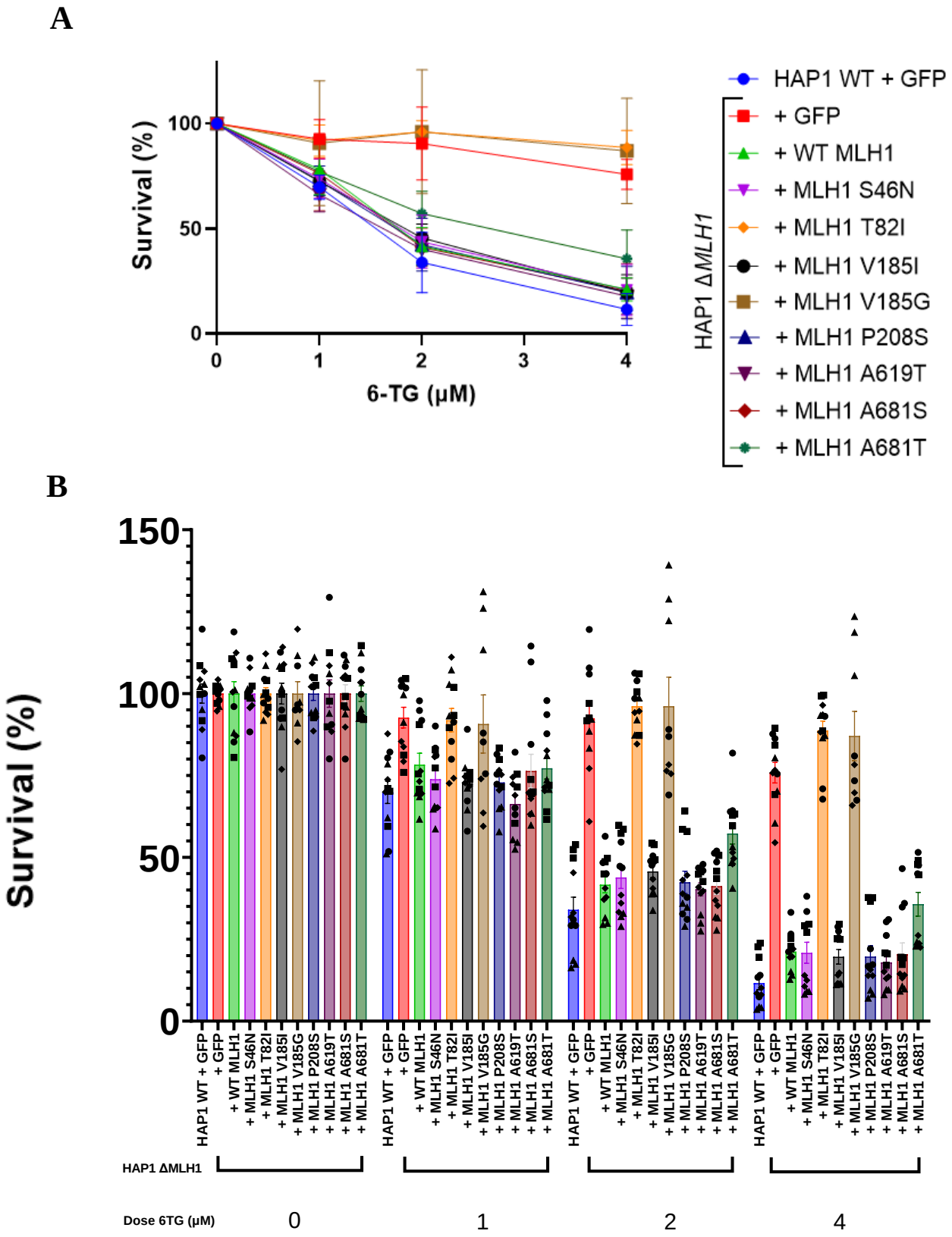


Figure 6-10: MLH1 constructs containing variants present in complementing orthologues can restore 6-TG sensitivity

(A) Survival curves demonstrating the colony survival % compared to untreated cells in response to increasing doses of 6-TG for HAP1 WT and HAP1 Δ*MLH1* cells complemented with WT or variant GFP-tagged MLH1 constructs or GFP only controls. Data is plotted as the mean of N=4. Error bars indicate SEM. **(B)** Bar chart displaying all technical data points from N=4. Error bars indicate SEM.

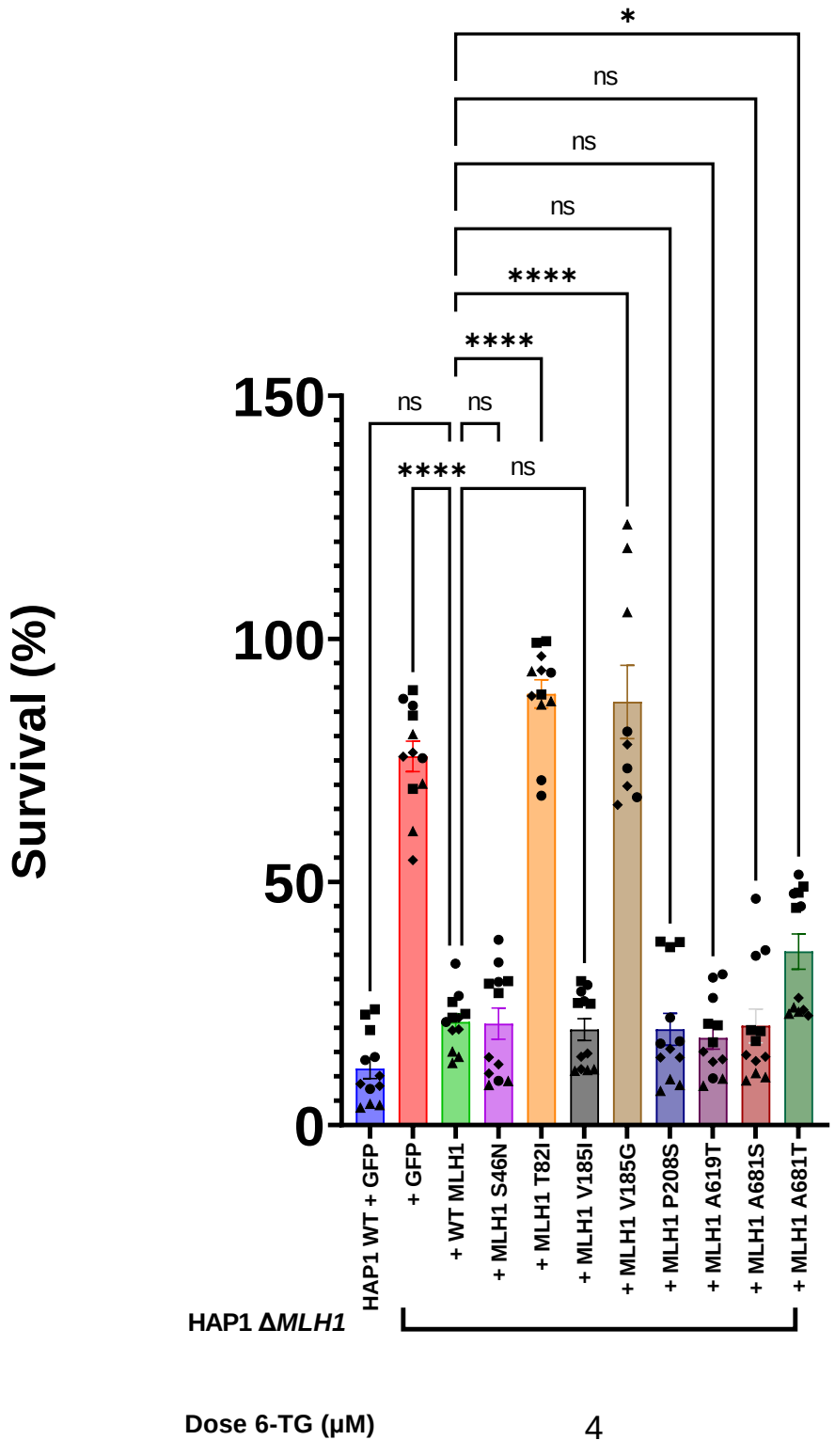


Figure 6-11: Statistical analysis of MLH1 variant effects in response to 4μM 6-TG

One-way ANOVA was performed on all values of N=4 at 4μM 6-TG and included a Dunnett's multiple comparisons test comparing HAP1 Δ*MLH1* + GFP-WT MLH1 to all other cell lines. Results from Dunnett's multiple comparisons test are displayed. ns = non-significant difference, **** = p <0.0001, * = p 0.0167.

6.3.9 MLH1 constructs containing variants present in complementing MLH1 orthologues can restore PMS2 stability

As the interaction between MLH1 and PMS2 is critical for PMS2 stability (Trojan et al., 2002), the expression level of MLH1 protein variants and their ability to restore PMS2 stability was assessed via western blot **Figure 6-12.** Importantly, HAP1 Δ*MLH1* cells complemented with MLH1 constructs containing variants present in complementing orthologues were able to restore PMS2 stability to the same extent as WT MLH1 **Figure 6-12**. Notably, A681T was expressed at similar levels to the WT protein but cells lines expressing this construct displayed significant reduction in PMS2 stability **Figure 6-12**. Immunoprecipitation experiments using lysates prepared from HAP1 Δ*MLH1* cells transduced with lentivirus encoding GFP-MLH1 A681S or A681T showed that whilst both variants were able to pull down PMS2, overall levels of PMS2 were significantly reduced in HAP1 Δ*MLH1* cells transduced with MLH1 A681T **Figure 6-13**. As the interaction between MLH1 and PMS2 is critical for PMS2 stability (Trojan et al., 2002) and A681T showed WT expression levels, the data suggest that A681T results in reduced binding to PMS2, reducing PMS2 stability, rather than a reduction in MLH1 stability causing loss of PMS2 stability. In contrast, A681S appears to show WT levels of binding to PMS2 and thus WT PMS2 levels are restored **Figure 6-13**.

Overall, the data presented demonstrate the ability of a sequence alignment of complementing orthologues to identify variants tolerated throughout the human MLH1 protein. Importantly, it is able to identify functionally tolerated variants at residues where pathogenic variation is known to occur.

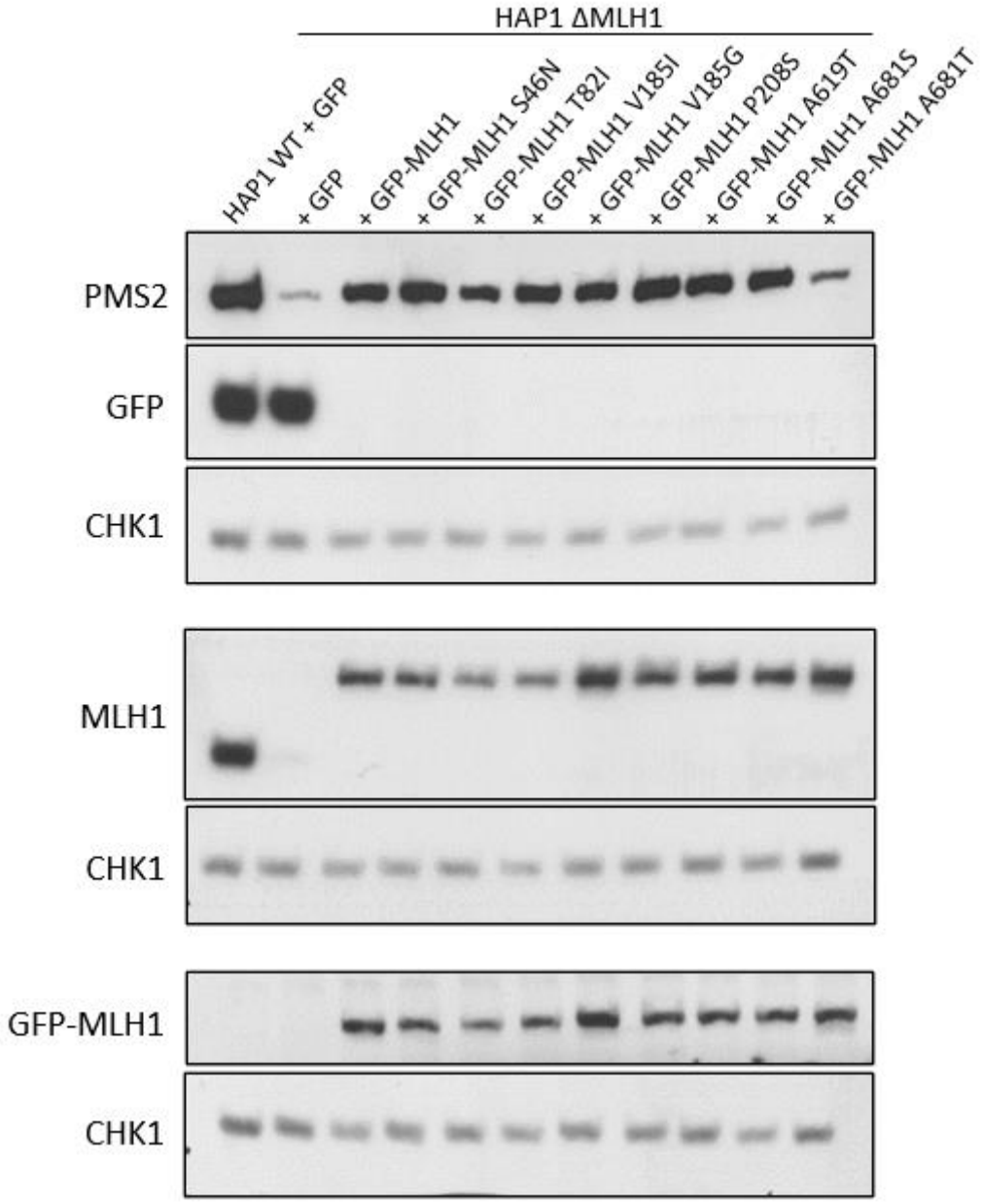


Figure 6-12: MLH1 constructs containing variants present in complementing orthologues can restore PMS2 stability

Western blot for PMS2 and GFP of lysates from HAP1 WT or HAP1 Δ*MLH1* cells constitutively expressing GFP, GFP-tagged WT or GFP-variant MLH1 constructs. CHK1 was used as a loading control. N=2.

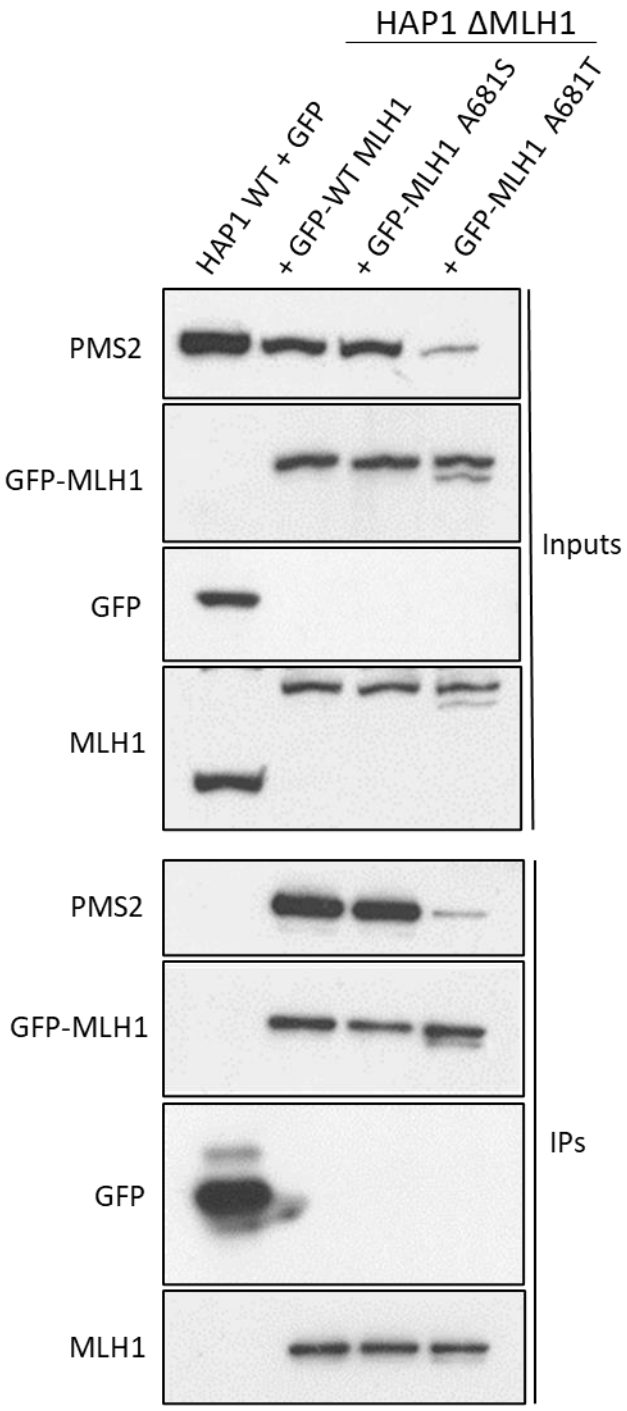


Figure 6-13: Immunoprecipitation of PMS2 with A681S and A681T MLH1

Western blot of a GFP-trap immunoprecipitation on lysates prepared from HAP1 WT or HAP1 Δ*MLH1* cells expressing GFP or GFP tagged WT, A681S or A681T MLH1. Probing for PMS2, GFP and MLH1. Inputs demonstrate overall protein levels in the lysate and IPs demonstrate the amount of protein pulled out by the respective MLH1 construct (or GFP control). N=1.

6.3.10 Prediction of requirement of PTMs reported in PhosphoSitePlus

My primary aim was to determine that the sequence alignment of complementing MLH1 orthologues could identify functional variants in the MLH1 protein. However, having previously shown the ability of the method to identify non-essential PTM sites, the sequence alignment of complementing MLH1 orthologues was searched for variation to non-modifiable residues at position of reported PTMs in the human MLH1 protein MLH1, catalogued on PhosphoSitePlus. This revealed 17 PTMs reported in human MLH1 present as non-modifiable residues in mouse and/or chicken MLH1 **Table 6-10**, suggesting that they are not required for MLH1 function. Only one such PTM had previously been assigned a functional role in MLH1. Specifically, it was claimed that activation of MLH1 in response to doxorubicin requires the presence of BRCA1, dependent on the ATM/ATR-mediated phosphorylation of MLH1 on S406 (Romeo et al., 2011). However, as explained previously, the authors provided no evidence of defects of MLH1 activity, i.e. survival in response to 6-TG or endonuclease activity of MutL complexes. Furthermore, doxorubicin does not induce mismatches, rather DSBs via topoisomerase II inhibition (Thorn et al., 2011). Importantly, loss of BRCA1 is not known to give rise to an MSI phenotype (Sokol et al., 2022, Looij et al., 2001). Therefore, this and the data presented here suggest that phosphorylation of this residue is not required for MLH1 function in response to 6-TG or in stabilising PMS2. All other PTMs which have diverged in complementing orthologues were all identified via high-throughput methodologies and thus lack any assigned functional role in MLH1. MLH1 is acetylated on K33, K241, K361 and K377, with deacetylation of these residues by HDAC6 blocking MutLα-MutSα complex formation and conferring 6-TG resistance (Zhang et al., 2019b). Notably, all were conserved to a lysine - or a residue which can be acetylated - in complementing orthologues **Figure 6-8**, supporting these findings. To my knowledge, no additional functional PTMs have been

reported in the MLH1 protein. This and the data presented here suggest that regulation via PTMs is not a prominent feature of MLH1 in response to 6-TG and PMS2 stabilisation.

Position	Residue	Modification	Residue in mouse	Residue in chicken	Encoded across an exon boundary?	Proposed function
2	S	Phosphorylation	A	A	No	None
87	S	Phosphorylation	T	K	No	None
236	K	Ubiquitination	K	A	No	None
370	S	Phosphorylation	S	A	No	None
402	K	Acetylation	S	N	No	None
402	K	Ubiquitination	S	N	No	None
406	S	Phosphorylation	S	A	No	Required for BRCA1-mediated activation of MLH1 in response to doxorubicin (Romeo et al., 2011)
416	K	SUMOylation	R	K	No	None
443	K	Acetylation	E	L	No	None
443	K	Ubiquitination	E	L	No	None
446	S	Phosphorylation	N	D	No	None
453	K	Ubiquitination	M	M	No	None
459	S	Phosphorylation	A	S	No	None
461	K	Acetylation	K	H	No	None
467	S	Phosphorylation	S	A	No	None
486	S	Phosphorylation	S	N	No	None

Table 6-10: Results from searching the sequence alignment of complementing MLH1 orthologues for variation to non-modifiable residues at reported PTM sites.

6.4 Discussion

6.4.1 Summary of the work presented in this chapter

In this chapter the complementation of human cells with orthologues of mouse and chicken MLH1 was demonstrated via their ability to bind to known interactors PMS2, MSH6 and EXO1 **Figure 6-2** and their ability to restore 6-TG sensitivity **Figure 6-4**, **Figure 6-6** and PMS2 stability in two MLH1 deficient human cell lines **Figure 6-5**, **Figure 6-7**. In contrast, an MLH1 orthologue from frog could not complement human cells. It showed reduced affinity for EXO1 **Figure 6-2**, and reduced 6-TG sensitivity restoration in two MLH1 deficient human cell lines **Figure 6-4**, **Figure 6-6**. Comparison of the sequence alignment of complementing orthologues to the gnomAD database did not reveal any variants present at frequencies of > 0.01% existed in any of the complementing orthologues **Table 6-3, Table 6-4.** Furthermore, no existing benign variants on the ClinVar database could be identified in complementing orthologues **Table 6-6.** Importantly,, no pathogenic ClinVar variants were found in complementing orthologues **Table 6-7.** VUS present in complementing orthologues were shown to be functional in the human protein evidenced by restoration of PMS2 stability **Figure 6-12**, **Figure 6-13** and 6-TG sensitivity **Figure 6-10**, **Figure 6-11** in an MLH1-deficient human cell line. This indicates functionality of almost 60 MLH1 variants with uncertain classifications on ClinVar due to their presence in complementing orthologues **Table 6-9**. Importantly, functional variation was demonstrated *in vivo* at sites of known pathogenic variants using PMS2 restoration and 6-TG sensitivity assays **Figure 6-10**, **Figure 6-11, Figure 6-12**, **Figure 6-13**. Additionally, analysis of PTMs reported for MLH1 on PhosphoSitePlus revealed several PTMs which were present as non-modifiable residues in complementing orthologues **Table 6-10.** Overall it has been shown that orthologue complementation can be applied to a protein

with high clinical significance and thus provides a database for functional variation in the human MLH1 protein.

6.4.2 A discussion of the findings presented in this chapter

Previous examples of orthologue complementation using the MLH1 protein mostly revolve around the complementation of yeast with human MLH1 cDNA or genes. Human MLH1 is unable to restore yeast MMR and has a dominant negative effect on the yeast MMR system when expressed ectopically (Shimodaira et al., 1998, Takahashi et al., 2007). However, replacement of the entire genes of both *mlh1* and *pms1* but not *mlh1* alone in yeast with their human equivalents restores MMR to WT levels (Vogelsang et al., 2009). It appears that different regions of the MLH1 protein are more tolerant to change than other areas. Human-yeast fusion proteins consisting of N-terminal human fused to C-terminal yeast MLH1 can complement yeast, whereas the inverse cannot (Ellison et al., 2001). To my knowledge no other examples of MLH1 orthologue complementation exist. However, previous work shows that introduction of variants into the mouse MLH1 protein is an effective way of studying their impact in the human protein (Houlleberghs et al., 2020). Here, the mouse and chicken MLH1 orthologues could complement human cells by binding to known interactors MSH6, EXO1 and PMS2 **Figure 6-2**, restoring PMS2 stability **Figure 6-3**, **Figure 6-5**, **Figure 6-7** and sensitivity to 6-TG **Figure 6-4**, **Figure 6-6** expression in 2 MLH1-deficient cell lines. In contrast, frog MLH1 showed reduced affinity for EXO1 **Figure 6-2** and failed to restore PMS2 stability in HCT116 and 6-TG sensitivity in both HCT116 and HAP1 *ΔMLH1* **Figure 6-4**, **Figure 6-6**. However, in the case of the IP and restoration of PMS2 experiments, the lack of quantitative assessment of these experiments presents a clear limitation of the work. Differences in expression of the MLH1 orthologues were observed in both cases. Thus greater clarity regarding the proportion of MLH1 orthologue present to amount of protein pulled down and/or

PMS2 level restored would enable stronger conclusions to be drawn. Techniques described in section 4.4.2 could be used to better understand the findings presented in this chapter.

The loss of complementation observed with frog MLH1 could be assigned to multiple factors. Firstly, loss of interaction with EXO1 could be a way in which the partial complementation of frog MLH1 arose. EXO1 participates in the repair of single-base mismatches and single-base IDLs and *Exo1*-deficient mice do not show the same mutation rates or cancer phenotypes compared to *Mlh1*-deficient mice (Wei et al., 2003, Prolla et al., 1998). Furthermore, *Exo1*-deficient mice show higher 50% survival rates (17 months) than *Mlh1*-deficient mice (6 months) (Wei et al., 2003, Prolla et al., 1998). Furthermore, deletion of *exo1* only causes a 2-4 fold increase in mutation rate compared to *msh2*-null yeast models (Tishkoff et al., 1997). Additionally, EXO1-null cells display lower mutations rates than MLH1 or MSH2-null cells (Zou et al., 2021). Furthermore, the requirement of EXO1 in post-replicative MMR is not essential and EXO1-independent MMR pathways are known (Goellner et al., 2015). However, deletion of *exo1* increases the percentage of cells displaying Mlh1-Pms1 foci, suggesting that the Exo1 based pathway is faster *in vivo*, and thus more frequently used (Hombauer et al., 2011). The reason behind loss of interaction is challenging to specify as the region of MLH1 required for this interaction (AA 410-650, (Schmutte et al., 2001) shows varying degrees of conservation **Figure 6-1**. Collectively, loss of human EXO1 binding by frog MLH1 may result in a partial MMR defect due to EXO1 independent MMR pathways and lack of requirement for EXO1 for repair of all types of mismatch (Goellner et al., 2015).

Frog MLH1 could bind to PMS2 in immunoprecipitation-based assays **Figure 6-2**, and was able to restore PMS2 stability in 293T and HAP1 Δ*MLH1* cell lines **Figure 6-3**, **Figure 6-5**. However, it failed to restore PMS2 stability in HCT116 cells **Figure 6-7**. This is likely due to lower expression of MLH1 in the HCT116 compared to HAP1 Δ*MLH1*. Again, quantitative assessment of MLH1 orthologue expression vs PMS2 restoration in this cell line would provide

further clarity around this conclusion. Furthermore, frog MLH1 showed greater restoration of 6-TG sensitivity in HAP1 Δ*MLH1* compared to HCT116 **Figure 6-4**, **Figure 6-6**. Therefore, in HAP1 Δ*MLH1*, the defect in 6-TG sensitivity is likely due to other factors asides from PMS2 stability restoration, such as loss of EXO1 interaction. However, in HCT116, the greater defect my arise due to partial loss of PMS2 stability in combination with loss of EXO1 binding. However, despite the likelihood that the partial complementation with frog MLH1 might be due to loss of EXO1 interaction, this cannot be said for certain due to variation existing at other residues outside of the EXO1-interaction domain.

In contrast to XPA and XRCC4, gnomAD variants present homozygously in at least one individual were not present in any of the complementing orthologues **Table 6-3**. Expansion of the analysis to variants present at frequencies >0.01% regardless of zygosity, in line with InSiGHT recommendations (Thompson et al., 2014), also failed to reveal any such variants in complementing orthologues. However, this step is only performed as an initial assessment of whether the alignment contains known benign variants and does not negate using a sequence alignment of complementing orthologues to detect functional amino acid variants in the human protein. Variants in gnomAD were detected in complementing orthologues **Table 6-5**, but all were present at frequencies below 0.004% and are thus not classified as benign according to INSiGHT criteria (Thompson et al., 2014, Richards et al., 2015). However, their presence in complementing orthologues demonstrates that they are functional in the human protein. Comparison of the sequence alignment to the ClinVar database did not reveal any variants classified as benign as being present in complementing orthologues **Table 6-6**. Notably, 90% of the residues at which benign variants are found were highly conserved in complementing orthologues **Table 6-6**. In comparison, 96% of the residues where pathogenic variants occur showed conservation in complementing orthologues **Table 6-7**. Importantly, no pathogenic variants were found in complementing orthologues **Table 6-7**. Overall, this demonstrates that

both benign and pathogenic variation can occur at highly conserved residues, meaning that caution is advised when solely using *in silico* variant-prediction techniques solely based on evolutionary conservation. This highlights the need to determine which orthologues are functional in human cells to enable accurate identification of benign variants.

59 ClinVar MLH1 variants without classifications as benign or pathogenic are present in complementing orthologues **Table 6-9**. The N-terminal variants S46N and V185I lie within the ATPase domain of MLH1 (AA 25-207, (Wu et al., 2015) and mutations which disrupt MLH1 ATPase activity are known to impair sensitivity to 6-TG (Johnson et al., 2010). S46N and V185I MLH1 could both restore 6-TG sensitivity to the same extent as WT MLH1, **Figure 6-11**, and therefore demonstrate functionality in the human protein. In contrast, the pathogenic T82I and V185G MLH1 variants showed significantly impaired ability to restore 6-TG sensitivity, **Figure 6-11**. This is in line with previous studies showing an inability of these variants to restore MMR capacity in human cells, yeast and *in vitro* (Raevaara et al., 2005, Shimodaira et al., 1998, Takahashi et al., 2007, Trojan et al., 2002, Kondo et al., 2003, Bouvet et al., 2019, Hinrichsen et al., 2013). Furthermore, S46N, T82I, V185I and V185G were all able to restore PMS2 stability in human cells **Figure 6-12**. This is expected as all variants lie outside of the PMS2-binding domain of MLH1 (Guerrette et al., 1999). The finding that V185I but not V185G MLH1 restores 6-TG sensitivity reveals that functional variation can exist at sites of known pathogenic variants in the MLH1 protein. As explained previously, valine is highly biochemically similar to isoleucine with both containing hydrophobic side chains differing by only one CH_3 group. In contrast, glycine does not contain a hydrophobic side chain and is thus biochemically different to valine. V185 forms part of an alpha helix (Wu et al., 2015, Jumper et al., 2021) and glycine residues are known to frequently disrupt alpha helix formation (O'Neil and DeGrado, 1990).

P208S lies just outside of the putative ATPase domain of human MLH1 (Wu et al., 2015). Multiple submitters had identified this variant in clinical testing with all giving 'uncertain significance' classifications citing lack of direct functional evidence (https://www.ncbi.nlm.nih.gov/clinvar/variation/141121/?new_evidence=false). Introduction of this variant into the human MLH1 protein had no impact on its ability to restore sensitivity to 6-TG, **Figure 6-11** or stabilise PMS2 **Figure 6-12**, providing functional evidence that this variant is tolerated by the human protein. To my knowledge, no prior assessment of this variant or any other variants at this position have been performed. However, it was identified in one individual with T-cell acute lymphoblastic leukaemia, but no functional analysis was performed with this variant meaning its contribution to the phenotype is unknown (Richter-Pechańska et al., 2017). Additionally, this variant was identified in breast cancer patients, but also in healthy controls, suggesting that it is a passenger mutation in this case (Consortium, 2021). Overall, it is clearly tolerated by the human MLH1 protein in the assays performed here. It must be noted that P208R and P208L are also catalogued on ClinVar as VUS. However, it cannot be said whether these variants are functional as neither are present in any of the complementing orthologues used here.

A619T lies within the region of MLH1 required for interaction with PMS2 and the other MutL family members MLH3 and PMS1 (Kondo et al., 2001). A619 appears to experience significant variation in individuals within the clinic with A619S, A619T, A619V, A619D and A619P all catalogued in ClinVar after identification in at least one individual. Here, A619T was identified in chicken MLH1 and was shown to restore PMS2 stability and 6-TG sensitivity, **Figure 6-11**, **Figure 6-12**. No other variation existed in complementing orthologues at this position, and therefore I cannot offer insight into other functional variation at this residue. Interestingly, A619P is classified as 'likely pathogenic' after being identified in a German family meeting the Amsterdam II criteria (Hardt et al., 2011). Furthermore, A619P shows only 37% MMR

activity compared to WT MLH1 and loss of interaction with PMS2 (Drost et al., 2019). However, it must be re-iterated that proline is biochemically distinct from threonine and that proline is the most frequently deleterious amino acid variant due to its conformational rigidity (Gray et al., 2017). Evidently, this residue is important for PMS2 interaction and certain variants at this position disrupt this. Here, A619T does not show any of these defects highlighting again that functional variation can exist at sites known to be important for protein function if present in a complementing orthologue.

Lastly, orthologue complementation revealed that A681S is tolerated by human MLH1, in contrast to the pathogenic A681T variant at this position. To date, the literature surrounding the impact of A681T on MLH1 protein function is mixed. Some studies have shown that this variant has proficient MMR function *in vitro* (Raevaara et al., 2005, Hinrichsen et al., 2013), but others have shown moderately reduced *in vitro* activity (Takahashi et al., 2007, Drost et al., 2019), and a reduced dominant mutator effect in yeast (Shimodaira et al., 1998, Hardt et al., 2011, Wanat et al., 2007). This variant is known to co-segregate with disease in multiple LS families (Barnetson et al., 2008, Froggatt et al., 1996) and tumours from these patients display high levels of MSI (Raevaara et al., 2005, Barnetson et al., 2008). Importantly, this variant disrupts MMR *in vivo*, assessed via transfection of HCT116, followed by exposure to MNNG and assessment of colony formation (Bouvet et al., 2019), highly similar to the methodology utilised here. Here, the A681T variant significantly disrupted sensitivity to 6-TG *in vivo*, **Figure 6-11**. The reasons underlying the discrepancies of functional impact of A681T are difficult to narrow down. However, the *in vivo* evidence from functional studies and patient samples, some *in vitro* and yeast studies and the data presented here collectively suggests moderately disrupted MMR capacity for this variant. In addition to disrupted MMR activity, MLH1 deficient cell lines stably expressing A681T could not restore PMS2 stability to the same extent as WT **Figure 6-12**, **Figure 6-13**. This is concordant with GST studies and some

Y2H based studies showing that this variant disrupts interaction with PMS2 (Guerrette et al., 1999, Kondo et al., 2003, Hardt et al., 2011, Kondo et al., 2001). However, A681T has been shown to bind to PMS2 in one Y2H screen (Drost et al., 2019) and restores PMS2 stability in 293T cells (Raevaara et al., 2005). Again, the reason underlying the discrepancy in results is challenging to pinpoint, especially as two separate Y2H screens showed discordant results. Structurally, A681 has been implicated in the MutL homologue dimerisation domain of MLH1 by some studies (Pang et al., 1997, Guerrette et al., 1999, Kondo et al., 2001, Dai et al., 2021) but excluded by others (Kosinski et al., 2010, Cutalo et al., 2006, Gueneau et al., 2013). A681 forms part of an alpha helix in Mlh1 which was recently shown to mediate the interaction between Mlh1 and Mlh3 (Dai et al., 2021). Importantly, sequence alignments of MLH1 orthologues ranging from human to yeast revealed that this residue was present as only an alanine or serine, suggesting minimal variation is tolerated at this position **Figure 10-9**. Indeed, here it was shown that A681S has no impact on MLH1's ability to sensitise to 6-TG, **Figure 6-11** or on its ability to interact with PMS2 and restore PMS2 stability **Figure 6-12**, **Figure 6-13**. Biochemically, alanine is a hydrophobic amino acid with a CH_3 side chain. In contrast, threonine is a polar amino acid with its side chain carbon bonded to a hydrogen, hydroxyl group and a CH_3 group. Serine is biochemically similar to threonine being a polar amino acid but its side chain carbon is bonded to two hydrogen and a hydroxyl group, meaning it does not contain a CH_3 group on its side chain. To date, one unpublished crystal structure of the C-terminus of MLH1 (residues 486-751) has been produced (PDB Entry – 3RBN) and AlphaFold predictions of its structure also exist (Jumper et al., 2021). Assessment of the AlphaFold structure of MLH1 reveals that A681 forms part of an α helix with the N and C-termini of this residue forming hydrogen bonds with S677 and S685 respectively (PDB Entry – 3RBN), (Jumper et al., 2021) **Figure 10-10** However, the side chain of this residue does not appear to make contact with surrounding residues **Figure 10-10.** AlphaFold is not yet effective at accurately modelling the

effect of individual amino acid variants on protein structure (Pak et al., 2021) (https://alphafold.ebi.ac.uk/faq) and thus it is challenging to determine what impact alteration of the side chain at this position would have on protein structure. Some studies have demonstrated that A681T results in loss of MLH1 protein stability (Hinrichsen et al., 2013, Hardt et al., 2011), whereas others have shown no reduction in protein stability (Raevaara et al., 2005, Xie et al., 2010). Here, A681T does not affect MLH1 protein stability **Figure 6-12**, **Figure 6-13**. Notably, normal MLH1 protein levels have been observed in some CRCs carrying this mutation (Alvarez et al., 2010). Furthermore, whilst loss of MLH1 results in loss of PMS2, the opposite is not always true. Indeed, there are multiple cases whereby IHC of tumour tissue reveals that PMS2 expression is lost but MLH1 is still detectable (Dudley et al., 2015, Jong et al., 2004, Kato et al., 2016). Therefore the likely mechanism of action of A681T is to disrupt binding to PMS2 without destabilising the MLH1 protein, resulting in a moderate reduction in MMR activity. In contrast, A681S clearly permits PMS2 binding, but the detailed structural mechanisms behind this remain unclear. It must be reiterated that all experiments assessing the effect of MLH1 variants on PMS2 restoration and PMS2 binding performed in this chapter lack any quantitative assessment of protein expression/binding. Therefore, it remains possible that differences in expression of each MLH1 variant construct may underlie the levels of PMS2 restoration/binding observed. Assays described in section 4.4.2 could be utilized to strengthen the conclusions presented here.

The ACMG/AMP and latest InSiGHT guidelines on variant interpretation dictate that identifying a variant at the site of a known pathogenic variant can be considered as moderate evidence of pathogenicity (PM5) (Richards et al., 2015, Thompson et al., 2014). The evidence presented here suggests that this rule should be treated with caution as it has been shown that functional variation can exist at site of known pathogenic variants in three separate cases, dependent on the variant present.

In addition to the variants directly tested in the human protein, many more are present in complementing orthologues and are thus tolerated by the human protein **Table 6-9**. In most cases, no prior functional data has been produced for these variants **Table 6-9**. However, there are a few cases where prior functional data has been produced. S2A was identified in 1 CRC patient with loss of MLH1 and PMS2 proteins with the patient meeting the Bethesda but not the Amsterdam criteria for LS diagnosis (Jiang et al., 2019). The authors concluded that this variant did not meet the criteria for classification as likely pathogenic (Jiang et al., 2019). Here, it is present in both mouse and chicken MLH1 **Table 6-9** and therefore is tolerated by the human protein. T364A was identified in MSI-H breast cancer in which MLH1 expression was lost (Curia et al., 1999). However, more recently it was identified in LS patients but considered non-pathogenic (Lellis et al., 2013). Here it is present in chicken MLH1 **Table 6-9** and is thus tolerated by the human protein according to our method. H381Y has been reported in an individual with early-onset gastric cancer showing high MSI (Bacani et al., 2005). However, this individual also had additional mutations in MLH1, including a nonsense mutation resulting in W597stop (Bacani et al., 2005). Therefore, it cannot be said as to whether H381Y is responsible for the MSI phenotype, especially as a premature stop would result in loss of the MLH1 C-terminus, required for MutL homologue interaction (Guerrette et al., 1999). Importantly, H381Y does not impact MLH1 stability or PMS2 binding in a Y2H assay (Wang et al., 2012), concordant with the results shown here as it is found in mouse MLH1 **Table 6-9**. S401N and D418G have been previously reported in individuals with colorectal cancer (Pearlman et al., 2019, Yurgelun et al., 2017). In one study, S401N was identified as a germline mutation in a patient who had both this and a somatic nonsense mutation E331X in both alleles via LOH (Pearlman et al., 2019). Therefore, the CRC phenotype in this individual could be attributed to the premature stop codon in both MLH1 alleles. In the other study, patients with CRC were screened using a 25 gene panel and the S401N and D418G variants were identified

in separate patients and classified as VUS (Yurgelun et al., 2017). No additional information was provided regarding the functional impact of these variants and it is possible that additional variants existed in these individuals which contributed to the CRC phenotype (Yurgelun et al., 2017). Here, both these variants are present in chicken MLH1 **Table 6-9** meaning that they are tolerated by the human protein and likely do not contribute to the CRC phenotype observed. In addition, R423T was present in chicken MLH1 **Table 6-9** and previous work showed that this variant had WT expression levels when expressed in HEK293, and restored sensitivity to 5-FU and MMS in HCT116 to the same extent as WT (Arora et al., 2017), concordant with results shown here. N444S is present in mouse MLH1 but has previously been reported as a VUS in one individual with colorectal cancer displaying MSI (Rodríguez–Soler et al., 2013). Remarkably, this patient also carried the A681T MLH1 variant, which is classified as pathogenic on ClinVar and was demonstrated to be non-functional in the human protein in this chapter. Therefore, the MSI phenotype is likely not attributed to N444S but rather A681T. Lastly, R522Q was identified in chicken MLH1 **Table 6-9** but has also been found in patients with CRC and breast cancer (Niessen et al., 2006; Yurgelun et al., 2015). However, these individuals were heterozygous for this variant in both cases (Niessen et al., 2006; Yurgelun et al., 2015) and this variant has also been identified in healthy controls as part of a breast cancer mutation study (B. C. A. Consortium, 2021). Furthermore, colorectal cancer patients carrying this mutation do not exhibit loss of MLH1 protein expression (Niessen et al., 2006). Therefore, the evidence suggests that this is perhaps a carrier mutation in patients with cancer. To my knowledge, no other functional information has been produced regarding any other variants without classifications of benign or pathogenic present in complementing orthologues. It must be noted that computational predictions using the MAPP + PolyPhen2 score in accordance with INSiGHT recommendations all agree with the findings presented here.

Lastly, comparison of the sequence alignment to PTMs for MLH1 catalogued in PhosphoSitePlus revealed several residues with reported PTMs that had undergone variation to non-modifiable residues in complementing orthologues **Table 6-10.** The modification of the majority of these residues was reported through high-throughput screens (Santamaria et al., 2011, Mertins et al., 2013, Cptac et al., 2016, Akimov et al., 2018, Hendriks et al., 2014, Lumpkin et al., 2017, Huttlin et al., 2010) and therefore lack prior functional assessment or any assignable role in MLH1 protein function. However, one diverged PTM has been assigned a direct role in MLH1 function **Table 6-10.** Specifically, phosphorylation of S406 by ATM/ATR was shown to be required for interaction with BRCA1 and subsequent activation of MLH1 in response to doxorubicin (Romeo et al., 2011). However, the authors showed no demonstrable role for this PTM in survival in response to doxorubicin only showing an increase in this PTM in response to doxorubicin via western blot. Therefore, the importance of this PTM for overall MLH1 protein function remains to be elucidated. The evidence presented here suggests this PTM is not required for either MLH1 MMR activity or PMS2 binding/stabilisation.

6.4.3 Future experiments based on the work presented in this chapter

The key next step of the work presented in this chapter is to quantitatively assess the interactions between MLH1 orthologues and the binding partners PMS2, EXO1 and MSH6. Subsequently, analysis could be carried out to quantitatively determine the expression levels of MLH1 orthologues in human cells as well as MLH1 constructs expressing variants tested in this chapter and how this correlates with PMS2 restoration. Mechanistically, it could be interesting to identify the mechanism behind the loss of complementation observed with frog MLH1. One hypothesis presented in the discussion is that the reduced interaction with EXO1 may underlie the partial complementation phenotype. The region in MLH1 required for this interaction is known (Schmutte et al., 2001) but the residues required within this domain are

currently unknown. Firstly, one could replace the corresponding EXO1 binding domain in frog (AA 410-650) with the human equivalent. If this fusion protein demonstrated full complementation of human cells then variants present in frog MLH1 within this region could be systematically reintroduced to try and determine residues behind loss of binding. This would offer good insight into the MLH1-EXO1 interaction, building on previous studies.

Overall, the results presented in this chapter have demonstrated that orthologue complementation can be used to identify tolerated variants in the MLH1 protein. Furthermore, it can also identify PTM sites which are not required for MLH1 function *in vivo*. The sequence alignment therefore provides a database of functional variation in the human MLH1 protein. Future studies identifying variation in the MLH1 protein are advised to use this as a reference when interpreting MLH1 variant impact.

7 Orthologue complementation of FANCL

7.1 Introduction

FANCL was first identified as *PHF9* via FANCA immunoprecipitation experiments which aimed to pull down and identify novel FA core complex factors responsible for FANCD2 monoubiquitination (Meetei et al., 2003). Sequence-based predictions of PHF9 structure hypothesised three WD40 repeats and a PHD-type zinc-finger motif (Meetei et al., 2003). PHD-type zinc-finger motifs are known to have ubiquitin ligase activity (Lu et al., 2002). PHF9 shows ubiquitin ligase activity *in vitro*, which was disrupted upon mutation of its PHD-type zinc-finger domain (Meetei et al., 2003). siRNA depletion of *PHF9* led to significantly lower levels of both basal and mitomycin C (MMC)-induced monoubiquitinated FANCD2 and increased chromosomal aberrations in response to MMC (Meetei et al., 2003). Crucially, little to no PHF9 protein was detected in the EUFA868 cell line from an individual with FA of unassigned complementation group (Meetei et al., 2003). These cells display hypersensitivity to MMC and absence of FANCD2 monoubiquitination on K561, but this is corrected upon complementation with *PHF9* cDNA (Meetei et al., 2003). The authors concluded that this individual represented a novel FA complementation group, FA-L, leading to PHF9 being termed FANCL, responsible for the ubiquitination of FANCD2 (Meetei et al., 2003, Meetei et al., 2004). In addition to FANCD2, FANCL also ubiquitinates FANCI, which forms a heterodimer with FANCD2 (Smogorzewska et al., 2007, Alpi et al., 2008, Longerich et al., 2009). Mechanistically, this is dependent on the ubiquitination of FANCD2 by FANCL, as cells expressing a ubiquitination defective FANCD2 K561R mutant do not display FANCL-mediated ubiquitination of FANCI on K523 (Smogorzewska et al., 2007, Alpi et al., 2008, Longerich et al., 2009). FANCL's E3-ubiquitin ligase activity *in vitro* and *in vivo* is dependent on its ubiquitin-conjugating enzyme UBE2T (Machida et al., 2006), now termed FANCT (Rickman et al., 2015, Hira et al., 2015, Virts et al., 2015). UBE2T-deficiency causes increased

chromosomal aberrations in response to MMC, mimicking FANCL deficiency (Machida et al., 2006). Complementation of UBE2T-deficient cells restores protection against MMC as well as FANCD2 monoubiquitination (Rickman et al., 2015). Within the FA core complex, FANCL forms a dimer of trimers in complex with FAAP100 and FANCB to catalyse the coordinated ubiquitination of the FANCD2-FANCI heterodimer (Twest et al., 2017). A FANCL-FANCB-FAAP100 complex stimulates the *in vitro* ubiquitination activity of FANCL on FANCD2 by approximately 5-fold over FANCL alone (Rajendra et al., 2014). Furthermore, FANCL-FANCB-FAAP100 is sufficient to maintain residual levels of FANCD2 monoubiquitination in cells defective for other FA core complex subunits (Rajendra et al., 2014, Huang et al., 2014). Furthermore, FANCD2-FANCI must be DNA bound for ubiquitination to occur (Twest et al., 2017).

Mechanistically, ubiquitinated FANCD2-FANCI has been proposed to recruit structure specific nucleases and associated proteins to the site of the ICL to mediate its removal (Zhang and Walter, 2014, Dan et al., 2021). SLX4 in combination with SLX4IP works to recruit XPF-ERCC1 and MUS81-EME1 to the site of the ICL (Klein Douwel et al., 2014, Zhang and Walter, 2014, Zhang et al., 2019a). Ubiquitinated FANCD2 has been shown to mediate the recruitment of SLX4 via SLX4's UBZ motif (Yamamoto et al., 2011). Furthermore, the binding of SLX4 and XPF to ICLs in *Xenopus* egg extracts requires ubiquitinated FANCD2 (Klein Douwel et al., 2014). However, the recruitment of SLX4 to sites of ICLs has been shown to be independent of FANCD2 in some cases (Lachaud et al., 2014). Furthermore, the recruitment of FAN1 has been shown to be dependent on ubiquitinated FANCD2 and the UBZ domain of FAN1 (Kratz et al., 2010, Liu et al., 2010b, MacKay et al., 2010, Smogorzewska et al., 2010). However, this was questioned by the finding that mutations disrupting the UBZ domain of FAN1 are functional for ICL resistance, implying a FANCD2-ub-independent recruitment pathway (Zhou et al., 2012b). Ubiquitin-binding proteins typically recognise a

hydrophobic patch on the ubiquitin moiety (Dikic et al., 2009). Recent cryo-EM studies show that the hydrophobic patches on both ubiquitin molecules of the ubiquitinated FANCD2-FANCI heterodimer bound to DNA are concealed within the interface of FANCD2-FANCI (Wang et al., 2020a). This implies that they are not easily accessible to the ubiquitin binding proteins of the ICL repair pathway (Lemonidis et al., 2022). Concomitantly, ubiquitination of FANCD2-FANCI had no effect on the interaction of this complex with several DNA repair proteins including SLX4 and FAN1 (Tan et al., 2020). It is now known that FANCD2 ubiquitination induces conformational changes in the FANCD2-FANCI heterodimer, significantly increasing its affinity for dsDNA (Alcón et al., 2020, Wang et al., 2020a, Rennie et al., 2020). Some findings suggest that this complex loses its preference for ICLs, becoming a sliding clamp upon which subsequent repair reactions can be coordinated (Wang et al., 2020a). Overall, FANCL-mediated ubiquitination of FANCD2-FANCI is crucial to the ICL repair pathway, but the specific mechanisms behind its importance are still being elucidated (Lemonidis et al., 2022).

Structurally, human FANCL comprises an ELF domain (residues 1-105), a central domain (residues 109-294) and a RING domain (residues 299-373) (Hodson et al., 2011, Hodson et al., 2014, Gurtan et al., 2006, Miles et al., 2015) in contrast to the initial sequence-based prediction of three WD40 repeats (residues 75-124; residues 147-202; residues 243-275) and a PHD finger domain (residues 307-363) (Meetei et al., 2003). The ELF domain of FANCL has been shown to bind to ubiquitin (Miles et al., 2015). Notably, the interaction is not important for recognition of the core complex or the interaction between FANCL and UBE2T or for FANCD2 monoubiquitination *in vitro* (Miles et al., 2015). Despite this, complementation of FANCL null cells with FANCL constructs mutated at residues important for ubiquitin binding does not restore FANCD2 monoubiquitination in response to MMC, suggesting an important role for FANCL binding to ubiquitin *in vivo* (Miles et al., 2015). The central domain of FANCL

comprises a ubiquitin conjugating (UBC) fold fused to a RWD fold (URD) (Hodson et al., 2011). The C-terminal end of the URD mediates the interaction between FANCL and FANCD2-FANCI (Hodson et al., 2011). The RING domain of FANCL binds to UBE2T (Hodson et al., 2011, Hodson et al., 2014). Initial studies showed what were thought to be the WD40 repeats of FANCL as crucial for binding to the FA core complex (Gurtan et al., 2006). Deletion of any of the WD40 repeats results in loss of binding to FANCA and loss of restoration of FANCD2 monoubiquitination upon complementation of a *FANCL*$^{-/-}$ cell line (Gurtan et al., 2006). More recent work has established in detail how FANCL forms part of the core complex. Two FANCL-FANCB-FAAP100 complexes form a symmetric dimer at the centre of the FA core complex mediated by interactions between the two FANCB molecules (Swuec et al., 2017, Rajendra et al., 2014). This dimer forms a scaffold for the recruitment of two FANCC-FANCE-FANCF trimeric assemblies (Swuec et al., 2017). As FANCE was reported to bind to the FANCD2-FANCI heterodimer, it was proposed that FANCC-FANCE-FANCF might act as a molecular matchmaker to bridge between the FANCD2-FANCI mono-ubiquitination substrate and the FANCL-FANCB-FAAP100 E3 ligase machinery (Swuec et al., 2017; van Twest et al., 2017). The cryo-EM structure of the whole FA complex confirmed previous observations that the FA core complex comprises two central dimers of the FANCB and FAAP100 subunits, flanked by two copies of FANCL interacting with the FANCB-FAAP100 coiled-coil (Shakeel et al., 2019, Swuec et al., 2017). Furthermore, FANCL contacts the substrate recognition module (FANCC-FANCE-FANCF); the RING and ELF domains of FANCL contacted FANCE and FANCF and the URD domain contacted FANCC and FANCE (Shakeel et al., 2019). Lastly, the two FANCL subunits are in different conformations at opposite ends of the complex (base and top) (Shakeel et al., 2019). It is proposed that the RING domain of the base FANCL subunit may promote substrate binding with FANCE, whereas the top FANCL subunit may be the active E3 which promotes ubiquitin transfer to FANCD2 (Shakeel et al., 2019).

Notably, this purified complex did not ubiquitinate FANCI, contradicting previous models (Twest et al., 2017, Swuec et al., 2017, Shakeel et al., 2019). It was suggested that this might require an additional activation step (Shakeel et al., 2019). This was revealed recently with the finding that ATR phosphorylation of FANCI primes the FANCD2-FANCI complex for ubiquitination (Sijacki et al., 2022). To date, only one PTM of FANCL has been assigned a role in FANCL function. Specifically, phosphorylation on T187 by GSK3β prevents its degradation by the proteasome in response to ubiquitination (Dao et al., 2013). In addition to this, other PTM sites on FANCL have been reported including ubiquitination, phosphorylation and SUMOylation, but all were identified via high-throughput methodologies and are yet to be functionally studied *in vivo* (Mertins et al., 2013, Akimov et al., 2018, Hendriks et al., 2014, Udeshi et al., 2013, Kim et al., 2011).

As explained in section 2.2.5, biallelic LOF of *FANCL* gives rise to FA. Notably, LOF of *FANCL* makes up just 0.2% of FA cases, in contrast to *FANCA* (57%), *FANCC* (11%) and *FANCG* (11%) (Neveling et al., 2009). The most frequently mutated *FANC* genes are all peripheral members of the FA core complex (Swuec et al., 2017, Shakeel et al., 2019) and loss of any of these components reduces but does not completely abrogate FANCD2-FANCI monoubiquitination *in vivo* and *in vitro* (Rajendra et al., 2014, Huang et al., 2014). In contrast, disruption to the FANCL-FANCB-FAAP100 module results in complete loss of residual FANCD2-FANCI monoubiquitination (Rajendra et al., 2014, Huang et al., 2014). Concomitantly, deletion of either *FANCG* or *FANCE* in produces mild sensitivity to ICLs in contrast to deletions of either *FANCB* or *FANCL* which causes severe sensitivity to ICLs (Huang et al., 2014). This information combined with the FA core complex structure suggests that mutations to peripheral components likely do not disrupt catalytic module formation and are thus better tolerated (Shakeel et al., 2019). In contrast, variants in FANCL or FANCB may disrupt its formation and thus are less tolerated and associated with more severe phenotypes

(Shakeel et al., 2019). Indeed, FA patients with LOF mutations in *FANCL* typically display a severe FA phenotype including bone marrow failure and leukaemia (Donovan et al., 2020, Wu et al., 2017, Ameziane et al., 2012). In some cases, early post-natal death was observed with severe deformities to the new-born (Vetro et al., 2015). Concomitantly, deletion of *FANCL* in mice is embryonic lethal in some strains (Agoulnik et al., 2002). It must be mentioned that patients with hypomorphic mutations in *FANCL* displayed only mild FA symptoms including café-au-lait spots and stunted growth (Ali et al., 2009). Notably, some mutations in *FANCL* have been found in patient cancer cells who do not have FA (Reinhold et al., 2012, Gao et al., 2013, Cerami et al., 2012). Recently, an analysis of the impact of some of these mutations on FANCL protein function was carried out (Frost et al., 2020). Multiple missense mutations in *FANCL* were found to disrupt its stability conferring sensitivity to MMC and loss of *in vitro* ubiquitination of FANCD2 (Frost et al., 2020). The authors posit that such mutations may have contributed to the progression of cancer in the patients they were found in (Frost et al., 2020). Currently, 208 missense mutations in FANCL on ClinVar are without sole classifications of either benign or pathogenic. Obtaining a better understanding of the impact of these variants may provide further insight into the mechanisms of FANCL disruption in FA and cancer.

All vertebrates are known to contain the following FA core complex proteins FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL and FAAP100, demonstrating that the FA gene network has remained intact since the last common ancestor of all vertebrate lineages (Titus et al., 2006). However, the complement of FA proteins varies amongst non-vertebrate organisms (McVey, 2010). *Drosophila melanogaster*, *Dictyostelium discoideum, Caenorhabditis elegans*, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* all contain a FANCM orthologue (McVey, 2010) and *Dictyostelium discoideum* also contains a FANCE orthologue (Zhang et al., 2009). In contrast to many of the FA core complex components, FANCD2, FANCI and FANCL are widely conserved across evolution (McVey,

2010), excluding *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* (Collis et al., 2006, Marek and Bale, 2006, Knipscheer et al., 2009, Zhang et al., 2009, Youds et al., 2009, Lambert et al., 2003, Ben-Yehoyada et al., 2009, Meier et al., 2009). The differences in conservation of FA proteins are sometimes reflected in the requirement of the FA proteins for repair of ICLs between organisms (McVey, 2010, Daee and Myung, 2012). In *Saccharomyces cerevisiae*, deficiency of the FANCM orthologue mph1 does not result in sensitivity to crosslinking agents, in contrast to human cells (Meetei et al., 2005, Schürer et al., 2004). Furthermore, in *Dictyostelium*, loss of either *FncL* (*FANCL*), *FncI* (*FANCI*), *FncD2* (*FANCD2*) or *FncM* (*FANCM*) results in mild sensitivity to cross-linking agents (Zhang et al., 2009). In contrast, in *Schizosaccharomyces pombe*, deficiency of the FANCM orthologue Fml1 results in hypersensitivity to cisplatin (Sun et al., 2008). Additionally, knockdown of *FANCD2* or *FANCL* in *Drosophila* by RNA interference results in hypersensitivity to a number of crosslinking agents, and depletion of *FANCD2* also causes chromosome fusions and defects in the S-phase checkpoint (Marek and Bale, 2006). Concomitantly, depletion of *FCD-2* (*FANCD2*) in *Caenorhabditis elegans* results in hypersensitivity to crosslinking agents (Collis et al., 2006). The ubiquitin ligase activity of FANCL has also been identified in many of the organisms described above. FANCL was shown to ubiquitinate FANCD2 in *Drosophila* (Marek and Bale, 2006) and was hypothesised to be the E3 ligase responsible for this in *Caenorhabditis elegans* (Collis et al., 2006). Notably, FncL was also found to be responsible for FncD2 ubiquitination in *Dictyostelium* (Zhang et al., 2009) highlighting that whilst this may not always be required for ICL repair, the E3 ligase activity of FANCL is conserved across evolution.

Overall, FANCL presents itself as an ideal candidate for orthologue complementation. The ubiquitin ligase activity of FANCL is crucial for ICL repair and loss of FANCL gives provides a clear assayable phenotype in extreme sensitivity to crosslinking agents and loss of FANCD2-

I monoubiquitination (Meetei et al., 2003, Lemonidis et al., 2022), in contrast to other FA components (Huang et al., 2014, Rajendra et al., 2014). Several PTMs have been described for FANCL but almost all lack any assigned functional role. Furthermore, mutations to FANCL give rise to FA, with the severity of the phenotype dependent on the impact of the mutation on FANCL function (Donovan et al., 2020, Wu et al., 2017, Ameziane et al., 2012, Ali et al., 2009). Additionally, some mutations in FANCL have been found in patient cancer cells who do not have FA (Reinhold et al., 2012, Gao et al., 2013, Cerami et al., 2012, Frost et al., 2020), with the majority of missense mutations in FANCL on ClinVar still without definitive classifications. Crucially, FANCL is widely conserved across evolution, in contrast to the majority of other core complex components (Collis et al., 2006, Marek and Bale, 2006, Knipscheer et al., 2009, Zhang et al., 2009, Youds et al., 2009, Lambert et al., 2003, Ben-Yehoyada et al., 2009, Meier et al., 2009) with its E3 ligase activity conserved down to at least *Dictyostelium* (Zhang et al., 2009). Overall, it presents itself as a suitable candidate for orthologue complementation.

7.2 Aims and Objectives

In this chapter, the complementation ability of vertebrate orthologues of FANCL will be assessed via FA core complex binding ability, cisplatin survival assays, and FANCD2-FANCI ubiquitination assays. Once orthologues which complement human cells have been identified, sequence alignments of complementing orthologues will be constructed and searched for variants to identify functional amino acid variants and non-essential PTM. Variants and non-essential PTM sites will then tested in the human protein. Lastly, the data will be compared to existing publications on variant effects and public variant databases.

7.3 Results

7.3.1 Selection of vertebrate FANCL orthologues for complementation of human cells

To assess the ability of FANCL orthologues to complement human cells, orthologues from distinct taxonomic groups with varying sequence divergence from human were selected. Orthologue sequences for *Homo sapiens* (human) *Mus musculus* (mouse) *Gallus gallus* (chicken) *Xenopus tropicalis* (frog) and *Danio rerio* (zebrafish) FANCL were obtained from https://www.ncbi.nlm.nih.gov/ with accession numbers displayed in **Table 3-1**. Multiple accession numbers existed for each orthologue and thus the entries showing the closest lengths to human were selected for each species. Sequences from each protein of interest were then aligned using the T-Coffee algorithm (https://www.ebi.ac.uk/Tools/msa/tcoffee) (Notredame et al., 2000) and the conservation of identical or similar residues was indicated using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html) with conservation marked via conservation of ≥ 80% residues **Figure 7-1**. The percentage identity to human FANCL was then calculated using the % identity matrix output option from the T-Coffee alignment **Table 7-1.**



Figure 7-1: Sequence alignment of vertebrate FANCL orthologues used in this study with conservation shading

Sequence alignment of vertebrate FANCL orthologues used in this study. Sequences were aligned using the T-Coffee algorithm (Notredame et al., 2000) and conservation was scored using Multiple Align Show (https://bioinformatics.org/SMS/multi_align.html). Areas of black shading indicate 80% identical conservation of a residue, areas of grey shading indicate 80% conservation to at least an amino acid with similar biochemical properties. Areas of white shading indicate less than 80% conservation. Numbers indicate AA position.

FANCL Orthologue	Accession Number	% Identity to Human Orthologue
Human	NP_001108108.1	100
Mouse	NP_080199.1	79
Chicken	NP_001029113.1	70
Frog	NP_001007482.1	68
Fish	NP_998147.1	58

Table 7-1: Accession numbers and % identity to human orthologue for vertebrate FANCL orthologues used

7.3.2 FANCL orthologues show decreasing affinity to the human FA core complex

Loss of FANCL association with the FA core complex abrogates its ubiquitin ligase activity, and thus its primary cellular function (Gurtan et al., 2006). Therefore, the ability of FANCL orthologues to associate with the core complex was determined. cDNAs encoding the vertebrate FANCL orthologues in **Table 7-1** were separately cloned into pcDNA5-FRT-TO-neo-mClover3. mClover3 is a GFP variant containing multiple point mutations to improve photostability and brightness (Bajar et al., 2016). These constructs were then transfected into 293FT cells along with a pcDNA5-FRT-TO-neo-mClover3 empty vector control. Immunoprecipitations were then performed to determine the ability of FANCL orthologues to pull down FANCA, a member of the FA core complex (Shakeel et al., 2019). This revealed that the FANCL orthologues used showed decreasing affinity for the FA core complex with increasing sequence divergence from human **Figure 7-2.** Expression levels of all constructs apart from fish FANCL appeared to be similar after Ponceau S staining **Figure 7-2**. Fish FANCL appeared to show reduced expression; the reasons for this remain unclear **Figure 7-2**.

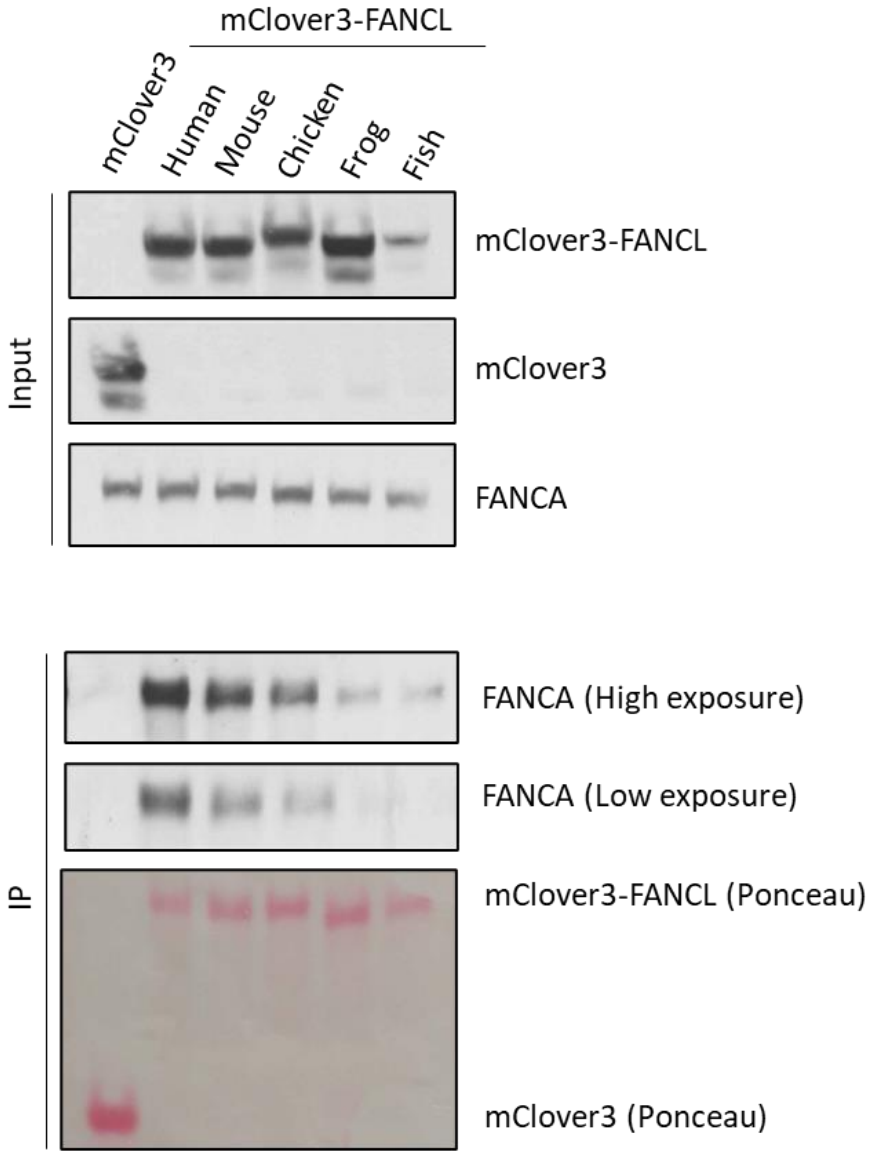


Figure 7-2: **FANCL orthologues show decreasing affinity to the human FA core complex with increasing sequence divergence to human**

Western blot depicting GFP-pulldowns from 293FT cells transfected with either mClover3 only or mClover3-tagged human, mouse, chicken, frog and fish FANCL orthologues showing that with increasing sequence divergence to human, affinity for human FANCA decreases. Ponceau staining demonstrates expression of mClover 3 and mClover3-tagged FANCL orthologues and respective molecular weights. N=3. 1 repeat performed by Miss. Roos Vincken.

7.3.3 FANCL orthologues show reduced FANCD2 ubiquitination activity

As mentioned previously, FANCL's association with the core complex is required for its FANCD2-FANCI ubiquitination activity (Gurtan et al., 2006). Because of this, orthologues which show reduced affinity for human FANCA should also show reduced FANCD2-FANCI ubiquitination activity. Therefore, an assessment of FANCL orthologues ability to ubiquitinate FANCD2 was performed. To assess this, mClover3-FANCL orthologues were cloned into pLenti-PGK via Gateway cloning. Lentiviruses expressing mClover3-tagged FANCL orthologues were then produced and used to transduce HAP1 Δ*FANCL* cells. Prior to me undertaking this project, a stable pool of HAP1 Δ*FANCL* cells stably expressing mClover3-tagged human FANCL had been produced in our lab by Miss. Roos Vincken. Therefore, I produced stable pools of HAP1 Δ*FANCL* cells stably expressing mClover3-tagged mouse, chicken, frog and fish FANCL and mClover3 alone, via transduction with lentivirus followed by selection in selective medium. Stable pools were confirmed via FACS to assess overall GFP+ cells of the population after selection, with all pools showing >80% GFP+ **Figure 10-11**. Stable pools were then treated with 1 mM of hydroxyurea (HU) for eight hours to induce replication fork stalling and S-phase arrest, capturing the peak of ubiquitinated FANCD2 in the cell cycle (Smogorzewska et al., 2007). The level of FANCD2 ubiquitination was then determined via western blot. Human FANCL could restore FANCD2 ubiquitination to the same extent as in WT cells **Figure 7-3**. In contrast, mouse, chicken, frog and fish FANCL could only partially, or could not restore FANCD2 ubiquitination to WT levels **Figure 7-3**. These findings are perhaps the result of the reduced affinity for the human FA core complex shown by FANCL orthologues in **Figure 7-2**. However, despite selection via selective medium followed by FACS to obtain a stable pool, FANCL orthologues showed significantly reduced expression compared to human FANCL **Figure 7-3**. Therefore, the reduced FANCD2 ubiquitination observed in

cells expressing FANCL orthologues could be the result of reduced expression of FANCL, rather than reduced affinity for the FA core complex.

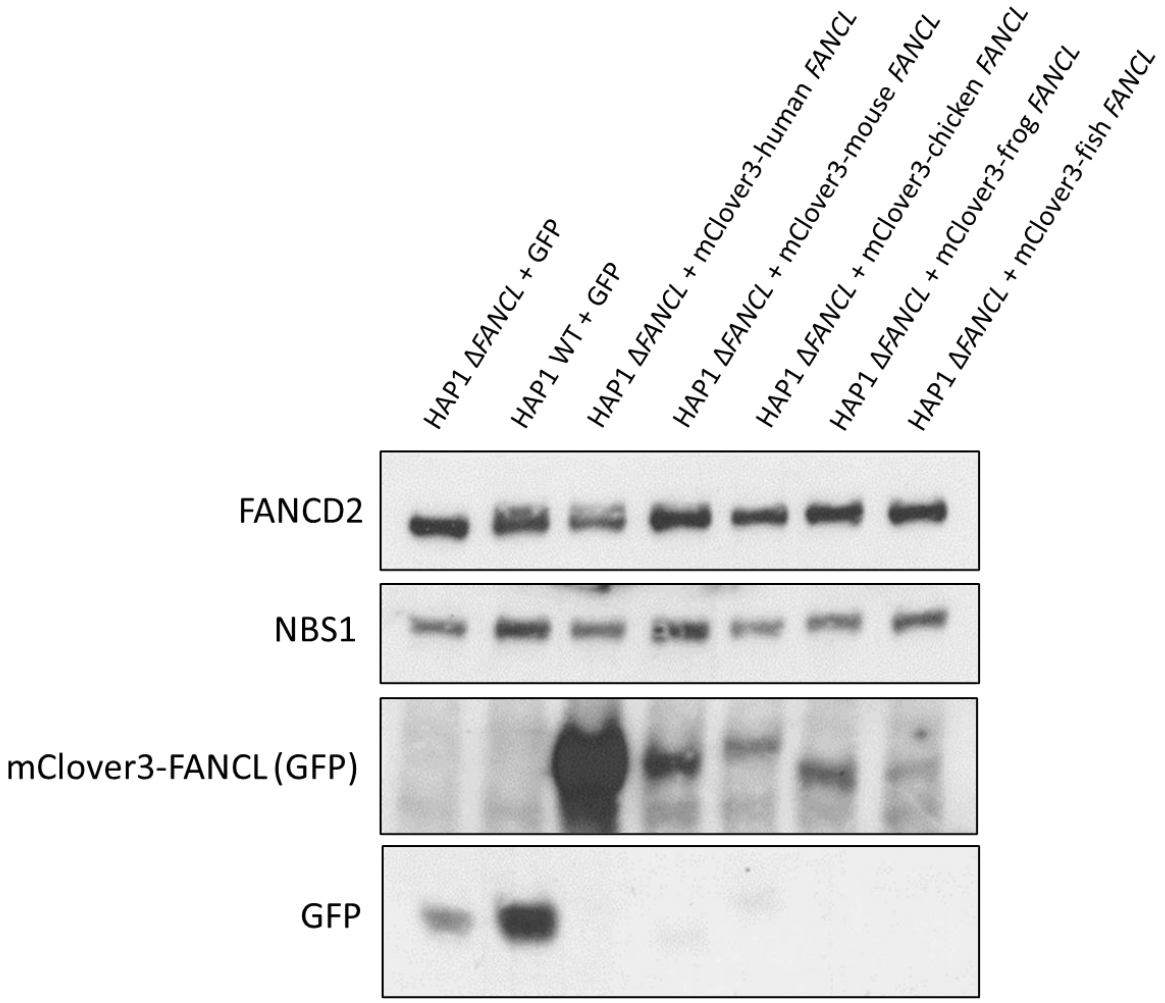


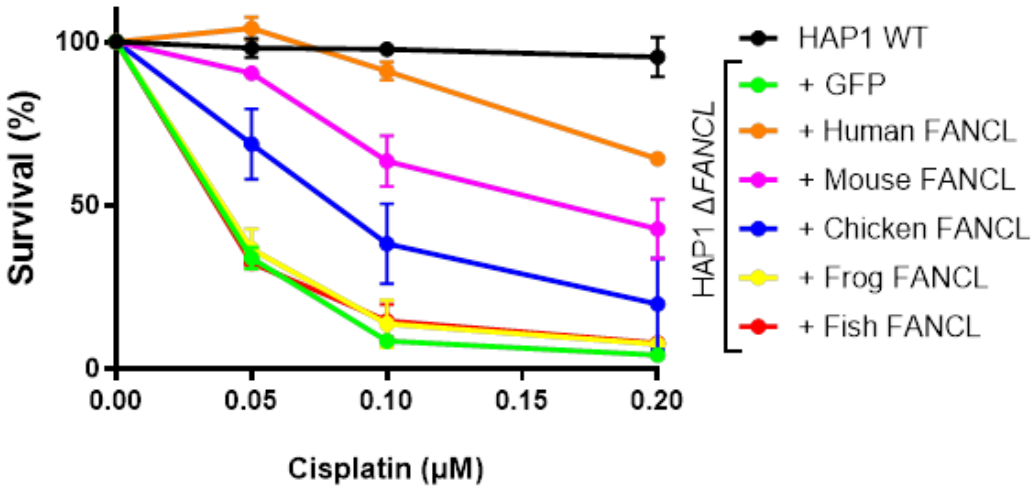
Figure 7-3: FANCL orthologues show reduced or absent FANCD2 ubiquitination ability

Western blot for FANCD2 ubiquitination using lysates prepared from HAP1 WT and HAP1 Δ*FANCL* cells stably expressing GFP or mClover3-tagged FANCL orthologues treated with 1 mM HU for 8 hours. A band shift indicates ubiquitination of FANCD2. GFP antibody was used to probe for mClover3 and mClover3-tagged FANCL orthologues. NBS1 was used as a loading control. N=2.

7.3.4 FANCL orthologues do not restore protection against cisplatin to WT levels

Having determined that vertebrate FANCL orthologues show reduced FANCD2 ubiquitination activity, it was then determined whether they could restore protection against cross linking agents, as loss of FANCL function gives rise to extreme sensitivity to ICLs (Huang et al., 2014, Meetei et al., 2003). To assess this, stable pools of HAP1 Δ*FANCL* cells expressing mClover3-tagged FANCL orthologues were sorted via FACS to obtain exact cell numbers for plating and seeded as single cells. 24 hours after seeding, cells were treated with varying doses of cisplatin for 4 hours. Colony formation was then assessed 1 week after seeding to determine overall survival. All of mouse, chicken, frog and fish FANCL failed to restore protection against cisplatin to the same extent as human FANCL **Figure 7-4**.The sensitivity to cisplatin observed upon complementation with FANCL orthologues correlates well with the reduced interaction with the core complex, reduced expression levels and reduced FANCD2 ubiquitination activity **Figure 7-2**, **Figure 7-3**. Overall, the FANCL orthologues utilised here cannot be used to form sequence alignments for functional variant identification. Potential reasons the for lack of complementation observed are discussed in the next section.

A



B

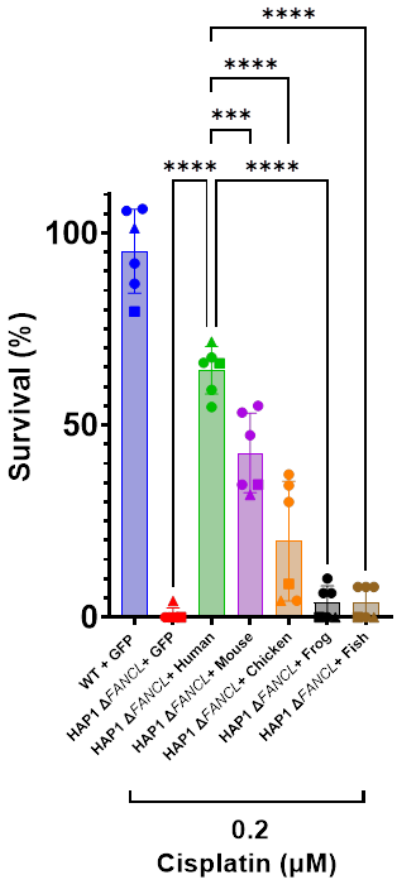


Figure 7-4: FANCL orthologues do not restore protection against cisplatin in human cells

A: Survival curves showing colony survival % of HAP1 WT or ΔFANCL HAP1 cells stably expressing GFP-tagged FANCL orthologues from human, mouse, chicken, frog and fish in response to increasing doses of cisplatin. Data is plotted as a mean of N=2. Error bars indicate SEM. **B:** One-way ANOVA was performed on all data points of N=2 at 0.2 μM cisplatin and included a Dunnett's multiple comparisons test comparing HAP1 Δ*FANCL* + mClover3-WT FANCL to HAP1 Δ*FANCL* + mClover3-mouse, chicken, frog and fish FANCL. Results from Dunnett's multiple comparisons test are displayed. ns = non-significant difference, *** = p 0.0006, **** = p <0.0001.

7.4 Discussion

7.4.1 Summary of the work presented in this chapter

In contrast to XPA, XRCC4 and MLH1, complementation of human cells with the vertebrate FANCL orthologues used here could not be achieved. Mouse, chicken, frog and fish FANCL showed reduced affinity for FANCA of the FA core complex **Figure 7-2**. Furthermore, mouse, chicken, frog and fish FANCL showed reduced expression and FANCD2 ubiquitination activity in human cells **Figure 7-3**. Finally, none of mouse, chicken, frog or fish FANCL could restore protection against cisplatin in human cells **Figure 7-4**. Therefore, sequence alignments to identify functional variation in the human protein could not be generated. In the previous chapters, complementation was observed with orthologues showing sequence similarity to human of 62% for XPA (fish), 47% for XRCC4 (frog) and 80% for MLH1 (chicken). Here, complementation was lost with an orthologue showing 79% similarity to the human protein. Thus, it remains possible that complementation could be achieved with FANCL orthologues demonstrating >79% similarity to the human protein. Furthermore, the minimum sequence similarity required for complementation appears to differ significantly between proteins.

7.4.2 A discussion of the work presented in this chapter

The FA core complex is well conserved throughout vertebrates including those utilised here (Titus et al., 2006). Furthermore, the FANCL and FANCD2-FANCI machinery is conserved throughout both vertebrates and invertebrates, excluding yeasts (Collis et al., 2006, Marek and Bale, 2006, Knipscheer et al., 2009, Zhang et al., 2009, Youds et al., 2009, Lambert et al., 2003, Ben-Yehoyada et al., 2009, Meier et al., 2009). Therefore the loss of binding to the FA core complex cannot be attributed to lack of requirement of this complex in vertebrate organisms.

There are many possible reasons for the lack of orthologue complementation observed in this chapter. FANCL orthologues showed reduced affinity for the FA core complex **Figure 7-2**.

Furthermore, FANCD2 ubiquitination was significantly reduced or absent and protection against cisplatin was attenuated in cells stably expressing FANCL orthologues **Figure 7-3**, **Figure 7-4**. However, as explained in previous chapters, no quantitative assessment of the interactions was performed. FANCL orthologues appear to display reduced expression in human cells after lentiviral integration **Figure 7-3**. Therefore, it remains possible that FANCL orthologues have equal ability to restore FANCD2 ubiquitylation and protection against cisplatin if the levels of expression are equal. However, it is also possible that FANCL orthologues are unstable in human cells, resulting in reduced expression and loss of complementation. The association of FANCL with the FA core complex is required for full FANCD2 ubiquitination activity (Gurtan et al., 2006, Rajendra et al., 2014), crucial for ICL repair (Huang et al., 2014). Therefore the reduced affinity of FANCL orthologues for the FA complex may underlie the phenotypes observed here. Therefore, one way in which clarity could be gained on the reasons underpinning loss of complementation would be to perform quantitative assessment of the interaction between FANCL orthologues and human FANCA, using the methods described in section 4.4.2. The regions previously believed to be the WD40 repeats of FANCL (residues 75-124; residues 147-202; residues 243-275) are required for binding to the FA core complex, and deletion of any of these repeats disrupts binding to FANCA and FANCD2 ubiquitination (Gurtan et al., 2006). Divergence of amino acids in these regions is present in all orthologues **Figure 7-1**, which may underlie the reduced affinity for the core complex. However, the loss of complementation cannot be assigned to alteration of a single amino acid in any given orthologue as extensive variation exists in these regions between species. Crosslinking mass-spectrometry has identified multiple lysines within FANCL which are important for intra and intermolecular crosslinks between itself and FANCB (Shakeel et al., 2019, Swuec et al., 2017) **Table 7-2.** Notably, at least one of these residues has diverged to a residue with different biochemical properties in the FANCL orthologues used here **Table**

7-2. This could underlie the reduction in the stability of FANCL orthologues **Figure 7-3** and/or the reduced affinity for the FA core complex **Figure 7-2**. Additional residues have been identified via crosslinking mass-spectrometry which mediate FANCL's interaction with FANCE, FANCF and FANCC (Shakeel et al., 2019). However, the data describing which residues were involved remains inaccessible due to a fault with the server storing the data (https://www.ebi.ac.uk/pride/directLink.do?experimentAccessionNumber=PXD014282), preventing inclusion in this discussion. Mutation to residues in *Drosophila* FANCL required for binding to ubiquitin abrogates the *in vivo* ubiquitination activity of FANCL (Miles et al., 2015). The homologous residues in human are Q47, K49, Q72 and D79 (Miles et al., 2015). Q47 and Q72 are conserved throughout the orthologues utilised here **Figure 7-1**. However, K49 is present as a glycine in frog, losing the negative charge. Furthermore, D79 is present as an asparagine in frog and a threonine in fish **Figure 7-1**, therefore losing the negative charge at this position. The divergence of these two residues may underlie the loss of ubiquitination activity of these orthologues **Figure 7-3**. It remains possible that mouse and chicken FANCL could still bind to ubiquitin, however this interaction is not required for interaction with the FA core complex (Miles et al., 2015) and thus *in vivo* ubiquitination activity may still be lost.

Residue (in human FANCL)	Contacts	Mouse	Chicken	Frog	Fish	Reference
K22	FANCB, FANCL	K	G	G	K	(Swuec et al., 2017), (Shakeel et al., 2019)
K48	FANCL	K	K	G	K	(Shakeel et al., 2019)
K125	FANCL	K	K	K	K	(Shakeel et al., 2019)
K137	FANCL	K	K	Q	Q	(Shakeel et al., 2019, Swuec et al., 2017)
K139	FANCL	K	K	K	K	(Shakeel et al., 2019, Swuec et al., 2017)
K153	FANCL	K	K	K	K	(Shakeel et al., 2019, Swuec et al., 2017)
K155	FANCL	K	N	N	K	(Swuec et al., 2017)
K215	FANCL	K	K	K	K	(Shakeel et al., 2019)
K266	FANCL	K	N	N	T	(Swuec et al., 2017)
K289	FANCB	K	K	K	K	(Swuec et al., 2017)
K305	FANCL	E	K	K	K	(Shakeel et al., 2019)
K369	FANCL, FANCB	K	K	K	K	(Shakeel et al., 2019, Swuec et al., 2017)
K374	FANCL	K	K	K	K	(Shakeel et al., 2019)
% of these residues conserved		92% (12/13)	85% (11/13)	62% (8/13)	85% (11/13)	

Table 7-2: Conservation of residues required for intramolecular and intermolecular interactions of FANCL

To date, no pathogenic variants for FANCL are catalogued on Clinvar meaning that the presence of pathogenic variants in complementing orthologues could not be assessed. Expansion to variants with classifications of likely pathogenic only revealed M1 variants

resulting in loss of the start codon. Therefore, the presence of currently established pathogenic variants cannot be attributed to loss of complementation with any of the orthologues used here.

Overall, FANCL orthologues from vertebrates were not able to complement human cells, preventing their use in sequence alignments for functional variant identification. This highlights that whilst the methodology is effective for some proteins, it is limited by a lack of complementation in others.

7.4.3 Future work based on the findings presented in this chapter

Firstly, quantification of the interaction of FANCL orthologues with FANCA as well as quantifying the level of expression and FANCD2 ubiquitination activity would provide more detailed clarity regarding their complementation ability. Furthermore, the lack of complementation observed with the orthologues used in this chapter has prevented identification of functional variants in the human protein. Therefore, it would be interesting to identify orthologues that could complement human cells (if any) by first assessing orthologues more identical to human than mouse FANCL (>79% similarity) e.g. primate orthologues. Alternatively, Sections of mouse FANCL could be replaced with human FANCL or vice versa, to try and identify the region of mouse FANCL responsible for loss of complementation. However, one region alone may not be responsible and divergence in multiple domains may be contributing to the phenotype. Mechanistically, one could mutate the lysine residues responsible for FA core complex binding which have diverged in mouse back to lysine, to see if the association with the human FA core complex could be increased. Otherwise, it is very challenging to propose a specific experiment to try and restore complementation, asides from the creation of human-mouse FANCL fusion orthologues. Overall, the use of orthologues with higher similarity to human FANCL than those used here offer the best chance of being able to identify functional amino acid variants in the human protein.

8 Discussion

8.1 Summary of the work presented in this thesis

The primary aim of this study was to determine whether systematic orthologue complementation of human cells could be achieved with proteins from distinct DDR pathways, and if so, could sequence alignments of complementing orthologues reveal functional amino acid variants tolerated by the human protein as well as non-essential PTM sites. In Chapter 3, I showed that vertebrate orthologues of XPA could bind to a known interactor (XPC) and restore protection against UV radiation in an XPA-deficient cell line. I then showed that invertebrate orthologues could not complement human cells, demonstrated via reduced affinity for a known binding partner and lack of restoration of protection against UV radiation. Subsequently, I demonstrated that a sequence alignment of complementing XPA orthologues could be used to identify functional variants in the human protein as well as non-essential PTM sites. In Chapter 4, this was corroborated with the demonstration that mouse, chicken and frog but not fish XRCC4 orthologues could bind to human XRCC4 and restore full protection against IR in an XRCC4-deficient human cell line, showing that this phenomenon is not exclusive to XPA. Subsequently, the sequence alignment of these complementing orthologues identified non-essential PTM sites in the human XRCC4 protein. Building on this, in Chapter 5 I demonstrated orthologue complementation using mouse and chicken MLH1. These orthologues were able to bind to known interactors MSH6, PMS2 and EXO1, as well as restoring PMS2 stability and 6-TG sensitivity in multiple cell lines. In contrast, frog MLH1 could not perform all of these abilities to the same extent as human MLH1, showing reduced affinity for EXO1 and loss of 6-TG sensitivity and PMS2 restoration in MLH1 deficient cell lines. Using a sequence alignment of complementing orthologues, almost 60 VUS were identified with the demonstration that VUS in this alignment could be tolerated by the human protein using 6-TG sensitivity and PMS2 restoration assays. Crucially, it was shown that

functional variants were present at the sites of known pathogenic variants V185G and A681T in the human protein, specifically V185I and A681S. Both V185I and A681S were able to restore full 6-TG sensitivity and PMS2 expression whereas V185G and A681T showed defects in at least one of these phenotypes. Finally, in Chapter 6, I showed that vertebrate FANCL orthologues were unable to complement human cells, in contrast to the findings in previous chapters. None of moue, chicken, frog of fish FANCL were able to restore protection against cisplatin and showed reduced affinity for the FA core complex alongside reduced FANCD2 ubiquitination activity. This highlights that whilst the method is effective for some proteins, for others it does not work as well due to lack of complementation.. Overall, orthologue complementation provides a novel methodology which allows the identification of functional amino acid variants and non-essential PTM sites. The sequence alignments presented here can be used as a reference database for the future interpretation of variants and PTM sites identified in the protein of interest.

8.2 An assessment of orthologue complementation

Detailed discussions of each protein were provided at the end of each results chapter. Here, I will discuss the advantages and limitations of using orthologue complementation to determine functional amino acid variants in the human protein, in the context of the current literature.

8.2.1 Strengths of orthologue complementation

The results presented in this thesis demonstrate that orthologue complementation provides an efficient means of identifying a high number of variants which are functional in the human protein, shown via multiple functional assays. For example, the sequence alignment of complementing MLH1 orthologues contains over 150 variants which will be tolerated by the human protein, demonstrated by binding to known interactors, protein stability and 6-TG sensitivity restoration. Generating 150 individual constructs each with unique variants and then

assessing their impact on protein stability and function would be experimentally demanding. Because of this, there is often a trade-off between the number of variants assessed and the number of functional assays performed. Concomitantly, studies which utilise individual construct approaches have often analysed < 100 MLH1 variants, of which a fraction are functionally understood controls (Drost et al., 2010, Nyström-Lahti et al., 2002, Plotz et al., 2006, Raevaara et al., 2005, Blasi et al., 2006, Bouvet et al., 2019, Kosinski et al., 2010, Brieger et al., 2002, Andersen et al., 2012, Borràs et al., 2012). More recently, approaches such as deep-mutational scanning have enabled the high-throughput assessment of the majority of all possible variants in MSH2 (Jia et al., 2021). However, on a similar note, this study was limited in its functional assessment of these variants to 6-TG sensitivity alone, and did not determine protein stability or variant impact on MSH3 or MSH6 binding (Jia et al., 2021). Furthermore, the saturation genome editing screen of *BRCA1* (Findlay et al., 2018a) solely looked at overall survival of a cell line containing a specific *BRCA1* missense mutation. It did not determine the effect a variant has in response to known functions of BRCA1 such as sensitivity to olaparib, RAD51 foci formation or binding to known interactors such as BARD1 (Findlay et al., 2018a). Furthermore, it did not perform assessment of the exon 11 region, which comprises over 60% of the coding sequence of *BRCA1* (Khachibi et al., 2015). Therefore, orthologue complementation offers a suitable middle-ground in that it is able to assess a large number of variants, but still provides detailed functional insight through the ease of performing multiple functional assays. This has important implications with the possibility of novel drug targets being discovered whose mechanism of action depends on loss of a particular element of protein function.

The ACMG/AMP framework dictates that pathogenic variation at any given site is moderate evidence that any subsequent variant would also be pathogenic (Richards et al., 2015). A clear strength of the methodology presented here is its ability to discern functional variants at sites

of known pathogenic variation. On a similar note, a sequence alignment of complementing orthologues provides a reference point for functional variation in the human protein. Importantly, complementing orthologues contain variants which as of yet are not present on the ClinVar database. In this sense, the sequence alignments are futureproof, in that they potentially contain as yet unidentified functional variants in the human protein. Clinicians or diagnostic laboratories may refer to the data presented here upon discovery of a novel variant in the proteins tested. On a similar note, researchers identifying novel PTM sites in any of the proteins tested here may refer to the sequence alignment to determine if that PTM is required for the functions analysed in this thesis. Overall, orthologue complementation is not restricted to look at variants already identified, a common feature of previous studies (Drost et al., 2010, Nyström-Lahti et al., 2002, Plotz et al., 2006, Raevaara et al., 2005, Blasi et al., 2006, Bouvet et al., 2019, Kosinski et al., 2010, Brieger et al., 2002, Andersen et al., 2012, Borràs et al., 2012, Murray et al., 2015, Mellon et al., 2002, Porter et al., 2005). Rather, it provides a reference for as yet unidentified functional variants in the human protein.

8.2.2 Weaknesses of orthologue complementation

The output of the methodology presented here is clearly limited to the variants present in complementing orthologues. This must be considered by researchers interested in specific variants as the variant of interest may not be present in any orthologues of the protein. This presents a disadvantage to orthologue complementation over recent techniques such as saturation mutagenesis (Findlay et al., 2018), deep-mutational scanning (Jia et al., 2021), prime-editing (Hayward et al., 2021) or base editor screens (Cuella-Martin et al., 2021, Hanna et al., 2021) which are capable of introducing a large number of variants into the human protein.

Another disadvantage or the methodology described here is that it is based on the transient transfection or lentiviral integration of cDNA constructs. This prevents the assessment of the impact variants have on pre-mRNA splicing. In contrast, CRISPR-based approaches to study

variant effects are capable of studying the impact variants have on this process (Findlay et al., 2018a, Cuella-Martin et al., 2021, Hanna et al., 2021, Hayward and Ciccia, 2021). This is important as missense mutations which impact splicing are often pathogenic(Daguenet et al., 2015). Because of this, it is essential to exclude any variants on or surrounding exon boundaries from methodologies which cannot assess pre-mRNA splicing, as performed here.

The methodology is restricted by the fact that it is challenging to identify which orthologues will complement prior to assessment. Using % identity to human is not an effective predictor as the work presented here showed that proteins show differing tolerances to variation regarding complementation of human cells. For example, complementation of the MLH1 protein was not obtained with an orthologue showing 76% identity to human MLH1; in contrast, complementation was obtained with an XRCC4 orthologue showing 39% similarity to human. Some proteins appear to be less tolerant to variation than others, which could be a consideration when designing prediction algorithms which incorporate evolutionary conservation as a metric (Ghosh et al., 2017).

Another limitation of the methodology presented here is that it does not directly determine which variants are pathogenic. Due to the sheer number of variants between the human orthologue and an orthologue which cannot complement human cells, it cannot be said which variant(s) are responsible for loss of complementation. However, as the method determines which variants are benign, this does exclude variants as pathogenic, narrowing down the number of variants which need to be assessed in future studies. One way in which an orthologue could be excluded prior to complementation studies could be identifying a known pathogenic mutation in the orthologue. Studies on MLH1 reveal that introduction of pathogenic mutations into the corresponding position in mouse MLH1 and MSH6 perturb their function in mouse embryonic stem cells (Houlleberghs et al., 2017, Houlleberghs et al., 2020).

8.3 Future work

In addition to the future work proposed in the individual chapter discussions, the obvious progression is to apply this methodology to additional clinically relevant proteins. During my DPhil, I began work on orthologue complementation of BRCA1, and this is currently being finished in the laboratory and will form part of the final manuscript. I hope that future studies also utilise this methodology to provide insight into their proteins of interest, from a clinical and/or mechanistic point of view. Additional experiments that could prove interesting include the creation of a 'pseudo-orthologue': a human protein containing all possible variants, determined from their presence in complementing orthologues. Furthermore, pathogenic variants could be introduced into fully complementing orthologues to further validate that they behave in the same manner as the human protein. As mentioned previously, this has already been performed for MLH1 and MSH6 by introducing the mutation in the human protein into the corresponding position in the mouse orthologue (Houlleberghs et al., 2017, Houlleberghs et al., 2020). Finally, during my DPhil I collaborated with Malitha Ratnaweera (another DPhil student) to create fully searchable databases containing information on the variants present in complementing orthologues and all associated information from databases such as ClinVar and PhosphoSitePlus. Upon publication of the manuscript, these will be made public and will provide a reference point for anyone interested in functional variation in any of the proteins where orthologue complementation was achieved. Overall, this should make a significant stride forward in our understanding of clinical variation in these proteins but also how they are regulated in human cells.

9 References

2001a. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860--921.

2001b. The Sequence of the Human Genome. *Science,* 291**,** 1304--1351.

2010. A map of human genome variation from population-scale sequencing. *Nature,* 467**,** 1061--1073.

AALTONEN, L. A., PELTOMÄKI, P., LEACH, F. S., SISTONEN, P., PYLKKÄNEN, L., MECKLIN, J.-P., JÄRVINEN, H., POWELL, S. M., JEN, J., HAMILTON, S. R., PETERSEN, G. M., KINZLER, K. W., VOGELSTEIN, B. & CHAPELLE, A. D. L. 1993. Clues to the Pathogenesis of Familial Colorectal Cancer. *Science,* 260**,** 812--816.

ABILDGAARD, A. B., STEIN, A., NIELSEN, S. V., SCHULTZ-KNUDSEN, K., PAPALEO, E., SHRIKHANDE, A., HOFFMANN, E. R., BERNSTEIN, I., GERDES, A.-M., TAKAHASHI, M., ISHIOKA, C., LINDORFF-LARSEN, K. & HARTMANN-PETERSEN, R. 2019. Computational and cellular studies reveal structural destabilization and degradation of MLH1 variants in Lynch syndrome. *eLife,* 8**,** e49138.

ABKEVICH, V., TIMMS, K. M., HENNESSY, B. T., POTTER, J., CAREY, M. S., MEYER, L. A., SMITH-MCCUNE, K., BROADDUS, R., LU, K. H., CHEN, J., TRAN, T. V., WILLIAMS, D., ILIEV, D., JAMMULAPATI, S., FITZGERALD, L. M., KRIVAK, T., DELOIA, J. A., GUTIN, A., MILLS, G. B. & LANCHBURY, J. S. 2012. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *British Journal of Cancer,* 107**,** 1776--1782.

ACHARYA, S., WILSON, T., GRADIA, S., KANE, M. F., GUERRETTE, S., MARSISCHKY, G. T., KOLODNER, R. & FISHEL, R. 1996. hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proceedings of the National Academy of Sciences,* 93**,** 13629--13634.

ADAM, S., ROSSI, S. E., MOATTI, N., ZOMPIT, M. D. M., XUE, Y., NG, T. F., ÁLVAREZ-QUILÓN, A., DESJARDINS, J., BHASKARAN, V., MARTINO, G., SETIAPUTRA, D., NOORDERMEER, S. M., OHSUMI, T. K., HUSTEDT, N., SZILARD, R. K., CHAUDHARY, N., MUNRO, M., VELOSO, A., MELO, H., YIN, S. Y., PAPP, R., YOUNG, J. T. F., ZINDA, M., STUCKI, M. & DUROCHER, D. 2021. The CIP2A–TOPBP1 axis safeguards chromosome stability and is a synthetic lethal target for BRCA-mutated cancer. *Nature Cancer,* 2**,** 1357--1371.

ADAMOVICH, A. I., BANERJEE, T., WINGO, M., DUNCAN, K., NING, J., RODRIGUES, F. M., HUANG, K.-L., LEE, C., CHEN, F., DING, L. & PARVIN, J. D. 2019. Functional analysis of BARD1 missense variants in homology-directed repair and damage sensitivity. *PLoS Genetics,* 15**,** e1008049.

ADAMS, K. E., MEDHURST, A. L., DART, D. A. & LAKIN, N. D. 2006. Recruitment of ATR to sites of ionising radiation-induced DNA damage requires ATM and components of the MRN protein complex. *Oncogene,* 25**,** 3894--3904.

ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. *Nature Methods,* 7**,** 248--249.

AGOULNIK, A. I., LU, B., ZHU, Q., TRUONG, C., TY, M. T., ARANGO, N., CHADA, K. K. & BISHOP, C. E. 2002. A novel gene, Pog, is necessary for primordial germ cell proliferation in the mouse and underlies the germ cell deficient mutation, gcd. *Human Molecular Genetics,* 11**,** 3047--3053.

AHN, J. Y., SCHWARZ, J. K., PIWNICA-WORMS, H. & CANMAN, C. E. 2000. Threonine 68 phosphorylation by ataxia telangiectasia mutated is required for efficient activation of Chk2 in response to ionizing radiation. *Cancer research,* 60**,** 5934--6.

AHNESORG, P., SMITH, P. & JACKSON, S. P. 2006. XLF Interacts with the XRCC4-DNA Ligase IV Complex to Promote DNA Nonhomologous End-Joining. *Cell,* 124**,** 301--313.

AKIMOV, V., BARRIO-HERNANDEZ, I., HANSEN, S. V. F., HALLENBORG, P., PEDERSEN, A.-K., BEKKER-JENSEN, D. B., PUGLIA, M., CHRISTENSEN, S. D. K., VANSELOW, J. T., NIELSEN, M. M.,

KRATCHMAROVA, I., KELSTRUP, C. D., OLSEN, J. V. & BLAGOEV, B. 2018. UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites. *Nature Structural \& Molecular Biology,* 25**,** 631--640.

ALCÓN, P., SHAKEEL, S., CHEN, Z. A., RAPPSILBER, J., PATEL, K. J. & PASSMORE, L. A. 2020. FANCD2–FANCI is a clamp stabilized on DNA by monoubiquitination of FANCD2 during DNA repair. *Nature Structural \& Molecular Biology,* 27**,** 240--248.

ALI, A. M., KIRBY, M., JANSEN, M., LACH, F. P., SCHULTE, J., SINGH, T. R., BATISH, S. D., AUERBACH, A. D., WILLIAMS, D. A. & MEETEI, A. R. 2009. Identification and characterization of mutations in FANCL gene: A second case of Fanconi anemia belonging to FA-L complementation group. *Human Mutation,* 30**,** E761--E770.

ALPI, A. F., PACE, P. E., BABU, M. M. & PATEL, K. J. 2008. Mechanistic Insight into Site-Restricted Monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. *Molecular Cell,* 32**,** 767--777.

ALT, FREDERICK W., ZHANG, Y., MENG, F.-L., GUO, C. & SCHWER, B. 2013. Mechanisms of Programmed DNA Lesions and Genomic Instability in the Immune System. *Cell,* 152**,** 417--429.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology,* 215**,** 403--410.

ALVAREZ, K., HURTADO, C., HEVIA, M. A., WIELANDT, A. M., FUENTE, M. D. L., CHURCH, J., CARVALLO, P. & LÓPEZ-KÖSTNER, F. 2010. Spectrum of MLH1 and MSH2 Mutations in Chilean Families With Suspected Lynch Syndrome. *Diseases of the Colon \& Rectum,* 53**,** 450--459.

AMENDOLA, L. M., JARVIK, G. P., LEO, M. C., MCLAUGHLIN, H. M., AKKARI, Y., AMARAL, M. D., BERG, J. S., BISWAS, S., BOWLING, K. M., CONLIN, L. K., COOPER, G. M., DORSCHNER, M. O., DULIK, M. C., GHAZANI, A. A., GHOSH, R., GREEN, R. C., HART, R., HORTON, C., JOHNSTON, J. J., LEBO, M. S., MILOSAVLJEVIC, A., OU, J., PAK, C. M., PATEL, R. Y., PUNJ, S., RICHARDS, C. S., SALAMA, J., STRANDE, N. T., YANG, Y., PLON, S. E., BIESECKER, L. G. & REHM, H. L. 2016. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *The American Journal of Human Genetics,* 98**,** 1067--1076.

AMEZIANE, N., SIE, D., DENTRO, S., ARIYUREK, Y., KERKHOVEN, L., JOENJE, H., DORSMAN, J. C., YLSTRA, B., GILLE, J. J. P., SISTERMANS, E. A. & WINTER, J. P. D. 2012. Diagnosis of Fanconi Anemia: Mutation Analysis by Next-Generation Sequencing. *Anemia,* 2012**,** 132856.

ANAND, R., RANJHA, L., CANNAVO, E. & CEJKA, P. 2016. Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. *Molecular Cell,* 64**,** 940--950.

ANDERSEN, S. D., LIBERTI, S. E., LÜTZEN, A., DROST, M., BERNSTEIN, I., NILBERT, M., DOMINGUEZ, M., NYSTRÖM, M., HANSEN, T. V. O., CHRISTOFFERSEN, J. W., JÄGER, A. C., WIND, N. D., NIELSEN, F. C., TØRRING, P. M. & RASMUSSEN, L. J. 2012. Functional characterization of MLH1 missense variants identified in lynch syndrome patients. *Human Mutation,* 33**,** 1647--1655.

ANDRES, S. N., MODESTI, M., TSAI, C. J., CHU, G. & JUNOP, M. S. 2007. Crystal Structure of Human XLF: A Twist in Nonhomologous DNA End-Joining. *Molecular Cell,* 28**,** 1093--1101.

ANDRES, S. N., VERGNES, A., RISTIC, D., WYMAN, C., MODESTI, M. & JUNOP, M. 2012. A human XRCC4–XLF complex bridges DNA. *Nucleic Acids Research,* 40**,** 1868--1878.

ANTONY, E., KHUBCHANDANI, S., CHEN, S. & HINGORANI, M. M. 2006. Contribution of Msh2 and Msh6 subunits to the asymmetric ATPase and DNA mismatch binding activities of Saccharomyces cerevisiae Msh2–Msh6 mismatch repair protein. *DNA Repair,* 5**,** 153--162.

ANZALONE, A. V., RANDOLPH, P. B., DAVIS, J. R., SOUSA, A. A., KOBLAN, L. W., LEVY, J. M., CHEN, P. J., WILSON, C., NEWBY, G. A., RAGURAM, A. & LIU, D. R. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature,* 576**,** 149--157.

ARGUESO, J. L., KIJAS, A. W., SARIN, S., HECK, J., WAASE, M. & ALANI, E. 2003. Systematic Mutagenesis of the Saccharomyces cerevisiae MLH1 Gene Reveals Distinct Roles for Mlh1p in Meiotic Crossing Over and in Vegetative and Meiotic Mismatch Repair. *Molecular and Cellular Biology,* 23**,** 873--886.

ARGUESO, J. L., WANAT, J., GEMICI, Z. & ALANI, E. 2004. Competing Crossover Pathways Act During Meiosis in Saccharomyces cerevisiae. *Genetics,* 168**,** 1805--1816.

ARNOLD, S., BUCHANAN, D. D., BARKER, M., JASKOWSKI, L., WALSH, M. D., BIRNEY, G., WOODS, M. O., HOPPER, J. L., JENKINS, M. A., BROWN, M. A., TAVTIGIAN, S. V., GOLDGAR, D. E., YOUNG, J. P. & SPURDLE, A. B. 2009. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. *Human Mutation,* 30**,** 757--770.

ARORA, S., HUWE, P. J., SIKDER, R., SHAH, M., BROWNE, A. J., LESH, R., NICOLAS, E., DESHPANDE, S., HALL, M. J., DUNBRACK, R. L. & GOLEMIS, E. A. 2017. Functional analysis of rare variants in mismatch repair proteins augments results from computation-based predictive methods. *Cancer Biology \& Therapy,* 18**,** 519--533.

AU, K. G., WELSH, K. & MODRICH, P. 1992. Initiation of methyl-directed mismatch repair. *Journal of Biological Chemistry,* 267**,** 12142--12148.

BACANI, J., ZWINGERMAN, R., NICOLA, N. D., SPENCER, S., WEGRYNOWSKI, T., MITCHELL, K., HAY, K., REDSTON, M., HOLOWATY, E., HUNTSMAN, D., POLLETT, A., RIDDELL, R. & GALLINGER, S. 2005. Tumor Microsatellite Instability in Early Onset Gastric Cancer. *The Journal of Molecular Diagnostics,* 7**,** 465--477.

BAJAR, B. T., WANG, E. S., LAM, A. J., KIM, B. B., JACOBS, C. L., HOWE, E. S., DAVIDSON, M. W., LIN, M. Z. & CHU, J. 2016. Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Scientific Reports,* 6**,** 20889.

BAKER, S. M., PLUG, A. W., PROLLA, T. A., BRONNER, C. E., HARRIS, A. C., YAO, X., CHRISTIE, D.-M., MONELL, C., ARNHEIM, N., BRADLEY, A., ASHLEY, T. & LISKAY, R. M. 1996. Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over. *Nature Genetics,* 13**,** 336--342.

BALMUS, G., BARROS, A. C., WIJNHOVEN, P. W. G., LESCALE, C., HASSE, H. L., BOROVIAK, K., SAGE, C. L., DOE, B., SPEAK, A. O., GALLI, A., JACOBSEN, M., DERIANO, L., ADAMS, D. J., BLACKFORD, A. N. & JACKSON, S. P. 2016. Synthetic lethality between PAXX and XLF in mammalian development. *Genes \& Development,* 30**,** 2152--2157.

BALMUS, G., PILGER, D., COATES, J., DEMIR, M., SCZANIECKA-CLIFT, M., BARROS, A. C., WOODS, M., FU, B., YANG, F., CHEN, E., OSTERMAIER, M., STANKOVIC, T., PONSTINGL, H., HERZOG, M., YUSA, K., MARTINEZ, F. M., DURANT, S. T., GALANTY, Y., BELI, P., ADAMS, D. J., BRADLEY, A., METZAKOPIAN, E., FORMENT, J. V. & JACKSON, S. P. 2019. ATM orchestrates the DNA-damage response to counter toxic non-homologous end-joining at broken replication forks. *Nature Communications,* 10**,** 87.

BAMSHAD, M. J., NG, S. B., BIGHAM, A. W., TABOR, H. K., EMOND, M. J., NICKERSON, D. A. & SHENDURE, J. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics,* 12**,** 745--755.

BAN, C. & YANG, W. 1998. Crystal Structure and ATPase Activity of MutL Implications for DNA Repair and Mutagenesis. *Cell,* 95**,** 541--552.

BANKMANN, M., PRAKASH, L. & PRAKASH, S. 1992. YeastRAD14 and human xeroderma pigmentosum group A DNA-repair genes encode homologous proteins. *Nature,* 355**,** 555--558.

BARETTI, M. & LE, D. T. 2018. DNA mismatch repair in cancer. *Pharmacology \& Therapeutics,* 189**,** 45--62.

BARNABA, N. & LAROCQUE, J. R. 2021. Targeting cell cycle regulation via the G2-M checkpoint for synthetic lethality in melanoma. *Cell Cycle,* 20**,** 1041--1051.

BARNETSON, R. A., CARTWRIGHT, N., VLIET, A. V., HAQ, N., DREW, K., FARRINGTON, S., WILLIAMS, N., WARNER, J., CAMPBELL, H., PORTEOUS, M. E. & DUNLOP, M. G. 2008. Classification of ambiguous mutations in DNA mismatch repair genes identified in a population-based study of colorectal cancer. *Human Mutation,* 29**,** 367--374.

BASKAR, R., DAI, J., WENLONG, N., YEO, R. & YEOH, K.-W. 2014. Biological response of cancer cells to radiation treatment. *Frontiers in Molecular Biosciences,* 1**,** 24.

BASS, T. E., LUZWICK, J. W., KAVANAUGH, G., CARROLL, C., DUNGRAWALA, H., GLICK, G. G., FELDKAMP, M. D., PUTNEY, R., CHAZIN, W. J. & CORTEZ, D. 2016. ETAA1 acts at stalled replication forks to maintain genome integrity. *Nature Cell Biology,* 18**,** 1185--1195.

BAUDAT, F., IMAI, Y. & MASSY, B. D. 2013. Meiotic recombination in mammals: localization and regulation. *Nature Reviews Genetics,* 14**,** 794--806.

BAUMANN, P. & WEST, S. C. 1998. Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. *Trends in Biochemical Sciences,* 23**,** 247--251.

BEARD, W. A., HORTON, J. K., PRASAD, R. & WILSON, S. H. 2019. Eukaryotic Base Excision Repair: New Approaches Shine Light on Mechanism. *Annual Review of Biochemistry,* 88**,** 137--162.

BĘBENEK, A. & ZIUZIA-GRACZYK, I. 2018. Fidelity of DNA replication—a matter of proofreading. *Current Genetics,* 64**,** 985--996.

BEBENEK, K. & KUNKEL, T. A. 2004. Functions of DNA Polymerases. *Advances in Protein Chemistry,* 69**,** 137--165.

BECK, H., NÄHSE-KUMPF, V., LARSEN, M. S. Y., O'HANLON, K. A., PATZKE, S., HOLMBERG, C., MEJLVANG, J., GROTH, A., NIELSEN, O., SYLJUÅSEN, R. G. & SØRENSEN, C. S. 2012. Cyclin-Dependent Kinase Suppression by WEE1 Kinase Protects the Genome through Control of Replication Initiation and Nucleotide Consumption. *Molecular and Cellular Biology,* 32**,** 4226--4236.

BECK, H., NÄHSE, V., LARSEN, M. S. Y., GROTH, P., CLANCY, T., LEES, M., JØRGENSEN, M., HELLEDAY, T., SYLJUÅSEN, R. G. & SØRENSEN, C. S. 2010. Regulators of cyclin-dependent kinases are crucial for maintaining genome integrity in S phase. *The Journal of Cell Biology,* 188**,** 629--638.

BECKER, J. R., CLIFFORD, G., BONNET, C., GROTH, A., WILSON, M. D. & CHAPMAN, J. R. 2021. BARD1 reads H2A lysine 15 ubiquitination to direct homologous recombination. *Nature,* 596**,** 433--437.

BEE, L., NASCA, A., ZANOLINI, A., CENDRON, F., D'ADAMO, P., COSTA, R., LAMPERTI, C., CELOTTI, L., GHEZZI, D. & ZEVIANI, M. 2015. A nonsense mutation of human XRCC4 is associated with adult-onset progressive encephalocardiomyopathy. *EMBO Molecular Medicine,* 7**,** 918--929.

BEN-YEHOYADA, M., WANG, L. C., KOZEKOV, I. D., RIZZO, C. J., GOTTESMAN, M. E. & GAUTIER, J. 2009. Checkpoint Signaling from a Single DNA Interstrand Crosslink. *Molecular Cell,* 35**,** 704--715.

BENSON, A. B., VENOOK, A. P., AL-HAWARY, M. M., ARAIN, M. A., CHEN, Y.-J., CIOMBOR, K. K., COHEN, S., COOPER, H. S., DEMING, D., FARKAS, L., GARRIDO-LAGUNA, I., GREM, J. L., GUNN, A., HECHT, J. R., HOFFE, S., HUBBARD, J., HUNT, S., JOHUNG, K. L., KIRILCUK, N., KRISHNAMURTHI, S., MESSERSMITH, W. A., MEYERHARDT, J., MILLER, E. D., MULCAHY, M. F., NURKIN, S., OVERMAN, M. J., PARIKH, A., PATEL, H., PEDERSEN, K., SALTZ, L., SCHNEIDER, C., SHIBATA, D., SKIBBER, J. M., SOFOCLEOUS, C. T., STOFFEL, E. M., STOTSKY-HIMELFARB, E., WILLETT, C. G., GREGORY, K. M. & GURSKI, L. A. 2021. Colon Cancer, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network,* 19**,** 329--359.

BENTLEY, N. J., HOLTZMAN, D. A., FLAGGS, G., KEEGAN, K. S., DEMAGGIO, A., FORD, J. C., HOEKSTRA, M. & CARR, A. M. 1996. The Schizosaccharomyces pombe rad3 checkpoint gene. *The EMBO Journal,* 15**,** 6641--6651.

BERNSTEIN, N. K., WILLIAMS, R. S., RAKOVSZKY, M. L., CUI, D., GREEN, R., KARIMI-BUSHERI, F., MANI, R. S., GALICIA, S., KOCH, C. A., CASS, C. E., DUROCHER, D., WEINFELD, M. & GLOVER, J. N. M.

2005. The Molecular Architecture of the Mammalian DNA Repair Enzyme, Polynucleotide Kinase. *Molecular Cell,* 17**,** 657--670.

BESENBACHER, S., LIU, S., IZARZUGAZA, J. M. G., GROVE, J., BELLING, K., BORK-JENSEN, J., HUANG, S., ALS, T. D., LI, S., YADAV, R., RUBIO-GARCÍA, A., LESCAI, F., DEMONTIS, D., RAO, J., YE, W., MAILUND, T., FRIBORG, R. M., PEDERSEN, C. N. S., XU, R., SUN, J., LIU, H., WANG, O., CHENG, X., FLORES, D., RYDZA, E., RAPACKI, K., SØRENSEN, J. D., CHMURA, P., WESTERGAARD, D., DWORZYNSKI, P., SØRENSEN, T. I. A., LUND, O., HANSEN, T., XU, X., LI, N., BOLUND, L., PEDERSEN, O., EIBERG, H., KROGH, A., BØRGLUM, A. D., BRUNAK, S., KRISTIANSEN, K., SCHIERUP, M. H., WANG, J., GUPTA, R., VILLESEN, P. & RASMUSSEN, S. 2015. Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nature Communications,* 6**,** 5969.

BIAN, Y., SONG, C., CHENG, K., DONG, M., WANG, F., HUANG, J., SUN, D., WANG, L., YE, M. & ZOU, H. 2014. An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *Journal of Proteomics,* 96**,** 253--262.

BIESECKER, L. G., HARRISON, S. M. & GROUP, T. C. S. V. I. W. 2018. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genetics in Medicine,* 20**,** 1687--1688.

BLACKFORD, A. N. & 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., NIEMINUSZCZY, J., SCHWAB, R. A., GALANTY, Y., JACKSON, S. P. & NIEDZWIEDZ, W. 2015. TopBP1 interacts with BLM to maintain genome stability but is dispensable for preventing BLM degradation. *Molecular cell,* 57**,** 1133--41.

BLACKFORD, A. N. & STUCKI, M. 2020. How Cells Respond to DNA Breaks in Mitosis. *Trends in Biochemical Sciences,* 45**,** 321--331.

BLASI, M. F., VENTURA, I., AQUILINA, G., DEGAN, P., BERTARIO, L., BASSI, C., RADICE, P. & BIGNAMI, M. 2006. A Human Cell-Based Assay to Evaluate the Effects of Alterations in the MLH1 Mismatch Repair Gene. *Cancer Research,* 66**,** 9036--9044.

BLOCK, W. D. & LEES-MILLER, S. P. 2005. Putative homologues of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) and other components of the non-homologous end joining machinery in Dictyostelium discoideum. *DNA Repair,* 4**,** 1061--1065.

BOECKMANN, B., BAIROCH, A., APWEILER, R., BLATTER, M.-C., ESTREICHER, A., GASTEIGER, E., MARTIN, M. J., MICHOUD, K., O'DONOVAN, C., PHAN, I., PILBOUT, S. & SCHNEIDER, M. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research,* 31**,** 365--370.

BOEING, S., WILLIAMSON, L., ENCHEVA, V., GORI, I., SAUNDERS, R. E., INSTRELL, R., AYGÜN, O., RODRIGUEZ-MARTINEZ, M., WEEMS, J. C., KELLY, G. P., CONAWAY, J. W., CONAWAY, R. C., STEWART, A., HOWELL, M., SNIJDERS, A. P. & SVEJSTRUP, J. Q. 2016. Multiomic Analysis of the UV-Induced DNA Damage Response. *Cell Reports,* 15**,** 1597--1610.

BOERSMA, V., MOATTI, N., SEGURA-BAYONA, S., PEUSCHER, M. H., TORRE, J. V. D., WEVERS, B. A., ORTHWEIN, A., DUROCHER, D. & JACOBS, J. J. L. 2015. MAD2L2 controls DNA repair at telomeres and DNA breaks by inhibiting 5' end resection. *Nature,* 521**,** 537--540.

BONAGURA, T. W., BABISCHKIN, J. S., PEPE, G. J. & ALBRECHT, E. D. 2022. Estrogen Receptors, Methods and Protocols. *Methods in Molecular Biology,* 2418**,** 77--93.

BOOHER, R. N., HOLMAN, P. S. & FATTAEY, A. 1997. Human Myt1 Is a Cell Cycle-regulated Kinase That Inhibits Cdc2 but Not Cdk2 Activity*. *Journal of Biological Chemistry,* 272**,** 22300--22306.

BOOM, V. V. D., CITTERIO, E., HOOGSTRATEN, D., ZOTTER, A., EGLY, J.-M., CAPPELLEN, W. A. V., HOEIJMAKERS, J. H. J., HOUTSMULLER, A. B. & VERMEULEN, W. 2004. DNA damage stabilizes interaction of CSB with the transcription elongation machinery. *The Journal of Cell Biology,* 166**,** 27--36.

BOONEN, R. A. C. M., RODRIGUE, A., STOEPKER, C., WIEGANT, W. W., VROLING, B., SHARMA, M., ROTHER, M. B., CELOSSE, N., VREESWIJK, M. P. G., COUCH, F., SIMARD, J., DEVILEE, P.,

MASSON, J.-Y. & ATTIKUM, H. V. 2019. Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. *Nature Communications,* 10**,** 5296.
BORRÀS, E., PINEDA, M., BRIEGER, A., HINRICHSEN, I., GÓMEZ, C., NAVARRO, M., BALMAÑA, J., CAJAL, T. R. Y., TORRES, A., BRUNET, J., BLANCO, I., PLOTZ, G., LÁZARO, C. & CAPELLÁ, G. 2012. Comprehensive functional assessment of MLH1 variants of unknown significance. *Human Mutation,* 33**,** 1576--1588.
BOTUYAN, M. V., LEE, J., WARD, I. M., KIM, J.-E., THOMPSON, J. R., CHEN, J. & 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.
BOUVET, D., BODO, S., MUNIER, A., GUILLERM, E., BERTRAND, R., COLAS, C., DUVAL, A., COULET, F. & MULERIS, M. 2019. Methylation Tolerance-Based Functional Assay to Assess Variants of Unknown Significance in the MLH1 and MSH2 Genes and Identify Patients With Lynch Syndrome. *Gastroenterology,* 157**,** 421--431.
BOUWMAN, P., ALY, A., ESCANDELL, J. M., PIETERSE, M., BARTKOVA, J., GULDEN, H. V. D., HIDDINGH, S., THANASOULA, M., KULKARNI, A., YANG, Q., HAFFTY, B. G., TOMMISKA, J., BLOMQVIST, C., DRAPKIN, R., ADAMS, D. J., NEVANLINNA, H., BARTEK, J., TARSOUNAS, M., GANESAN, S. & JONKERS, J. 2010. 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. *Nature Structural \& Molecular Biology,* 17**,** 688--695.
BOUWMAN, P., GULDEN, H. V. D., HEIJDEN, I. V. D., DROST, R., KLIJN, C. N., PRASETYANTI, P., PIETERSE, M., WIENTJENS, E., SEIBLER, J., HOGERVORST, F. B. L. & JONKERS, J. 2013. A High-Throughput Functional Complementation Assay for Classification of BRCA1 Missense Variants. *Cancer Discovery,* 3**,** 1142--1155.
BRATTAIN, M. G., FINE, W. D., KHALED, F. M., THOMPSON, J. & BRATTAIN, D. E. 1981. Heterogeneity of malignant cells from a human colonic carcinoma. *Cancer research,* 41**,** 1751--6.
BRIEGER, A., PLOTZ, G., RAEDLE, J., WEBER, N., BAUM, W., CASPARY, W. F., ZEUZEM, S. & TROJAN, J. 2005. Characterization of the nuclear import of human MutLα. *Molecular Carcinogenesis,* 43**,** 51--58.
BRIEGER, A., TROJAN, J., RAEDLE, J., PLOTZ, G. & ZEUZEM, S. 2002. Transient mismatch repair gene transfection for functional analysis of genetic hMLH1 and hMSH2 variants. *Gut,* 51**,** 677.
BRNICH, S. E., TAYOUN, A. N. A., COUCH, F. J., CUTTING, G. R., GREENBLATT, M. S., HEINEN, C. D., KANAVY, D. M., LUO, X., MCNULTY, S. M., STARITA, L. M., TAVTIGIAN, S. V., WRIGHT, M. W., HARRISON, S. M., BIESECKER, L. G., BERG, J. S., TAYOUN, A. N. A., BERG, J. S., BIESECKER, L. G., BRENNER, S. E., CUTTING, G. R., ELLARD, S., GREENBLATT, M. S., HARRISON, S. M., KARBASSI, I., KARCHIN, R., MESTER, J. L., O'DONNELL-LURIA, A., PESARAN, T., PLON, S. E., REHM, H., TAVTIGIAN, S. V. & TOPPER, S. 2019. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Medicine,* 12**,** 3.
BRONNER, C. E., BAKER, S. M., MORRISON, P. T., WARREN, G., SMITH, L. G., LESCOE, M. K., KANE, M., EARABINO, C., LIPFORD, J., LINDBLOM, A., TANNERGÅRD, P., BOLLAG, R. J., GODWIN, A. R., WARD, D. C., NORDENSKJ⃝LD, M., FISHEL, R., KOLODNER, R. & LISKAY, R. M. 1994. Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer. *Nature,* 368**,** 258--261.
BROOKS, K., OAKES, V., EDWARDS, B., RANALL, M., LEO, P., PAVEY, S., PINDER, A., BEAMISH, H., MUKHOPADHYAY, P., LAMBIE, D. & GABRIELLI, B. 2013. A potent Chk1 inhibitor is selectively cytotoxic in melanomas with high levels of replicative stress. *Oncogene,* 32**,** 788--796.
BROSH, R. M., BELLANI, M., LIU, Y. & SEIDMAN, M. M. 2017. Fanconi Anemia: A DNA repair disorder characterized by accelerated decline of the hematopoietic stem cell compartment and other features of aging. *Ageing Research Reviews,* 33**,** 67--75.
BROUWER, I., SITTERS, G., CANDELLI, A., HEEREMA, S. J., HELLER, I., DE, A. J. M., ZHANG, H., NORMANNO, D., MODESTI, M., PETERMAN, E. J. G. & WUITE, G. J. L. 2016. Sliding sleeves of XRCC4–XLF bridge DNA and connect fragments of broken DNA. *Nature,* 535**,** 566--569.

BROWN, A. L., LEE, C.-H., SCHWARZ, J. K., MITIKU, N., PIWNICA-WORMS, H. & CHUNG, J. H. 1999. A human Cds1-related kinase that functions downstream of ATM protein in the cellular response to DNA damage. *Proceedings of the National Academy of Sciences,* 96**,** 3745--3750.

BROWN, T. C. & JIRICNY, J. 1988. Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. *Cell,* 54**,** 705--711.

BRÜCKNER, A., POLGE, C., LENTZE, N., AUERBACH, D. & SCHLATTNER, U. 2009. Yeast Two-Hybrid, a Powerful Tool for Systems Biology. *International Journal of Molecular Sciences,* 10**,** 2763--2788.

BRUIN, C. D., MERICQ, V., ANDREW, S. F., DUYVENVOORDE, H. A. V., VERKAIK, N. S., LOSEKOOT, M., POROLLO, A., GARCIA, H., KUANG, Y., HANSON, D., CLAYTON, P., GENT, D. C. V., WIT, J. M., HWA, V. & DAUBER, A. 2015. An XRCC4 Splice Mutation Associated With Severe Short Stature, Gonadal Failure, and Early-Onset Metabolic Syndrome. *The Journal of Clinical Endocrinology \& Metabolism,* 100**,** E789--E798.

BRYANS, M., VALENZANO, M. C. & STAMATO, T. D. 1999. Absence of DNA ligase IV protein in XR-1 cells: evidence for stabilization by XRCC4. *Mutation Research/DNA Repair,* 433**,** 53--58.

BRYANT, H. E., SCHULTZ, N., THOMAS, H. D., PARKER, K. M., FLOWER, D., LOPEZ, E., KYLE, S., MEUTH, M., CURTIN, N. J. & HELLEDAY, T. 2005. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature,* 434**,** 913--917.

BUCHKO, G. W., DAUGHDRILL, G. W., LORIMIER, R. D., K., S. R. B., ISERN, N. G., LINGBECK, J. M., TAYLOR, J.-S., WOLD, M. S., GOCHIN, M., SPICER, L. D., LOWRY, D. F. & KENNEDY, M. A. 1999. Interactions of Human Nucleotide Excision Repair Protein XPA with DNA and RPA70ΔC327: Chemical Shift Mapping and 15N NMR Relaxation Studies †. *Biochemistry,* 38**,** 15116--15128.

BUCK, D., MALIVERT, L., CHASSEVAL, R. D., BARRAUD, A., FONDANÈCHE, M.-C., SANAL, O., PLEBANI, A., STÉPHAN, J.-L., HUFNAGEL, M., DEIST, F. L., FISCHER, A., DURANDY, A., VILLARTAY, J.-P. D. & REVY, P. 2006. Cernunnos, a Novel Nonhomologous End-Joining Factor, Is Mutated in Human Immunodeficiency with Microcephaly. *Cell,* 124**,** 287--299.

BUDZOWSKA, M., GRAHAM, T. G. W., SOBECK, A., WAGA, S. & WALTER, J. C. 2015. Regulation of the Rev1–pol ζ complex during bypass of a DNA interstrand cross-link. *The EMBO Journal,* 34**,** 1971--1985.

BUISSON, R., BOISVERT, JESSICA L., BENES, CYRIL H. & ZOU, L. 2015. Distinct but Concerted Roles of ATR, DNA-PK, and Chk1 in Countering Replication Stress during S Phase. *Molecular Cell,* 59**,** 1011--1024.

BUNICK, C. G., MILLER, M. R., FULLER, B. E., FANNING, E. & CHAZIN, W. J. 2006. Biochemical and Structural Domain Analysis of Xeroderma Pigmentosum Complementation Group C Protein †. *Biochemistry,* 45**,** 14965--14979.

BUNTING, S. F., CALLÉN, E., WONG, N., CHEN, H.-T., POLATO, F., GUNN, A., BOTHMER, A., FELDHAHN, N., FERNANDEZ-CAPETILLO, O., CAO, L., XU, X., DENG, C.-X., FINKEL, T., NUSSENZWEIG, M., STARK, J. M. & NUSSENZWEIG, A. 2010. 53BP1 Inhibits Homologous Recombination in Brca1-Deficient Cells by Blocking Resection of DNA Breaks. *Cell,* 141**,** 243--254.

BURCKHARDT, C. J., MINNA, J. D. & DANUSER, G. 2021. Co-immunoprecipitation and semi-quantitative immunoblotting for the analysis of protein-protein interactions. *STAR Protocols,* 2**,** 100644.

BURDETT, V., BAITINGER, C., VISWANATHAN, M., LOVETT, S. T. & MODRICH, P. 2001. In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proceedings of the National Academy of Sciences,* 98**,** 6765--6770.

BURKE, W., DALY, M., GARBER, J., BOTKIN, J., KAHN, M. J. E., LYNCH, P., MCTIERNAN, A., OFFIT, K., PERLMAN, J., PETERSEN, G., THOMSON, E. & VARRICCHIO, C. 1997. Recommendations for Follow-up Care of Individuals With an Inherited Predisposition to Cancer: II. BRCA1 and BRCA2. *Jama,* 277**,** 997--1003.

BUTKIEWICZ, D., RUSIN, M., HARRIS, C. C. & CHORAZY, M. 2000. Identification of four single nucleotide polymorphisms in DNA repair genes: XPA and XPB (ERCC3) in Polish population. *Human Mutation,* 15**,** 577--578.

BZYMEK, M., THAYER, N. H., OH, S. D., KLECKNER, N. & HUNTER, N. 2010. Double Holliday junctions are intermediates of DNA break repair. *Nature,* 464**,** 937--941.

CADET, J., DOUKI, T. & RAVANAT, J.-L. 2011. Measurement of oxidatively generated base damage in cellular DNA. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 711**,** 3--12.

CALDECOTT, K. W. 2014. Protein ADP-ribosylation and the cellular response to DNA strand breaks. *DNA Repair,* 19**,** 108--113.

CALLÉN, E., JANKOVIC, M., WONG, N., ZHA, S., CHEN, H.-T., DIFILIPPANTONIO, S., VIRGILIO, M. D., HEIDKAMP, G., ALT, F. W., NUSSENZWEIG, A. & NUSSENZWEIG, M. 2009. Essential Role for DNA-PKcs in DNA Double-Strand Break Repair and Apoptosis in ATM-Deficient Lymphocytes. *Molecular Cell,* 34**,** 285--297.

CALLIS, J. 2014. The Ubiquitination Machinery of the Ubiquitin System. *The Arabidopsis Book,* 12**,** e0174.

CAMENISCH, U., DIP, R., SCHUMACHER, S. B., SCHULER, B. & NAEGELI, H. 2006. Recognition of helical kinks by xeroderma pigmentosum group A protein triggers DNA excision repair. *Nature Structural \& Molecular Biology,* 13**,** 278--284.

CAMPBELL, C. S., HOMBAUER, H., SRIVATSAN, A., BOWEN, N., GRIES, K., DESAI, A., PUTNAM, C. D. & KOLODNER, R. D. 2014. Mlh2 Is an Accessory Factor for DNA Mismatch Repair in Saccharomyces cerevisiae. *PLoS Genetics,* 10**,** e1004327.

CANMAN, C. E., LIM, D.-S., CIMPRICH, K. A., TAYA, Y., TAMAI, K., SAKAGUCHI, K., APPELLA, E., KASTAN, M. B. & SILICIANO, J. D. 1998. Activation of the ATM Kinase by Ionizing Radiation and Phosphorylation of p53. *Science,* 281**,** 1677--1679.

CANNAVO, E. & CEJKA, P. 2014. Sae2 promotes dsDNA endonuclease activity within Mre11–Rad50–Xrs2 to resect DNA breaks. *Nature,* 514**,** 122--125.

CANNAVO, E., SANCHEZ, A., ANAND, R., RANJHA, L., HUGENER, J., ADAM, C., ACHARYA, A., WEYLAND, N., ARAN-GUIU, X., CHARBONNIER, J.-B., HOFFMANN, E. R., BORDE, V., MATOS, J. & CEJKA, P. 2020. Regulation of the MLH1–MLH3 endonuclease in meiosis. *Nature,* 586**,** 618--622.

CAO, L., XU, X., BUNTING, S. F., LIU, J., WANG, R.-H., CAO, L. L., WU, J. J., PENG, T.-N., CHEN, J., NUSSENZWEIG, A., DENG, C.-X. & FINKEL, T. 2009. A Selective Requirement for 53BP1 in the Biological Response to Genomic Instability Induced by Brca1 Deficiency. *Molecular Cell,* 35**,** 534--541.

CARRASSA, L., CHILÀ, R., LUPI, M., RICCI, F., CELENZA, C., MAZZOLETTI, M., BROGGINI, M. & DAMIA, G. 2012. Combined inhibition of Chk1 and Wee1: In vitro synergistic effect translates to tumor growth inhibition in vivo. *Cell Cycle,* 11**,** 2507--2517.

CARTER, T., VANCUROVÁ, I., SUN, I., LOU, W. & DELEON, S. 1990. A DNA-activated protein kinase from HeLa cell nuclei. *Molecular and Cellular Biology,* 10**,** 6460--6471.

CECCALDI, R., LIU, J. C., AMUNUGAMA, R., HAJDU, I., PRIMACK, B., PETALCORIN, M. I. R., O'CONNOR, K. W., KONSTANTINOPOULOS, P. A., ELLEDGE, S. J., BOULTON, S. J., YUSUFZAI, T. & D'ANDREA, A. D. 2015. Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature,* 518**,** 258--262.

CECCALDI, R., SARANGI, P. & D'ANDREA, A. D. 2016. The Fanconi anaemia pathway: new players and new functions. *Nature Reviews Molecular Cell Biology,* 17**,** 337--349.

CEJKA, P., STOJIC, L., MOJAS, N., RUSSELL, A. M., HEINIMANN, K., CANNAVÓ, E., PIETRO, M. D., MARRA, G. & JIRICNY, J. 2003. Methylation-induced G2/M arrest requires a full complement of the mismatch repair protein hMLH1. *The EMBO Journal,* 22**,** 2245--2254.

CEJKA, P. & SYMINGTON, L. S. 2021. DNA End Resection: Mechanism and Control. *Annual Review of Genetics,* 55**,** 285--307.

CERAMI, E., GAO, J., DOGRUSOZ, U., GROSS, B. E., SUMER, S. O., AKSOY, B. A., JACOBSEN, A., BYRNE, C. J., HEUER, M. L., LARSSON, E., ANTIPIN, Y., REVA, B., GOLDBERG, A. P., SANDER, C. & SCHULTZ, N. 2012. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery,* 2**,** 401--404.

CERVELLI, T., LODOVICHI, S., BELLÈ, F. & GALLI, A. 2020. Yeast-based assays for the functional characterization of cancer-associated variants of human DNA repair genes. *Microb Cell,* 7**,** 162-174.

CHAN, E. M., SHIBUE, T., MCFARLAND, J. M., GAETA, B., GHANDI, M., DUMONT, N., GONZALEZ, A., MCPARTLAN, J. S., LI, T., ZHANG, Y., LIU, J. B., LAZARO, J.-B., GU, P., PIETT, C. G., APFFEL, A., ALI, S. O., DEASY, R., KESKULA, P., NG, R. W. S., ROBERTS, E. A., REZNICHENKO, E., LEUNG, L., ALIMOVA, M., SCHENONE, M., ISLAM, M., MARUVKA, Y. E., LIU, Y., ROPER, J., RAGHAVAN, S., GIANNAKIS, M., TSENG, Y.-Y., NAGEL, Z. D., D'ANDREA, A., ROOT, D. E., BOEHM, J. S., GETZ, G., CHANG, S., GOLUB, T. R., TSHERNIAK, A., VAZQUEZ, F. & BASS, A. J. 2019. WRN helicase is a synthetic lethal target in microsatellite unstable cancers. *Nature,* 568**,** 551--556.

CHAN, P. A., DURAISAMY, S., MILLER, P. J., NEWELL, J. A., MCBRIDE, C., BOND, J. P., RAEVAARA, T., OLLILA, S., NYSTRÖM, M., GRIMM, A. J., CHRISTODOULOU, J., OETTING, W. S. & GREENBLATT, M. S. 2007. Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). *Human Mutation,* 28**,** 683--693.

CHANG, S., BISWAS, K., MARTIN, B. K., STAUFFER, S. & SHARAN, S. K. 2009. Expression of human BRCA1 variants in mouse ES cells allows functional analysis of BRCA1 mutations. *Journal of Clinical Investigation,* 119**,** 3160--3171.

CHAO, E. C., VELASQUEZ, J. L., WITHERSPOON, M. S. L., ROZEK, L. S., PEEL, D., NG, P., GRUBER, S. B., WATSON, P., RENNERT, G., ANTON-CULVER, H., LYNCH, H. & LIPKIN, S. M. 2008. Accurate classification of MLH1/MSH2 missense variants with multivariate analysis of protein polymorphisms–mismatch repair (MAPP-MMR). *Human Mutation,* 29**,** 852--860.

CHAPMAN, J. R., BARRAL, P., VANNIER, J.-B., BOREL, V., STEGER, M., TOMAS-LOBA, A., SARTORI, ALESSANDRO A., ADAMS, IAN R., BATISTA, FACUNDO D. & BOULTON, SIMON 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. & 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., TAYLOR, MARTIN R. G. & BOULTON, SIMON J. 2012. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Molecular Cell,* 47**,** 497--510.

CHATTERJEE, N. & WALKER, G. C. 2017. Mechanisms of DNA damage, repair, and mutagenesis: DNA Damage and Repair. *Environmental and Molecular Mutagenesis,* 58**,** 235--263.

CHEN, S., LEE, L., NAILA, T., FISHBAIN, S., WANG, A., TOMKINSON, A. E., LEES-MILLER, S. P. & HE, Y. 2021. Structural basis of long-range to short-range synaptic transition in NHEJ. *Nature,* 593**,** 294--298.

CHEN, S., WANG, W., LEE, S., NAFA, K., LEE, J., ROMANS, K., WATSON, P., GRUBER, S. B., EUHUS, D., KINZLER, K. W., JASS, J., GALLINGER, S., LINDOR, N. M., CASEY, G., ELLIS, N., GIARDIELLO, F. M., OFFIT, K., PARMIGIANI, G. & REGISTRY, F. T. C. C. F. 2006. Prediction of Germline Mutations and Cancer Risk in the Lynch Syndrome. *Jama,* 296**,** 1479--1487.

CHEN, Z., YUAN, Y., CHEN, X., CHEN, J., LIN, S., LI, X. & DU, H. 2020. Systematic comparison of somatic variant calling performance among different sequencing depth and mutation frequency. *Scientific Reports,* 10**,** 3501.

CHENG, K. C., CAHILL, D. S., KASAI, H., NISHIMURA, S. & LOEB, L. A. 1992. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. *The Journal of biological chemistry,* 267**,** 166--72.

CHEUNG-ONG, K., GIAEVER, G. & NISLOW, C. 2013. DNA-Damaging Agents in Cancer Chemotherapy: Serendipity and Chemical Biology. *Chemistry \& Biology,* 20**,** 648--659.

CHI, X., LI, Y. & QIU, X. 2020. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology,* 160**,** 233--247.

CHOI, J. Y., PARK, J.-M., YI, J. M., LEEM, S.-H. & KANG, T.-H. 2015. Enhanced nucleotide excision repair capacity in lung cancer cells by preconditioning with DNA-damaging agents. *Oncotarget,* 6**,** 22575--22586.

CHOI, M., SCHOLL, U. I., JI, W., LIU, T., TIKHONOVA, I. R., ZUMBO, P., NAYIR, A., BAKKALOĞLU, A. I., ÖZEN, S., SANJAD, S., NELSON-WILLIAMS, C., FARHI, A., MANE, S. & LIFTON, R. P. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences,* 106**,** 19096--19101.

CHRISTENSEN, S., ROEST, B. V. D., BESSELINK, N., JANSSEN, R., BOYMANS, S., MARTENS, J. W. M., YASPO, M.-L., PRIESTLEY, P., KUIJK, E., CUPPEN, E. & HOECK, A. V. 2019. 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. *Nature Communications,* 10**,** 4571.

CHRISTIE, S. M., FIJEN, C. & ROTHENBERG, E. 2022. V(D)J Recombination: Recent Insights in Formation of the Recombinase Complex and Recruitment of DNA Repair Machinery. *Frontiers in Cell and Developmental Biology,* 10**,** 886718.

CIBULSKIS, K., LAWRENCE, M. S., CARTER, S. L., SIVACHENKO, A., JAFFE, D., SOUGNEZ, C., GABRIEL, S., MEYERSON, M., LANDER, E. S. & GETZ, G. 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology,* 31**,** 213--219.

CICCIA, A. & ELLEDGE, S. J. 2010. The DNA Damage Response: Making It Safe to Play with Knives. *Molecular Cell,* 40**,** 179--204.

CIMPRICH, K. A., SHIN, T. B., KEITH, C. T. & 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.

CLARK, A. B., VALLE, F., DROTSCHMANN, K., GARY, R. K. & KUNKEL, T. A. 2000. Functional Interaction of Proliferating Cell Nuclear Antigen with MSH2-MSH6 and MSH2-MSH3 Complexes*. *Journal of Biological Chemistry,* 275**,** 36498--36501.

CLAUSEN, A. R., ZHANG, S., BURGERS, P. M., LEE, M. Y. & KUNKEL, T. A. 2013. Ribonucleotide incorporation, proofreading and bypass by human DNA polymerase δ. *DNA Repair,* 12**,** 121--127.

CLEMENTS, P. M., BRESLIN, C., DEEKS, E. D., BYRD, P. J., JU, L., BIEGANOWSKI, P., BRENNER, C., MOREIRA, M.-C., TAYLOR, A. M. R. & 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.

CLIBY, W. A., ROBERTS, C. J., CIMPRICH, K. A., STRINGER, C. M., LAMB, J. R., SCHREIBER, S. L. & FRIEND, S. H. 1998. Overexpression of a kinase-inactive ATR protein causes sensitivity to DNA-damaging agents and defects in cell cycle checkpoints. *The EMBO Journal,* 17**,** 159--169.

CLINE, M. S., LIAO, R. G., PARSONS, M. T., PATEN, B., ALQUADDOOMI, F., ANTONIOU, A., BAXTER, S., BRODY, L., COOK-DEEGAN, R., COFFIN, A., COUCH, F. J., CRAFT, B., CURRIE, R., DLOTT, C. C., DOLMAN, L., DUNNEN, J. T. D., DYKE, S. O. M., DOMCHEK, S. M., EASTON, D., FISCHMANN, Z., FOULKES, W. D., GARBER, J., GOLDGAR, D., GOLDMAN, M. J., GOODHAND, P., HARRISON, S., HAUSSLER, D., KATO, K., KNOPPERS, B., MARKELLO, C., NUSSBAUM, R., OFFIT, K., PLON, S. E., RASHBASS, J., REHM, H. L., ROBSON, M., RUBINSTEIN, W. S., STOPPA-LYONNET, D., TAVTIGIAN, S., THOROGOOD, A., ZHANG, C., ZIMMERMANN, M., AUTHORS, B. C., BURN, J., CHANOCK, S., RÄTSCH, G. & SPURDLE, A. B. 2018. BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. *PLoS Genetics,* 14**,** e1007752.

COHN, M. A., KOWAL, P., YANG, K., HAAS, W., HUANG, T. T., GYGI, S. P. & D'ANDREA, A. D. 2007. A UAF1-Containing Multisubunit Protein Complex Regulates the Fanconi Anemia Pathway. *Molecular Cell,* 28**,** 786--797.

COIN, F., OKSENYCH, V. & EGLY, J.-M. 2007. Distinct Roles for the XPB/p52 and XPD/p44 Subcomplexes of TFIIH in Damaged DNA Opening during Nucleotide Excision Repair. *Molecular Cell,* 26**,** 245--256.

COLLIS, S. J., BARBER, L. J., WARD, J. D., MARTIN, J. S. & BOULTON, S. J. 2006. C. elegans FANCD2 responds to replication stress and functions in interstrand cross-link repair. *DNA Repair,* 5**,** 1398--1406.

COMPE, E. & EGLY, J.-M. 2012. TFIIH: when transcription met DNA repair. *Nature Reviews Molecular Cell Biology,* 13**,** 343--354.

CONG, K., PENG, M., KOUSHOLT, A. N., LEE, W. T. C., LEE, S., NAYAK, S., KRAIS, J., VANDERVERE-CAROZZA, P. S., PAWELCZAK, K. S., CALVO, J., PANZARINO, N. J., TURCHI, J. J., JOHNSON, N., JONKERS, J., ROTHENBERG, E. & CANTOR, S. B. 2021. Replication gaps are a key determinant of PARP inhibitor synthetic lethality with BRCA deficiency. *Molecular Cell,* 81**,** 3128--3144.e7.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science,* 339**,** 819--823.

CONSORTIUM, B. C. A. 2021. Breast Cancer Risk Genes — Association Analysis in More than 113,000 Women. *New England Journal of Medicine,* 384**,** 428--439.

CONSORTIUM, E. A., LEK, M., KARCZEWSKI, K. J., MINIKEL, E. V., SAMOCHA, K. E., BANKS, E., FENNELL, T., O'DONNELL-LURIA, A. H., WARE, J. S., HILL, A. J., CUMMINGS, B. B., TUKIAINEN, T., BIRNBAUM, D. P., KOSMICKI, J. A., DUNCAN, L. E., ESTRADA, K., ZHAO, F., ZOU, J., PIERCE-HOFFMAN, E., BERGHOUT, J., COOPER, D. N., DEFLAUX, N., DEPRISTO, M., DO, R., FLANNICK, J., FROMER, M., GAUTHIER, L., GOLDSTEIN, J., GUPTA, N., HOWRIGAN, D., KIEZUN, A., KURKI, M. I., MOONSHINE, A. L., NATARAJAN, P., OROZCO, L., PELOSO, G. M., POPLIN, R., RIVAS, M. A., RUANO-RUBIO, V., ROSE, S. A., RUDERFER, D. M., SHAKIR, K., STENSON, P. D., STEVENS, C., THOMAS, B. P., TIAO, G., TUSIE-LUNA, M. T., WEISBURD, B., WON, H.-H., YU, D., ALTSHULER, D. M., ARDISSINO, D., BOEHNKE, M., DANESH, J., DONNELLY, S., ELOSUA, R., FLOREZ, J. C., GABRIEL, S. B., GETZ, G., GLATT, S. J., HULTMAN, C. M., KATHIRESAN, S., LAAKSO, M., MCCARROLL, S., MCCARTHY, M. I., MCGOVERN, D., MCPHERSON, R., NEALE, B. M., PALOTIE, A., PURCELL, S. M., SALEHEEN, D., SCHARF, J. M., SKLAR, P., SULLIVAN, P. F., TUOMILEHTO, J., TSUANG, M. T., WATKINS, H. C., WILSON, J. G., DALY, M. J. & MACARTHUR, D. G. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature,* 536**,** 285--291.

CONSORTIUM, I. H. G. S. 2004. Finishing the euchromatic sequence of the human genome. *Nature,* 431**,** 931--945.

CONSORTIUM*, T. C. E. S. 1998. Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science,* 282**,** 2012--2018.

COOPER, D. L., LAHUE, R. S. & MODRICH, P. 1993. Methyl-directed mismatch repair is bidirectional. *Journal of Biological Chemistry,* 268**,** 11823--11829.

COSTANZO, V., SHECHTER, D., LUPARDUS, P. J., CIMPRICH, K. A., GOTTESMAN, M. & GAUTIER, J. 2003. An ATR- and Cdc7-Dependent DNA Damage Checkpoint that Inhibits Initiation of DNA Replication. *Molecular Cell,* 11**,** 203--213.

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. & CORTEZ, D. 2013. ATR phosphorylates SMARCAL1 to prevent replication fork collapse. *Genes \& Development,* 27**,** 1610--1623.

COYNE, R. S., MCDONALD, H. B., EDGEMON, K. & BRODY, L. C. 2004. Functional Characterization of BRCA1 Sequence Variants using a Yeast Small Colony Phenotype Assay. *Cancer Biology \& Therapy,* 3**,** 453--457.

CPTAC, N. C. I., MERTINS, P., MANI, D. R., RUGGLES, K. V., GILLETTE, M. A., CLAUSER, K. R., WANG, P., WANG, X., QIAO, J. W., CAO, S., PETRALIA, F., KAWALER, E., MUNDT, F., KRUG, K., TU, Z., LEI, J. T., GATZA, M. L., WILKERSON, M., PEROU, C. M., YELLAPANTULA, V., HUANG, K.-L., LIN, C., MCLELLAN, M. D., YAN, P., DAVIES, S. R., TOWNSEND, R. R., SKATES, S. J., WANG, J., ZHANG, B., KINSINGER, C. R., MESRI, M., RODRIGUEZ, H., DING, L., PAULOVICH, A. G., FENYÖ, D., ELLIS, M. J. & CARR, S. A. 2016. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature,* 534**,** 55--62.

CREMONA, C. A. & BEHRENS, A. 2014. ATM signalling and cancer. *Oncogene,* 33**,** 3351--3360.

CRITCHLOW, S. E., BOWATER, R. P. & JACKSON, S. P. 1997. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Current Biology,* 7**,** 588--598.

CUADRADO, M., MARTINEZ-PASTOR, B., MURGA, M., TOLEDO, L. I., GUTIERREZ-MARTINEZ, P., LOPEZ, E. & FERNANDEZ-CAPETILLO, O. 2006. ATM regulates ATR chromatin loading in response to DNA double-strand breaks. *The Journal of Experimental Medicine,* 203**,** 297--303.

CUELLA-MARTIN, R., HAYWARD, S. B., FAN, X., CHEN, X., HUANG, J.-W., TAGLIALATELA, A., LEUZZI, G., ZHAO, J., RABADAN, R., LU, C., SHEN, Y. & CICCIA, A. 2021. Functional interrogation of DNA damage response variants with base editing screens. *Cell,* 184**,** 1081--1097.e19.

CUNNIFF, C., BASSETTI, J. A. & ELLIS, N. A. 2017. Bloom's Syndrome: Clinical Spectrum, Molecular Pathogenesis, and Cancer Predisposition. *Molecular Syndromology,* 8**,** 4--23.

CURIA, M. C., PALMIROTTA, R., ACETO, G., MESSERINI, L., VERÌ, M. C., CROGNALE, S., VALANZANO, R., FICARI, F., FRACASSO, P., STIGLIANO, V., TONELLI, F., CASALE, V., GUADAGNI, F., BATTISTA, P., MARIANI-COSTANTINI, R. & CAMA, A. 1999. Unbalanced germ-line expression of hMLH1 and hMSH2 alleles in hereditary nonpolyposis colorectal cancer. *Cancer research,* 59**,** 3570--5.

CUTALO, J. M., DARDEN, T. A., KUNKEL, T. A. & TOMER, K. B. 2006. Mapping the Dimer Interface in the C-Terminal Domains of the Yeast MLH1−PMS1 Heterodimer †. *Biochemistry,* 45**,** 15458--15467.

D'ANGIOLELLA, V., DONATO, V., FORRESTER, FRANCES M., JEONG, Y.-T., PELLACANI, C., KUDO, Y., SARAF, A., FLORENS, L., WASHBURN, MICHAEL P. & PAGANO, M. 2012. Cyclin F-Mediated Degradation of Ribonucleotide Reductase M2 Controls Genome Integrity and DNA Repair. *Cell,* 149**,** 1023--1034.

DAEE, D. L. & MYUNG, K. 2012. Fanconi-like crosslink repair in yeast. *Genome integrity,* 3**,** 7.

DAGUENET, E., DUJARDIN, G. & VALCÁRCEL, J. 2015. The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. *EMBO reports,* 16**,** 1640--1655.

DAI, J., SANCHEZ, A., ADAM, C., RANJHA, L., REGINATO, G., CHERVY, P., TELLIER-LEBEGUE, C., ANDREANI, J., GUÉROIS, R., ROPARS, V., DU, M.-H. L., MALOISEL, L., MARTINI, E., LEGRAND, P., THUREAU, A., CEJKA, P., BORDE, V. & CHARBONNIER, J.-B. 2021. Molecular basis of the dual role of the Mlh1-Mlh3 endonuclease in MMR and in meiotic crossover formation. *Proceedings of the National Academy of Sciences,* 118**,** e2022704118.

DAN, C., PEI, H., ZHANG, B., ZHENG, X., RAN, D. & DU, C. 2021. Fanconi anemia pathway and its relationship with cancer. *Genome Instability \& Disease,* 2**,** 175--183.

DAO, K.-H. T., ROTELLI, M. D., BROWN, B. R., YATES, J. E., RANTALA, J., TOGNON, C., TYNER, J. W., DRUKER, B. J. & BAGBY, G. C. 2013. The PI3K/Akt1 pathway enhances steady-state levels of FANCL. *Molecular Biology of the Cell,* 24**,** 2582--2592.

DAO, V. & MODRICH, P. 1998. Mismatch-, MutS-, MutL-, and Helicase II-dependent Unwinding from the Single-strand Break of an Incised Heteroduplex*. *Journal of Biological Chemistry,* 273**,** 9202--9207.

DAVIS, A. J., CHEN, B. P. C. & CHEN, D. J. 2014. DNA-PK: A dynamic enzyme in a versatile DSB repair pathway. *DNA Repair,* 17**,** 21--29.

DAVIS, A. J. & CHEN, D. J. 2013. DNA double strand break repair via non-homologous end-joining. *Translational cancer research,* 2**,** 130--143.

DEANS, A. J. & WEST, S. C. 2011. DNA interstrand crosslink repair and cancer. *Nature Reviews Cancer,* 11**,** 467--480.

DEMOGINES, A., EAST, A. M., LEE, J.-H., GROSSMAN, S. R., SABETI, P. C., PAULL, T. T. & SAWYER, S. L. 2010. Ancient and Recent Adaptive Evolution of Primate Non-Homologous End Joining Genes. *PLoS Genetics,* 6**,** e1001169.

DESHPANDE, RAJASHREE A., LEE, J.-H., ARORA, S. & PAULL, TANYA T. 2016. Nbs1 Converts the Human Mre11/Rad50 Nuclease Complex into an Endo/Exonuclease Machine Specific for Protein-DNA Adducts. *Molecular Cell,* 64**,** 593--606.

DESHPANDE, R. A. & WILSON, T. E. 2007. Modes of interaction among yeast Nej1, Lif1 and Dnl4 proteins and comparison to human XLF, XRCC4 and Lig4. *DNA Repair,* 6**,** 1507--1516.

DESOUKY, O., DING, N. & ZHOU, G. 2015. Targeted and non-targeted effects of ionizing radiation. *Journal of Radiation Research and Applied Sciences,* 8**,** 247--254.

DEV, H., CHIANG, T.-W. W., LESCALE, C., KRIJGER, I. D., MARTIN, A. G., PILGER, D., COATES, J., SCZANIECKA-CLIFT, M., WEI, W., OSTERMAIER, M., HERZOG, M., LAM, J., SHEA, A., DEMIR, M., WU, Q., YANG, F., FU, B., LAI, Z., BALMUS, G., BELOTSERKOVSKAYA, R., SERRA, V., O'CONNOR, M. J., BRUNA, A., BELI, P., PELLEGRINI, L., CALDAS, C., DERIANO, L., JACOBS, J. J. L., GALANTY, Y. & JACKSON, S. P. 2018. Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. *Nature Cell Biology,* 20**,** 954--965.

DICKERSON, R. E. 1971. The structure of cytochromec and the rates of molecular evolution. *Journal of Molecular Evolution,* 1**,** 26--45.

DIGIOVANNA, J. J. & KRAEMER, K. H. 2012. Shining a Light on Xeroderma Pigmentosum. *Journal of Investigative Dermatology,* 132**,** 785--796.

DIKIC, I., WAKATSUKI, S. & WALTERS, K. J. 2009. Ubiquitin-binding domains — from structures to functions. *Nature Reviews Molecular Cell Biology,* 10**,** 659--671.

DOBBS, T. A., TAINER, J. A. & LEES-MILLER, S. P. 2010. A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA Repair,* 9**,** 1307--1314.

DONOVAN, F. X., SOLANKI, A., MORI, M., CHAVAN, N., GEORGE, M., C, S. K., OKUNO, Y., MURAMASTSU, H., YOSHIDA, K., SHIMAMOTO, A., TAKAORI-KONDO, A., YABE, H., OGAWA, S., KOJIMA, S., YABE, M., RAMANAGOUDR-BHOJAPPA, R., SMOGORZEWSKA, A., MOHAN, S., RAJENDRAN, A., AUERBACH, A. D., TAKATA, M., CHANDRASEKHARAPPA, S. C. & VUNDINTI, B. R. 2020. A founder variant in the South Asian population leads to a high prevalence of FANCL Fanconi anemia cases in India. *Human Mutation,* 41**,** 122--128.

DROST, M., TIERSMA, Y., GLUBB, D., KATHE, S., HEES, S. V., CALLÉJA, F., ZONNEVELD, J. B. M., BOUCHER, K. M., RAMLAL, R. P. E., THOMPSON, B. A., RASMUSSEN, L. J., GREENBLATT, M. S., LEE, A., SPURDLE, A. B., TAVTIGIAN, S. V. & WIND, N. D. 2020. Two integrated and highly predictive functional analysis-based procedures for the classification of MSH6 variants in Lynch syndrome. *Genetics in Medicine,* 22**,** 847--856.

DROST, M., TIERSMA, Y., THOMPSON, B. A., FREDERIKSEN, J. H., KEIJZERS, G., GLUBB, D., KATHE, S., OSINGA, J., WESTERS, H., PAPPAS, L., BOUCHER, K. M., MOLENKAMP, S., ZONNEVELD, J. B., ASPEREN, C. J. V., GOLDGAR, D. E., WALLACE, S. S., SIJMONS, R. H., SPURDLE, A. B., RASMUSSEN, L. J., GREENBLATT, M. S., WIND, N. D. & TAVTIGIAN, S. V. 2019. A functional assay–based procedure to classify mismatch repair gene variants in Lynch syndrome. *Genetics in Medicine,* 21**,** 1486--1496.

DROST, M., ZONNEVELD, J. É. B. M., DIJK, L. V., MORREAU, H., TOPS, C. M., VASEN, H. F. A., WIJNEN, J. T. & WIND, N. D. 2010. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. *Human Mutation,* 31**,** 247--253.

DROTSCHMANN, K., CLARK, A. B., TRAN, H. T., RESNICK, M. A., GORDENIN, D. A. & KUNKEL, T. A. 1999. Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene. *Proceedings of the National Academy of Sciences,* 96**,** 2970--2975.

DRUMMOND, J. T., LI, G.-M., LONGLEY, M. J. & MODRICH, P. 1995. Isolation of an hMSH2-p160 Heterodimer That Restores DNA Mismatch Repair to Tumor Cells. *Science,* 268**,** 1909--1912.
DUAN, M., SPEER, R. M., ULIBARRI, J., LIU, K. J. & MAO, P. 2021. Transcription-coupled nucleotide excision repair: New insights revealed by genomic approaches. *DNA Repair,* 103**,** 103126.
DUCKETT, D. R., DRUMMOND, J. T., MURCHIE, A. I., REARDON, J. T., SANCAR, A., LILLEY, D. M. & MODRICH, P. 1996. Human MutSalpha recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct. *Proceedings of the National Academy of Sciences,* 93**,** 6443--6447.
DUDLEY, B., BRAND, R. E., THULL, D., BAHARY, N., NIKIFOROVA, M. N. & PAI, R. K. 2015. Germline MLH1 Mutations Are Frequently Identified in Lynch Syndrome Patients With Colorectal and Endometrial Carcinoma Demonstrating Isolated Loss of PMS2 Immunohistochemical Expression. *The American Journal of Surgical Pathology,* 39**,** 1114--1120.
DUPRÉ, A., BOYER-CHATENET, L. & GAUTIER, J. 2006. Two-step activation of ATM by DNA and the Mre11–Rad50–Nbs1 complex. *Nature Structural \& Molecular Biology,* 13**,** 451--457.
DUPRIE, M. L., PALACIO, T., CALIL, F. A., KOLODNER, R. D. & PUTNAM, C. D. 2022. Mlh1 interacts with both Msh2 and Msh6 for recruitment during mismatch repair. *DNA Repair,* 119**,** 103405.
DURATURO, F., LICCARDO, R., ROSA, M. D. & IZZO, P. 2019. Genetics, diagnosis and treatment of Lynch syndrome: Old lessons and current challenges. *Oncology Letters,* 17**,** 3048--3054.
DZANTIEV, L., CONSTANTIN, N., GENSCHEL, J., IYER, R. R., BURGERS, P. M. & MODRICH, P. 2004. A Defined Human System That Supports Bidirectional Mismatch-Provoked Excision. *Molecular Cell,* 15**,** 31--41.
EASTMAN, A. 1987. The formation, isolation and characterization of DNA adducts produced by anticancer platinum complexes. *Pharmacology \& Therapeutics,* 34**,** 155--166.
EASTON, D. F., FORD, D. & BISHOP, D. T. 1995. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *American journal of human genetics,* 56**,** 265--71.
EDELMANN, W., COHEN, P. E., KNEITZ, B., WINAND, N., LIA, M., HEYER, J., KOLODNER, R., POLLARD, J. W. & KUCHERLAPATI, R. 1999. Mammalian MutS homologue 5 is required for chromosome pairing in meiosis. *Nature Genetics,* 21**,** 123--127.
EDWARDS, P. & MONAHAN, K. J. 2022. Diagnosis and management of Lynch syndrome. *Frontline Gastroenterology,* 13**,** e80--e87.
EL-KHAMISY, S. F., MASUTANI, M., SUZUKI, H. & 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.
ELEZ, M., RADMAN, M. & MATIC, I. 2012. Stoichiometry of MutS and MutL at unrepaired mismatches in vivo suggests a mechanism of repair. *Nucleic Acids Research,* 40**,** 3929--3938.
ELLISON, A. R., LOFING, J. & BITTER, G. A. 2001. Functional analysis of human MLH1 and MSH2 missense variants and hybrid human–yeast MLH1 proteins in Saccharomyces cerevisiae. *Human Molecular Genetics,* 10**,** 1889--1900.
ELLISON, A. R., LOFING, J. & BITTER, G. A. 2004. Human MutL homolog (MLH1) function in DNA mismatch repair: a prospective screen for missense mutations in the ATPase domain. *Nucleic Acids Research,* 32**,** 5321--5338.
ELZE, L., MENSENKAMP, A. R., NAGTEGAAL, I. D., ZELST-STAMS, W. A. G. V., GROUP, D. L. S. L. S., DOMMERING, C. J., HOOGERBRUGGE, N., JONG, M. M. D., BLEEKER, F. E., LETER, E. M., LETTEBOER, T. G. W., NIELSEN, M., POST, R. S. V. D., WAPSTRA, B., VOER, R. M. D. & LIGTENBERG, M. J. L. 2021. Somatic Nonepigenetic Mismatch Repair Gene Aberrations Underly Most Mismatch Repair–Deficient Lynch-Like Tumors. *Gastroenterology,* 160**,** 1414--1416.e3.
ERWOOD, S., BILY, T. M. I., LEQUYER, J., YAN, J., GULATI, N., BREWER, R. A., ZHOU, L., PELLETIER, L., IVAKINE, E. A. & COHN, R. D. 2022. Saturation variant interpretation using CRISPR prime editing. *Nature Biotechnology***,** 1--11.

ESCRIBANO-DÍAZ, C., ORTHWEIN, A., FRADET-TURCOTTE, A., XING, M., YOUNG, JORDAN T. F., TKÁČ, J., COOK, MICHAEL A., ROSEBROCK, ADAM P., MUNRO, M., CANNY, MARELLA D., XU, D. & DUROCHER, D. 2013. A Cell Cycle-Dependent Regulatory Circuit Composed of 53BP1-RIF1 and BRCA1-CtIP Controls DNA Repair Pathway Choice. *Molecular Cell,* 49**,** 872--883.

EWALD, B., SAMPATH, D. & PLUNKETT, W. 2008. Nucleoside analogs: molecular mechanisms signaling cell death. *Oncogene,* 27**,** 6522--6537.

FALCK, J., COATES, J. & JACKSON, S. P. 2005. Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature,* 434**,** 605--611.

FAN, L., FUSS, J. O., CHENG, Q. J., ARVAI, A. S., HAMMEL, M., ROBERTS, V. A., COOPER, P. K. & TAINER, J. A. 2008. XPD Helicase Structures and Activities: Insights into the Cancer and Aging Phenotypes from XPD Mutations. *Cell,* 133**,** 789--800.

FAN, W. & LUO, J. 2010. SIRT1 Regulates UV-Induced DNA Repair through Deacetylating XPA. *Molecular Cell,* 39**,** 247--258.

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., MARTIN, N. M. B., JACKSON, S. P., SMITH, G. C. M. & ASHWORTH, A. 2005. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature,* 434**,** 917--921.

FARRUGIA, D. J., AGARWAL, M. K., PANKRATZ, V. S., DEFFENBAUGH, A. M., PRUSS, D., FRYE, C., WADUM, L., JOHNSON, K., MENTLICK, J., TAVTIGIAN, S. V., GOLDGAR, D. E. & COUCH, F. J. 2008. Functional Assays for Classification of BRCA2 Variants of Uncertain Significance. *Cancer Research,* 68**,** 3523--3531.

FARWELL, K. D., SHAHMIRZADI, L., EL-KHECHEN, D., POWIS, Z., CHAO, E. C., DAVIS, B. T., BAXTER, R. M., ZENG, W., MROSKE, C., PARRA, M. C., GANDOMI, S. K., LU, I., LI, X., LU, H., LU, H.-M., SALVADOR, D., RUBLE, D., LAO, M., FISCHBACH, S., WEN, J., LEE, S., ELLIOTT, A., DUNLOP, C. L. M. & TANG, S. 2015. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model–based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genetics in Medicine,* 17**,** 578--586.

FEKAIRI, S., SCAGLIONE, S., CHAHWAN, C., TAYLOR, E. R., TISSIER, A., COULON, S., DONG, M.-Q., RUSE, C., YATES, J. R., RUSSELL, P., FUCHS, R. P., MCGOWAN, C. H. & GAILLARD, P.-H. L. 2009. Human SLX4 Is a Holliday Junction Resolvase Subunit that Binds Multiple DNA Repair/Recombination Endonucleases. *Cell,* 138**,** 78--89.

FENG, S., ZHAO, Y., XU, Y., NING, S., HUO, W., HOU, M., GAO, G., JI, J., GUO, R. & XU, D. 2016. Ewing Tumor-associated Antigen 1 Interacts with Replication Protein A to Promote Restart of Stalled Replication Forks* * This work was supported in part by the National Basic Research Program of China Grant 2013CB911002 and the National Natural Science Foundation of China Grants 31271435 and 31370836. The authors declare that they have no conflicts of interest with the contents of this article. *Journal of Biological Chemistry,* 291**,** 21956--21962.

FERNANDEZ-LEIRO, R., BHAIROSING-KOK, D., KUNETSKY, V., LAFFEBER, C., WINTERWERP, H. H., GROOTHUIZEN, F., FISH, A., LEBBINK, J. H. G., FRIEDHOFF, P., SIXMA, T. K. & LAMERS, M. H. 2021. The selection process of licensing a DNA mismatch for repair. *Nature Structural \& Molecular Biology,* 28**,** 373--381.

FERRAO, P. T., BUKCZYNSKA, E. P., JOHNSTONE, R. W. & MCARTHUR, G. A. 2012. Efficacy of CHK inhibitors as single agents in MYC-driven lymphoma cells. *Oncogene,* 31**,** 1661--1672.

FINDLAY, G. M., DAZA, R. M., MARTIN, B., ZHANG, M. D., LEITH, A. P., GASPERINI, M., JANIZEK, J. D., HUANG, X., STARITA, L. M. & SHENDURE, J. 2018a. Accurate classification of BRCA1 variants with saturation genome editing. *Nature,* 562**,** 217--222.

FINDLAY, S., HEATH, J., LUO, V. M., MALINA, A., MORIN, T., COULOMBE, Y., DJERIR, B., LI, Z., SAMIEI, A., SIMO-CHEYOU, E., KARAM, M., BAGCI, H., RAHAT, D., GRAPTON, D., LAVOIE, E. G., DOVE, C., KHALED, H., KUASNE, H., MANN, K. K., KLEIN, K. O., GREENWOOD, C. M., TABACH, Y., PARK, M., CÔTÉ, J. F., MASSON, J. Y., MARÉCHAL, A. & ORTHWEIN, A. 2018b. SHLD2/FAM35A

co-operates with REV7 to coordinate DNA double-strand break repair pathway choice. *The EMBO Journal,* 37**,** e100158.

FISCHER, J. M. F., POPP, O., GEBHARD, D., VEITH, S., FISCHBACH, A., BENEKE, S., LEITENSTORFER, A., BERGEMANN, J., SCHEFFNER, M., FERRANDO-MAY, E., MANGERICH, A. & BÜRKLE, A. 2014. Poly(ADP-ribose)-mediated interplay of XPA and PARP1 leads to reciprocal regulation of protein function. *The FEBS Journal,* 281**,** 3625--3641.

FISHEL, R., LESCOE, M. K., RAO, M. R. S., COPELAND, N. G., JENKINS, N. A., GARBER, J., KANE, M. & KOLODNER, R. 1993. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell,* 75**,** 1027--1038.

FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J.-F., DOUGHERTY, B. A., MERRICK, J. M., MCKENNEY, K., SUTTON, G., FITZHUGH, W., FIELDS, C., GOCAYNE, J. D., SCOTT, J., SHIRLEY, R., LIU, L.-L., GLODEK, A., KELLEY, J. M., WEIDMAN, J. F., PHILLIPS, C. A., SPRIGGS, T., HEDBLOM, E., COTTON, M. D., UTTERBACK, T. R., HANNA, M. C., NGUYEN, D. T., SAUDEK, D. M., BRANDON, R. C., FINE, L. D., FRITCHMAN, J. L., FUHRMANN, J. L., GEOGHAGEN, N. S. M., GNEHM, C. L., MCDONALD, L. A., SMALL, K. V., FRASER, C. M., SMITH, H. O. & VENTER, J. C. 1995. Whole-Genome Random Sequencing and Assembly of Haemophilus influenzae Rd. *Science,* 269**,** 496--512.

FLORES-ROZAS, H., CLARK, D. & KOLODNER, R. D. 2000. Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex. *Nature Genetics,* 26**,** 375--378.

FOK, J. H. L., RAMOS-MONTOYA, A., VAZQUEZ-CHANTADA, M., WIJNHOVEN, P. W. G., FOLLIA, V., JAMES, N., FARRINGTON, P. M., KARMOKAR, A., WILLIS, S. E., CAIRNS, J., NIKKILÄ, J., BEATTIE, D., LAMONT, G. M., FINLAY, M. R. V., WILSON, J., SMITH, A., O'CONNOR, L. O., LING, S., FAWELL, S. E., O'CONNOR, M. J., HOLLINGSWORTH, S. J., DEAN, E., GOLDBERG, F. W., DAVIES, B. R. & CADOGAN, E. B. 2019. AZD7648 is a potent and selective DNA-PK inhibitor that enhances radiation, chemotherapy and olaparib activity. *Nature Communications,* 10**,** 5065.

FORBES, S. A., BEARE, D., BOUTSELAKIS, H., BAMFORD, S., BINDAL, N., TATE, J., COLE, C. G., WARD, S., DAWSON, E., PONTING, L., STEFANCSIK, R., HARSHA, B., KOK, C. Y., JIA, M., JUBB, H., SONDKA, Z., THOMPSON, S., DE, T. & CAMPBELL, P. J. 2017. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Research,* 45**,** D777--D783.

FORBES, S. A., BINDAL, N., BAMFORD, S., COLE, C., KOK, C. Y., BEARE, D., JIA, M., SHEPHERD, R., LEUNG, K., MENZIES, A., TEAGUE, J. W., CAMPBELL, P. J., STRATTON, M. R. & FUTREAL, P. A. 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Research,* 39**,** D945--D950.

FORMENT, J. V., BLASIUS, M., GUERINI, I. & JACKSON, S. P. 2011. Structure-Specific DNA Endonuclease Mus81/Eme1 Generates DNA Damage Caused by Chk1 Inactivation. *PLoS ONE,* 6**,** e23517.

FOWLER, D. M. & FIELDS, S. 2014. Deep mutational scanning: a new style of protein science. *Nature Methods,* 11**,** nmeth.3027.

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. & DUROCHER, D. 2013. 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature,* 499**,** 50--54.

FRANCIS, D. B., KOZLOV, M., CHAVEZ, J., CHU, J., MALU, S., HANNA, M. & CORTES, P. 2014. DNA Ligase IV regulates XRCC4 nuclear localization. *DNA Repair,* 21**,** 36--42.

FRAZER, J., NOTIN, P., DIAS, M., GOMEZ, A., MIN, J. K., BROCK, K., GAL, Y. & MARKS, D. S. 2021. Disease variant prediction with deep generative models of evolutionary data. *Nature,* 599**,** 91--95.

FRIT, P., ROPARS, V., MODESTI, M., CHARBONNIER, J.-B. & CALSOU, P. 2019. Plugged into the Ku-DNA hub: The NHEJ network. *Progress in biophysics and molecular biology,* 147**,** 62--76.

FROGGATT, N. J., BRASSETT, C., KOCH, D. J., EVANS, D. G., HODGSON, S. V., PONDER, B. A. & MAHER, E. R. 1996. Mutation screening of MSH2 and MLH1 mRNA in hereditary non-polyposis colon cancer syndrome. *Journal of Medical Genetics,* 33**,** 726.

FROST, M. G., ABOUKHEILI, A. M. M., TOTH, R. & WALDEN, H. 2020. Characterization of FANCL variants observed in patient cancer cells. *Bioscience Reports,* 40**,** BSR20191304.

FU, W., O'CONNOR, T. D., JUN, G., KANG, H. M., ABECASIS, G., LEAL, S. M., GABRIEL, S., RIEDER, M. J., ALTSHULER, D., SHENDURE, J., NICKERSON, D. A., BAMSHAD, M. J., PROJECT, N. E. S. & AKEY, J. M. 2013. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature,* 493**,** 216--220.

FUKUCHI, M., WANOTAYAN, R., LIU, S., IMAMICHI, S., SHARMA, M. K. & MATSUMOTO, Y. 2015. Lysine 271 but not lysine 210 of XRCC4 is required for the nuclear localization of XRCC4 and DNA ligase IV. *Biochemical and Biophysical Research Communications,* 461**,** 687--694.

GABRIEL, H., KORINTH, D., RITTHALER, M., SCHULTE, B., BATTKE, F., KAISENBERG, C., WÜSTEMANN, M., SCHULZE, B., FRIEDRICH-FREKSA, A., PFEIFFER, L., ENTEZAMI, M., SCHRÖER, A., BÜRGER, J., SCHWAIBOLD, E. M. C., LEBEK, H. & BISKUP, S. 2022. Trio exome sequencing is highly relevant in prenatal diagnostics. *Prenatal Diagnosis,* 42**,** 845--851.

GALANDE, A. A., SAIJO, M., GHASKADBI, S. S. & GHASKADBI, S. 2021. Xeroderma pigmentosum A homolog from Hydra partially complements DNA repair defect in human XPA-deficient cells. *Journal of Biosciences,* 46**,** 48.

GALLO, D., YOUNG, J. T. F., FOURTOUNIS, J., MARTINO, G., ÁLVAREZ-QUILÓN, A., BERNIER, C., DUFFY, N. M., PAPP, R., ROULSTON, A., STOCCO, R., SZYCHOWSKI, J., VELOSO, A., ALAM, H., BARUAH, P. S., FORTIN, A. B., BOWLAN, J., CHAUDHARY, N., DESJARDINS, J., DIETRICH, E., FOURNIER, S., FUGÈRE-DESJARDINS, C., RUGY, T. G. D., LECLAIRE, M.-E., LIU, B., BHASKARAN, V., MAMANE, Y., MELO, H., NICOLAS, O., SINGHANIA, A., SZILARD, R. K., TKÁČ, J., YIN, S. Y., MORRIS, S. J., ZINDA, M., MARSHALL, C. G. & DUROCHER, D. 2022. CCNE1 amplification is synthetic lethal with PKMYT1 kinase inhibition. *Nature,* 604**,** 749--756.

GAMMIE, A. E., ERDENIZ, N., BEAVER, J., DEVLIN, B., NANJI, A. & ROSE, M. D. 2007. Functional Characterization of Pathogenic Human MSH2 Missense Mutations in Saccharomyces cerevisiae. *Genetics,* 177**,** 707--721.

GAO, J., AKSOY, B. A., DOGRUSOZ, U., DRESDNER, G., GROSS, B., SUMER, S. O., SUN, Y., JACOBSEN, A., SINHA, R., LARSSON, E., CERAMI, E., SANDER, C. & SCHULTZ, N. 2013. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling,* 6**,** pl1.

GAO, S., FENG, S., NING, S., LIU, J., ZHAO, H., XU, Y., SHANG, J., LI, K., LI, Q., GUO, R. & XU, D. 2018. An OB-fold complex controls the repair pathways for DNA double-strand breaks. *Nature Communications,* 9**,** 3925.

GAO, Y., SUN, Y., FRANK, K. M., DIKKES, P., FUJIWARA, Y., SEIDL, K. J., SEKIGUCHI, J. M., RATHBUN, G. A., SWAT, W., WANG, J., BRONSON, R. T., MALYNN, B. A., BRYANS, M., ZHU, C., CHAUDHURI, J., DAVIDSON, L., FERRINI, R., STAMATO, T., ORKIN, S. H., GREENBERG, M. E. & ALT, F. W. 1998. A Critical Role for DNA End-Joining Proteins in Both Lymphogenesis and Neurogenesis. *Cell,* 95**,** 891--902.

GARAYCOECHEA, J. I., CROSSAN, G. P., LANGEVIN, F., MULDERRIG, L., LOUZADA, S., YANG, F., GUILBAUD, G., PARK, N., ROERINK, S., NIK-ZAINAL, S., STRATTON, M. R. & PATEL, K. J. 2018. Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. *Nature,* 553**,** 171--177.

GARCIA-HIGUERA, I., TANIGUCHI, T., GANESAN, S., MEYN, M. S., TIMMERS, C., HEJNA, J., GROMPE, M. & D'ANDREA, A. D. 2001. Interaction of the Fanconi Anemia Proteins and BRCA1 in a Common Pathway. *Molecular Cell,* 7**,** 249--262.

GARCIA, V., PHELPS, S. E. L., GRAY, S. & NEALE, M. J. 2011. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature,* 479**,** 241--244.

GAUDELLI, N. M., KOMOR, A. C., REES, H. A., PACKER, M. S., BADRAN, A. H., BRYSON, D. I. & LIU, D. R. 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature,* 551**,** 464--471.

GENSCHEL, J., BAZEMORE, L. R. & MODRICH, P. 2002. Human Exonuclease I Is Required for 5′ and 3′ Mismatch Repair∗. *Journal of Biological Chemistry,* 277**,** 13302--13311.

GENSCHEL, J., LITTMAN, S. J., DRUMMOND, J. T. & MODRICH, P. 1998. Isolation of MutSβ from Human Cells and Comparison of the Mismatch Repair Specificities of MutSβ and MutSα*. *Journal of Biological Chemistry,* 273**,** 19895--19901.

GENSCHEL, J. & MODRICH, P. 2003. Mechanism of 5′-Directed Excision in Human Mismatch Repair. *Molecular Cell,* 12**,** 1077--1086.

GEORGIOU, D., MONJE-GARCIA, L., MILES, T., MONAHAN, K. & RYAN, N. A. J. 2023. A Focused Clinical Review of Lynch Syndrome. *Cancer Management and Research,* 15**,** 67--85.

GERMAN, J. 1993. Bloom Syndrome. *Medicine,* 72**,** 393--406.

GERMANO, G., LAMBA, S., ROSPO, G., BARAULT, L., MAGRÌ, A., MAIONE, F., RUSSO, M., CRISAFULLI, G., BARTOLINI, A., LERDA, G., SIRAVEGNA, G., MUSSOLIN, B., FRAPOLLI, R., MONTONE, M., MORANO, F., BRAUD, F. D., AMIROUCHENE-ANGELOZZI, N., MARSONI, S., D'INCALCI, M., ORLANDI, A., GIRAUDO, E., SARTORE-BIANCHI, A., SIENA, S., PIETRANTONIO, F., NICOLANTONIO, F. D. & BARDELLI, A. 2017. Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth. *Nature,* 552**,** 116--120.

GHEZRAOUI, H., OLIVEIRA, C., BECKER, J. R., BILHAM, K., MORALLI, D., ANZILOTTI, C., FISCHER, R., DEOBAGKAR-LELE, M., SANCHIZ-CALVO, M., FUEYO-MARCOS, E., BONHAM, S., KESSLER, B. M., ROTTENBERG, S., CORNALL, R. J., GREEN, C. M. & CHAPMAN, J. R. 2018. 53BP1 cooperation with the REV7–shieldin complex underpins DNA structure-specific NHEJ. *Nature,* 560**,** 122--127.

GHOSH, R., HARRISON, S. M., REHM, H. L., PLON, S. E., BIESECKER, L. G. & GROUP, O. B. O. C. S. V. I. W. 2018. Updated recommendation for the benign stand-alone ACMG/AMP criterion. *Human Mutation,* 39**,** 1525--1530.

GHOSH, R., OAK, N. & PLON, S. E. 2017. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biology,* 18**,** 225.

GIACCIA, A., WEINSTEIN, R., HU, J. & STAMATO, T. D. 1985. Cell cycle-dependent repair of double-strand DNA breaks in a γ-ray-sensitive Chinese hamster cell. *Somatic Cell and Molecular Genetics,* 11**,** 485--491.

GIACCIA, A. J., DENKO, N., MACLAREN, R., MIRMAN, D., WALDREN, C., HART, I. & STAMATO, T. D. 1990. Human chromosome 5 complements the DNA double-strand break-repair deficiency and gamma-ray sensitivity of the XR-1 hamster variant. *American journal of human genetics,* 47**,** 459--69.

GIANFALDONI, S., GIANFALDONI, R., WOLLINA, U., LOTTI, J., TCHERNEV, G. & LOTTI, T. 2017. An Overview on Radiotherapy: From Its History to Its Current Applications in Dermatology. *Open Access Macedonian Journal of Medical Sciences,* 0**,** 521--525.

GIARDIELLO, F. M., ALLEN, J. I., AXILBUND, J. E., BOLAND, C. R., BURKE, C. A., BURT, R. W., CHURCH, J. M., DOMINITZ, J. A., JOHNSON, D. A., KALTENBACH, T., LEVIN, T. R., LIEBERMAN, D. A., ROBERTSON, D. J., SYNGAL, S. & REX, D. K. 2014. Guidelines on Genetic Evaluation and Management of Lynch Syndrome. *Diseases of the Colon \& Rectum,* 57**,** 1025--1048.

GIARDINA, T., ROBINSON, C., GRIEU-IACOPETTA, F., MILLWARD, M., IACOPETTA, B., SPAGNOLO, D. & AMANUEL, B. 2018. Implementation of next generation sequencing technology for somatic mutation detection in routine laboratory practice. *Pathology,* 50**,** 389--401.

GILBERT, W. & MAXAM, A. 1973. The Nucleotide Sequence of the lac Operator. *Proceedings of the National Academy of Sciences,* 70**,** 3581--3584.

GILISSEN, C., HOISCHEN, A., BRUNNER, H. G. & VELTMAN, J. A. 2011. Unlocking Mendelian disease using exome sequencing. *Genome Biology,* 12**,** 228.

GILLJAM, K. M., MÜLLER, R., LIABAKK, N. B. & OTTERLEI, M. 2012. Nucleotide Excision Repair Is Associated with the Replisome and Its Efficiency Depends on a Direct Interaction between XPA and PCNA. *PLoS ONE,* 7**,** e49199.

GLAZER, P. M., SARKAR, S. N., CHISHOLM, G. E. & SUMMERS, W. C. 1987. DNA mismatch repair detected in human cell extracts. *Molecular and Cellular Biology,* 7**,** 218--224.

GLICKMAN, B. W. & RADMAN, M. 1980. Escherichia coli mutator mutants deficient in methylation-instructed DNA mismatch correction. *Proceedings of the National Academy of Sciences,* 77**,** 1063--1067.

GOELLNER, E. M., PUTNAM, C. D. & KOLODNER, R. D. 2015. Exonuclease 1-dependent and independent mismatch repair. *DNA Repair,* 32**,** 24--32.

GOODWIN, S., MCPHERSON, J. D. & MCCOMBIE, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics,* 17**,** 333--351.

GORMAN, J., CHOWDHURY, A., SURTEES, J. A., SHIMADA, J., REICHMAN, D. R., ALANI, E. & GREENE, E. C. 2007. Dynamic Basis for One-Dimensional DNA Scanning by the Mismatch Repair Complex Msh2-Msh6. *Molecular Cell,* 28**,** 359--370.

GOTTIFREDI, V. & WIESMÜLLER, L. 2020. Current Understanding of RAD52 Functions: Fundamental and Therapeutic Insights. *Cancers,* 12**,** 705.

GOTTLIEB, T. M. & JACKSON, S. P. 1993. The DNA-dependent protein kinase: Requirement for DNA ends and association with Ku antigen. *Cell,* 72**,** 131--142.

GRADIA, S., ACHARYA, S. & FISHEL, R. 1997. The Human Mismatch Recognition Complex hMSH2-hMSH6 Functions as a Novel Molecular Switch. *Cell,* 91**,** 995--1005.

GRADIA, S., SUBRAMANIAN, D., WILSON, T., ACHARYA, S., MAKHOV, A., GRIFFITH, J. & FISHEL, R. 1999. hMSH2–hMSH6 Forms a Hydrolysis-Independent Sliding Clamp on Mismatched DNA. *Molecular Cell,* 3**,** 255--261.

GRAHAM, T. G. W., CARNEY, S. M., WALTER, J. C. & LOPARO, J. J. 2018a. A single XLF dimer bridges DNA ends during nonhomologous end joining. *Nature Structural \& Molecular Biology,* 25**,** 877--884.

GRAHAM, THOMAS G. W., WALTER, JOHANNES C. & LOPARO, JOSEPH J. 2016. Two-Stage Synapsis of DNA Ends during Non-homologous End Joining. *Molecular Cell,* 61**,** 850--858.

GRAHAM, W. J., PUTNAM, C. D. & KOLODNER, R. D. 2018b. The properties of Msh2–Msh6 ATP binding mutants suggest a signal amplification mechanism in DNA mismatch repair. *Journal of Biological Chemistry,* 293**,** 18055--18070.

GRAVEL, S., CHAPMAN, J. R., MAGILL, C. & 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. & LIEBER, M. R. 1997. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature,* 388**,** 492--495.

GRAWUNDER, U., ZIMMER, D. & LIEBER, M. R. 1998. DNA ligase IV binds to XRCC4 via a motif located between rather than within its BRCT domains. *Current Biology,* 8**,** 873--879.

GRAY, V. E., HAUSE, R. J. & FOWLER, D. M. 2017. Analysis of Large-Scale Mutagenesis Data To Assess the Impact of Single Amino Acid Substitutions. *Genetics,* 207**,** 53--61.

GRILLEY, M., WELSH, K. M., SU, S. S. & MODRICH, P. 1989. Isolation and Characterization of the Escherichia coli mutL Gene Product. *Journal of Biological Chemistry,* 264**,** 1000--1004.

GROELLY, F. J., FAWKES, M., DAGG, R. A., BLACKFORD, A. N. & TARSOUNAS, M. 2022. Targeting DNA damage response pathways in cancer. *Nature Reviews Cancer***,** 1--17.

GROOTHUIZEN, F. S., WINKLER, I., CRISTÓVÃO, M., FISH, A., WINTERWERP, H. H. K., REUMER, A., MARX, A. D., HERMANS, N., NICHOLLS, R. A., MURSHUDOV, G. N., LEBBINK, J. H. G., FRIEDHOFF, P. & SIXMA, T. K. 2015. MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA. *eLife,* 4**,** e06744.

GU, J., LU, H., TIPPIN, B., SHIMAZAKI, N., GOODMAN, M. F. & LIEBER, M. R. 2007. XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *The EMBO Journal,* 26**,** 1010--1023.

GU, Y., ROSENBLATT, J. & MORGAN, D. O. 1992. Cell cycle regulation of CDK2 activity by phosphorylation of Thr160 and Tyr15. *The EMBO Journal,* 11**,** 3995--4005.

GUARNÉ, A., RAMON-MAIQUES, S., WOLFF, E. M., GHIRLANDO, R., HU, X., MILLER, J. H. & YANG, W. 2004. Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *The EMBO Journal,* 23**,** 4134--4145.

GUDMUNDSSON, S., SINGER-BERK, M., WATTS, N. A., PHU, W., GOODRICH, J. K., SOLOMONSON, M., CONSORTIUM, G. A. D., REHM, H. L., MACARTHUR, D. G. & O'DONNELL-LURIA, A. 2022. Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation,* 43**,** 1012--1030.

GUEIDERIKH, A., MACZKOWIAK-CHARTOIS, F. & ROSSELLI, F. 2022. A new frontier in Fanconi anemia: From DNA repair to ribosome biogenesis. *Blood Reviews,* 52**,** 100904.

GUENEAU, E., DHERIN, C., LEGRAND, P., TELLIER-LEBEGUE, C., GILQUIN, B., BONNESOEUR, P., LONDINO, F., QUEMENER, C., DU, M.-H. L., MÁRQUEZ, J. A., MOUTIEZ, M., GONDRY, M., BOITEUX, S. & CHARBONNIER, J.-B. 2013. Structure of the MutLα C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. *Nature Structural \& Molecular Biology,* 20**,** 461--468.

GUERRETTE, S., ACHARYA, S. & FISHEL, R. 1999. The Interaction of the Human MutL Homologues in Hereditary Nonpolyposis Colon Cancer*. *Journal of Biological Chemistry,* 274**,** 6336--6341.

GUERTIN, A. D., MARTIN, M. M., ROBERTS, B., HURD, M., QU, X., MISELIS, N. R., LIU, Y., LI, J., FELDMAN, I., BENITA, Y., BLOECHER, A., TONIATTI, C. & SHUMWAY, S. D. 2012. Unique functions of CHK1 and WEE1 underlie synergistic anti-tumor activity upon pharmacologic inhibition. *Cancer Cell International,* 12**,** 45.

GUO, C., FISCHHABER, P. L., LUK-PASZYC, M. J., MASUDA, Y., ZHOU, J., KAMIYA, K., KISKER, C. & FRIEDBERG, E. C. 2003. Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. *The EMBO Journal,* 22**,** 6621--6630.

GUO, C., NAKAZAWA, Y., WOODBINE, L., BJÖRKMAN, A., SHIMADA, M., FAWCETT, H., JIA, N., OHYAMA, K., LI, T.-S., NAGAYAMA, Y., MITSUTAKE, N., PAN-HAMMARSTRÖM, Q., GENNERY, A. R., LEHMANN, A. R., JEGGO, P. A. & OGI, T. 2015. XRCC4 deficiency in human subjects causes a marked neurological phenotype but no overt immunodeficiency. *Journal of Allergy and Clinical Immunology,* 136**,** 1007--1017.

GUO, Z., KUMAGAI, A., WANG, S. X. & 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. *Genes \& Development,* 14**,** 2745--2756.

GUPTA, R., SOMYAJIT, K., NARITA, T., MASKEY, E., STANLIE, A., KREMER, M., TYPAS, D., LAMMERS, M., MAILAND, N., NUSSENZWEIG, A., LUKAS, J. & CHOUDHARY, C. 2018. DNA Repair Network Analysis Reveals Shieldin as a Key Regulator of NHEJ and PARP Inhibitor Sensitivity. *Cell,* 173**,** 972--988.e23.

GURLEY, K. E. & KEMP, C. J. 2001. Synthetic lethality between mutation in Atm and DNA-PKcs during murine embryogenesis. *Current Biology,* 11**,** 191--194.

GURTAN, A. M., STUCKERT, P. & D'ANDREA, A. D. 2006. The WD40 Repeats of FANCL Are Required for Fanconi Anemia Core Complex Assembly. *Journal of Biological Chemistry,* 281**,** 10896--10905.

HAAHR, P., HOFFMANN, S., TOLLENAERE, M. A. X., HO, T., TOLEDO, L. I., MANN, M., BEKKER-JENSEN, S., RÄSCHLE, M. & MAILAND, N. 2016. Activation of the ATR kinase by the RPA-binding protein ETAA1. *Nature Cell Biology,* 18**,** 1196--1207.

HAAPANIEMI, E., BOTLA, S., PERSSON, J., SCHMIERER, B. & TAIPALE, J. 2018. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine,* 24**,** 927--930.

HADA, M. & GEORGAKILAS, A. G. 2008. Formation of Clustered DNA Damage after High-LET Irradiation: A Review. *Journal of Radiation Research,* 49**,** 203--210.
HALL, J. M., LEE, M. K., NEWMAN, B., MORROW, J. E., ANDERSON, L. A., HUEY, B. & KING, M.-C. 1990. Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21. *Science,* 250**,** 1684--1689.
HAMMEL, M., YU, Y., FANG, S., LEES-MILLER, S. P. & TAINER, J. A. 2010. XLF Regulates Filament Architecture of the XRCC4·Ligase IV Complex. *Structure,* 18**,** 1431--1442.
HANADA, K., BUDZOWSKA, M., MODESTI, M., MAAS, A., WYMAN, C., ESSERS, J. & KANAAR, R. 2006. The structure-specific endonuclease Mus81–Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks. *The EMBO Journal,* 25**,** 4921--4932.
HANAHAN, D. & WEINBERG, ROBERT A. 2011. Hallmarks of Cancer: The Next Generation. *Cell,* 144**,** 646--674.
HANNA, R. E., HEGDE, M., FAGRE, C. R., DEWEIRDT, P. C., SANGREE, A. K., SZEGLETES, Z., GRIFFITH, A., FEELEY, M. N., SANSON, K. R., BAIDI, Y., KOBLAN, L. W., LIU, D. R., NEAL, J. T. & DOENCH, J. G. 2021. Massively parallel assessment of human variants with base editor screens. *Cell,* 184**,** 1064--1080.e20.
HANZLIKOVA, H., KALASOVA, I., DEMIN, A. A., PENNICOTT, L. E., CIHLAROVA, Z. & CALDECOTT, K. W. 2018. The Importance of Poly(ADP-Ribose) Polymerase as a Sensor of Unligated Okazaki Fragments during DNA Replication. *Molecular Cell,* 71**,** 319--331.e3.
HAO, P., LEBLANC, S. J., CASE, B. C., ELSTON, T. C., HINGORANI, M. M., ERIE, D. A. & WENINGER, K. R. 2020. Recurrent mismatch binding by MutS mobile clamps on DNA localizes repair complexes nearby. *Proceedings of the National Academy of Sciences,* 117**,** 17775--17784.
HARACSKA, L., UNK, I., JOHNSON, R. E., JOHANSSON, E., BURGERS, P. M. J., PRAKASH, S. & PRAKASH, L. 2001. Roles of yeast DNA polymerases δ and ζ and of Rev1 in the bypass of abasic sites. *Genes \& Development,* 15**,** 945--954.
HARDT, K., HEICK, S. B., BETZ, B., GOECKE, T., YAZDANPARAST, H., KÜPPERS, R., SERVAN, K., STEINKE, V., RAHNER, N., MORAK, M., HOLINSKI-FEDER, E., ENGEL, C., MÖSLEIN, G., SCHACKERT, H.-K., DOEBERITZ, M. V. K., POX, C., PROPPING, P., CONSORTIUM, G. H., HEGEMANN, J. H. & ROYER-POKORA, B. 2011. Missense variants in hMLH1 identified in patients from the German HNPCC consortium and functional studies. *Familial Cancer,* 10**,** 273.
HARTLEY, J. L., TEMPLE, G. F. & BRASCH, M. A. 2000. DNA Cloning Using In Vitro Site-Specific Recombination. *Genome Research,* 10**,** 1788--1795.
HAYWARD, S. B. & CICCIA, A. 2021. Towards a CRISPeR understanding of homologous recombination with high-throughput functional genomics. *Current Opinion in Genetics \& Development,* 71**,** 171--181.
HE, M., HU, X., CHEN, L., CAO, A. Y., YU, K.-D., SHI, T.-Y., KUANG, X.-Y., SHI, W.-B., LING, H., LI, S., QIAO, F., YAO, L., WEI, Q., DI, G.-H. & SHAO, Z.-M. 2014. A recessive variant of XRCC4 predisposes to non-BRCA1/2 breast cancer in chinese women and impairs the DNA damage response via dysregulated nuclear localization. *Oncotarget,* 5**,** 12218--12232.
HE, Z., HENRICKSEN, L. A., WOLD, M. S. & INGLES, C. J. 1995. RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature,* 374**,** 566--569.
HEALD, R., MCLOUGHLIN, M. & MCKEON, F. 1993. Human wee1 maintains mitotic timing by protecting the nucleus from cytoplasmically activated cdc2 kinase. *Cell,* 74**,** 463--474.
HEINEN, C. D., CYR, J. L., COOK, C., PUNJA, N., SAKATO, M., FORTIES, R. A., LOPEZ, J. M., HINGORANI, M. M. & FISHEL, R. 2011. Human MSH2 (hMSH2) Protein Controls ATP Processing by hMSH2-hMSH6*. *Journal of Biological Chemistry,* 286**,** 40287--40295.
HEKMAT-NEJAD, M., YOU, Z., YEE, M.-C., NEWPORT, J. W. & CIMPRICH, K. A. 2000. Xenopus ATR is a replication-dependent chromatin-binding protein required for the DNA replication checkpoint. *Current Biology,* 10**,** 1565--1573.

HEMMINKI, A., PELTOMÄKI, P., MECKLIN, J.-P., JÄRVINEN, H., SALOVAARA, R., NYSTRÖM-LAHTI, M., CHAPELLE, A. D. L. & AALTONEN, L. A. 1994. Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. *Nature Genetics,* 8**,** 405--410.

HENDRIKS, I. A., D'SOUZA, R. C. J., YANG, B., VRIES, M. V.-D., MANN, M. & VERTEGAAL, A. C. O. 2014. Uncovering global SUMOylation signaling networks in a site-specific manner. *Nature Structural \& Molecular Biology,* 21**,** 927--936.

HENLE, E. S. & LINN, S. 1997. Formation, Prevention, and Repair of DNA Damage by Iron/Hydrogen Peroxide*. *Journal of Biological Chemistry,* 272**,** 19095--19098.

HENNER, W. D., GRUNBERG, S. M. & HASELTINE, W. A. 1982. Sites and structure of gamma radiation-induced DNA strand breaks. *The Journal of biological chemistry,* 257**,** 11750--4.

HERRMANN, G., LINDAHL, T. & SCHÄR, P. 1998. Saccharomyces cerevisiae LIF1: a function involved in DNA double-strand break repair related to mammalian XRCC4. *The EMBO Journal,* 17**,** 4188--4198.

HEWITT, G., BOREL, V., SEGURA-BAYONA, S., TAKAKI, T., RUIS, P., BELLELLI, R., LEHMANN, L. C., SOMMEROVA, L., VANCEVSKA, A., TOMAS-LOBA, A., ZHU, K., COOPER, C., FUGGER, K., PATEL, H., GOLDSTONE, R., SCHNEIDER-LUFTMAN, D., HERBERT, E., STAMP, G., BROUGH, R., PETTITT, S., LORD, C. J., WEST, S. C., AHEL, I., AHEL, D., CHAPMAN, J. R., DEINDL, S. & BOULTON, S. J. 2021. Defective ALC1 nucleosome remodeling confers PARPi sensitization and synthetic lethality with HRD. *Molecular Cell,* 81**,** 767--783.e11.

HIETER, P. & BOYCOTT, K. M. 2014. Understanding Rare Disease Pathogenesis: A Grand Challenge for Model Organisms. *Genetics,* 198**,** 443--445.

HILTON, B., SHKRIABAI, N., MUSICH, P. R., KVARATSKHELIA, M., SHELL, S. & ZOU, Y. 2014. A new structural insight into XPA–DNA interactions. *Bioscience Reports,* 34**,** e00162.

HIMMELS, S.-F. & SARTORI, A. A. 2016. Controlling DNA-End Resection: An Emerging Task for Ubiquitin and SUMO. *Frontiers in Genetics,* 7**,** 152.

HINRICHSEN, I., BRIEGER, A., TROJAN, J., ZEUZEM, S., NILBERT, M. & PLOTZ, G. 2013. Expression Defect Size among Unclassified MLH1 Variants Determines Pathogenicity in Lynch Syndrome Diagnosis. *Clinical Cancer Research,* 19**,** 2432--2441.

HIRA, A., YOSHIDA, K., SATO, K., OKUNO, Y., SHIRAISHI, Y., CHIBA, K., TANAKA, H., MIYANO, S., SHIMAMOTO, A., TAHARA, H., ITO, E., KOJIMA, S., KURUMIZAKA, H., OGAWA, S., TAKATA, M., YABE, H. & YABE, M. 2015. Mutations in the Gene Encoding the E2 Conjugating Enzyme UBE2T Cause Fanconi Anemia. *The American Journal of Human Genetics,* 96**,** 1001--1007.

HODSON, C., COLE, A. R., LEWIS, L. P. C., MILES, J. A., PURKISS, A. & WALDEN, H. 2011. Structural Analysis of Human FANCL, the E3 Ligase in the Fanconi Anemia Pathway. *Journal of Biological Chemistry,* 286**,** 32628--32637.

HODSON, C., PURKISS, A., MILES, JENNIFER A. & WALDEN, H. 2014. Structure of the Human FANCL RING-Ube2T Complex Reveals Determinants of Cognate E3-E2 Selection. *Structure,* 22**,** 337--344.

HOFSTATTER, E. W., DOMCHEK, S. M., MIRON, A., GARBER, J., WANG, M., COMPONESCHI, K., BOGHOSSIAN, L., MIRON, P. L., NATHANSON, K. L. & TUNG, N. 2011. PALB2 mutations in familial breast and pancreatic cancer. *Familial Cancer,* 10**,** 225--231.

HOLLINGSWORTH, N. M., PONTE, L. & HALSEY, C. 1995. MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in Saccharomyces cerevisiae but not mismatch repair. *Genes \& Development,* 9**,** 1728--1739.

HOMBAUER, H., CAMPBELL, CHRISTOPHER S., SMITH, CATHERINE E., DESAI, A. & KOLODNER, RICHARD D. 2011. Visualization of Eukaryotic DNA Mismatch Repair Reveals Distinct Recognition and Repair Intermediates. *Cell,* 147**,** 1040--1053.

HORAK, P., GRIFFITH, M., DANOS, A. M., PITEL, B. A., MADHAVAN, S., LIU, X., CHOW, C., WILLIAMS, H., CARMODY, L., BARROW-LAING, L., RIEKE, D., KREUTZFELDT, S., STENZINGER, A., TAMBORERO, D., BENARY, M., RAJAGOPAL, P. S., IDA, C. M., LESMANA, H., SATGUNASEELAN, L., MERKER, J. D., TOLSTORUKOV, M. Y., CAMPREGHER, P. V., WARNER, J. L., RAO, S.,

NATESAN, M., SHEN, H., VENSTROM, J., ROY, S., TAO, K., KANAGAL-SHAMANNA, R., XU, X., RITTER, D. I., PAGEL, K., KRYSIAK, K., DUBUC, A., AKKARI, Y. M., LI, X. S., LEE, J., KING, I., RACA, G., WAGNER, A. H., LI, M. M., PLON, S. E., KULKARNI, S., GRIFFITH, O. L., CHAKRAVARTY, D. & SONKIN, D. 2022. Standards for the classification of pathogenicity of somatic variants in cancer (oncogenicity): Joint recommendations of Clinical Genome Resource (ClinGen), Cancer Genomics Consortium (CGC), and Variant Interpretation for Cancer Consortium (VICC). *Genetics in Medicine*.

HORNBECK, P. V., ZHANG, B., MURRAY, B., KORNHAUSER, J. M., LATHAM, V. & SKRZYPEK, E. 2015. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research,* 43**,** D512--D520.

HOULLEBERGHS, H., DEKKER, M., LUSSEVELD, J., PIETERS, W., RAVESTEYN, T. V., VERHOEF, S., HOFSTRA, R. M. W. & RIELE, H. T. 2020. Three-step site-directed mutagenesis screen identifies pathogenic MLH1 variants associated with Lynch syndrome. *Journal of Medical Genetics,* 57**,** 308.

HOULLEBERGHS, H., GOVERDE, A., LUSSEVELD, J., DEKKER, M., BRUNO, M. J., MENKO, F. H., MENSENKAMP, A. R., SPAANDER, M. C. W., WAGNER, A., HOFSTRA, R. M. W. & RIELE, H. T. 2017. Suspected Lynch syndrome associated MSH6 variants: A functional assay to determine their pathogenicity. *PLoS Genetics,* 13**,** e1006765.

HU, Q., BOTUYAN, M. V., ZHAO, D., CUI, G., MER, E. & MER, G. 2021. Mechanisms of BRCA1–BARD1 nucleosome recognition and ubiquitylation. *Nature,* 596**,** 438--443.

HUANG, R. & ZHOU, P.-K. 2021. DNA damage repair: historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. *Signal Transduction and Targeted Therapy,* 6**,** 254.

HUANG, Y., LEUNG, JUSTIN W. C., LOWERY, M., MATSUSHITA, N., WANG, Y., SHEN, X., HUONG, D., TAKATA, M., CHEN, J. & LI, L. 2014. Modularized Functions of the Fanconi Anemia Core Complex. *Cell Reports,* 7**,** 1849--1857.

HUEN, M. S. Y., GRANT, R., MANKE, I., MINN, K., YU, X., YAFFE, M. B. & CHEN, J. 2007. RNF8 Transduces the DNA-Damage Signal via Histone Ubiquitylation and Checkpoint Protein Assembly. *Cell,* 131**,** 901--914.

HUERTAS, P. 2010. DNA resection in eukaryotes: deciding how to fix the break. *Nature Structural \& Molecular Biology,* 17**,** 11--16.

HUERTAS, P. & AGUILERA, A. 2003. Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination. *Molecular Cell,* 12**,** 711--721.

HUMPHREY, J. S., SALIM, A., ERDOS, M. R., COLLINS, F. S., BRODY, L. C. & KLAUSNER, R. D. 1997. Human BRCA1 inhibits growth in yeast: Potential use in diagnostic testing. *Proceedings of the National Academy of Sciences,* 94**,** 5820--5825.

HUNTER, N. 2015. Meiotic Recombination: The Essence of Heredity. *Cold Spring Harbor Perspectives in Biology,* 7**,** a016618.

HUNTER, N. & BORTS, R. H. 1997. Mlh1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis. *Genes \& Development,* 11**,** 1573--1582.

HUTCHINSON, F. 1985. Chemical Changes Induced in DNA by Ionizing Radiation. *Progress in Nucleic Acid Research and Molecular Biology,* 32**,** 115--154.

HUTTLIN, E. L., JEDRYCHOWSKI, M. P., ELIAS, J. E., GOSWAMI, T., RAD, R., BEAUSOLEIL, S. A., VILLÉN, J., HAAS, W., SOWA, M. E. & GYGI, S. P. 2010. A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. *Cell,* 143**,** 1174--1189.

IACCARINO, I., PALOMBO, F., DRUMMOND, J., TOTTY, N. F., HSUAN, J. J., MODRICH, P. & JIRICNY, J. 1996. MSH6, a Saccharomyces cerevisiae protein that binds to mismatches as a heterodimer with MSH2. *Current Biology,* 6**,** 484--486.

IHRY, R. J., WORRINGER, K. A., SALICK, M. R., FRIAS, E., HO, D., THERIAULT, K., KOMMINENI, S., CHEN, J., SONDEY, M., YE, C., RANDHAWA, R., KULKARNI, T., YANG, Z., MCALLISTER, G., RUSS, C.,

REECE-HOYES, J., FORRESTER, W., HOFFMAN, G. R., DOLMETSCH, R. & KAYKAS, A. 2018. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature Medicine,* 24**,** 939--946.

IKEGAMI, M., KOHSAKA, S., UENO, T., MOMOZAWA, Y., INOUE, S., TAMURA, K., SHIMOMURA, A., HOSOYA, N., KOBAYASHI, H., TANAKA, S. & MANO, H. 2020. High-throughput functional evaluation of BRCA2 variants of unknown significance. *Nature Communications,* 11**,** 2573.

IKEGAMI, T., KURAOKA, I., SAIJO, M., KODO, N., KYOGOKU, Y., MORIKAWA, K., TANAKA, K. & SHIRAKAWA, M. 1998. Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA. *Nature Structural Biology,* 5**,** 701--706.

ILES, N., RULTEN, S., EL-KHAMISY, S. F. & CALDECOTT, K. W. 2007. APLF (C2orf13) Is a Novel Human Protein Involved in the Cellular Response to Chromosomal DNA Strand Breaks. *Molecular and Cellular Biology,* 27**,** 3793--3803.

INDIANI, C., MCINERNEY, P., GEORGESCU, R., GOODMAN, M. F. & O'DONNELL, M. 2005. A Sliding-Clamp Toolbelt Binds High- and Low-Fidelity DNA Polymerases Simultaneously. *Molecular Cell,* 19**,** 805--815.

IONOV, Y., PEINADO, M. A., MALKHOSYAN, S., SHIBATA, D. & PERUCHO, M. 1993. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature,* 363**,** 558--561.

IP, S. C. Y., RASS, U., BLANCO, M. G., FLYNN, H. R., SKEHEL, J. M. & WEST, S. C. 2008. Identification of Holliday junction resolvases from humans and yeast. *Nature,* 456**,** 357--361.

JACKSON, S. P. & BARTEK, J. 2009. The DNA-damage response in human biology and disease. *Nature,* 461**,** 1071.

JACOB, H. J., ABRAMS, K., BICK, D. P., BRODIE, K., DIMMOCK, D. P., FARRELL, M., GEURTS, J., HARRIS, J., HELBLING, D., JOERS, B. J., KLIEGMAN, R., KOWALSKI, G., LAZAR, J., MARGOLIS, D. A., NORTH, P., NORTHUP, J., ROQUEMORE-GOINS, A., SCHARER, G., SHIMOYAMA, M., STRONG, K., TAYLOR, B., TSAIH, S.-W., TSCHANNEN, M. R., VEITH, R. L., WENDT-ANDRAE, J., WILK, B. & WORTHEY, E. A. 2013. Genomics in Clinical Practice: Lessons from the Front Lines. *Science Translational Medicine,* 5**,** 194cm5.

JÄGER, A. C., RASMUSSEN, M., BISGAARD, H. C., SINGH, K. K., NIELSEN, F. C. & RASMUSSEN, L. J. 2001. HNPCC mutations in the human DNA mismatch repair gene hMLH1 influence assembly of hMutLα and hMLH1–hEXO1 complexes. *Oncogene,* 20**,** 3590--3595.

JANES, K. A. 2015. An analysis of critical factors for quantitative immunoblotting. *Science Signaling,* 8**,** rs2.

JÄRVINEN*, H. J., AARNIO*, M., MUSTONEN*, H., AKTAN–COLLAN‡, K., AALTONEN‡, L. A., PELTOMÄKI§, P., CHAPELLE§, A. D. L. & MECKLIN‖, J. P. 2000. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology,* 118**,** 829--834.

JEGGO, P. A. 1998. Identification of genes involved in repair of DNA double-strand breaks in mammalian cells. *Radiation research,* 150**,** S80--91.

JENSEN, R. B., CARREIRA, A. & KOWALCZYKOWSKI, S. C. 2010. Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature,* 467**,** 678--683.

JIA, X., BURUGULA, B. B., CHEN, V., LEMONS, R. M., JAYAKODY, S., MAKSUTOVA, M. & KITZMAN, J. O. 2021. Massively parallel functional testing of MSH2 missense variants conferring Lynch syndrome risk. *The American Journal of Human Genetics,* 108**,** 163--175.

JIANG, W., CAI, M. Y., LI, S. Y., BEI, J. X., WANG, F., HAMPEL, H., LING, Y. H., FRAYLING, I. M., SINICROPE, F. A., RODRIGUEZ-BIGAS, M. A., DIGNAM, J. J., KERR, D. J., ROSELL, R., MAO, M., LI, J. B., GUO, Y. M., WU, X. Y., KONG, L. H., TANG, J. H., WU, X. D., LI, C. F., CHEN, J. R., OU, Q. J., YE, M. Z., GUO, F. M., HAN, P., WANG, Q. W., WAN, D. S., LI, L., XU, R. H., PAN, Z. Z., DING, P. R. & GROUP, W. O. B. O. A. M. E. C. C. C. 2019. Universal screening for Lynch syndrome in a large consecutive cohort of Chinese colorectal cancer patients: High prevalence and unique molecular features. *International Journal of Cancer,* 144**,** 2161--2168.

JIN, J., CHEN, Z., LIU, J., DU, H. & ZHANG, G. 2022. Towards an accurate and robust analysis pipeline for somatic mutation calling. *Frontiers in Genetics,* 13**,** 979928.

JIRICNY, J. 2006. The multifaceted mismatch-repair system. *Nature Reviews Molecular Cell Biology,* 7**,** 335--346.

JOHNSON, J. R., ERDENIZ, N., NGUYEN, M., DUDLEY, S. & LISKAY, R. M. 2010. Conservation of functional asymmetry in the mammalian MutLα ATPase. *DNA Repair,* 9**,** 1209--1213.

JONG, A. E. D., PUIJENBROEK, M. V., HENDRIKS, Y., TOPS, C., WIJNEN, J., AUSEMS, M. G. E. M., MEIJERS-HEIJBOER, H., WAGNER, A., OS, T. A. M. V., BRÖCKER-VRIENDS, A. H. J. T., VASEN, H. F. A. & MORREAU, H. 2004. Microsatellite Instability, Immunohistochemistry, and Additional PMS2 Staining in Suspected Hereditary Nonpolyposis Colorectal Cancer. *Clinical Cancer Research,* 10**,** 972--980.

JÓNSSON, H., SULEM, P., KEHR, B., KRISTMUNDSDOTTIR, S., ZINK, F., HJARTARSON, E., HARDARSON, M. T., HJORLEIFSSON, K. E., EGGERTSSON, H. P., GUDJONSSON, S. A., WARD, L. D., ARNADOTTIR, G. A., HELGASON, E. A., HELGASON, H., GYLFASON, A., JONASDOTTIR, A., JONASDOTTIR, A., RAFNAR, T., FRIGGE, M., STACEY, S. N., MAGNUSSON, O. T., THORSTEINSDOTTIR, U., MASSON, G., KONG, A., HALLDORSSON, B. V., HELGASON, A., GUDBJARTSSON, D. F. & STEFANSSON, K. 2017. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature,* 549**,** 519--522.

JOWSEY, P., MORRICE, N. A., HASTIE, C. J., MCLAUCHLAN, H., TOTH, R. & ROUSE, J. 2007. Characterisation of the sites of DNA damage-induced 53BP1 phosphorylation catalysed by ATM and ATR. *DNA Repair,* 6**,** 1536--1544.

JUHÁSZ, S., SMITH, R., SCHAUER, T., SPEKHARDT, D., MAMAR, H., ZENTOUT, S., CHAPUIS, C., HUET, S. & TIMINSZKY, G. 2020. The chromatin remodeler ALC1 underlies resistance to PARP inhibitor treatment. *Science Advances,* 6.

JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ŽÍDEK, A., POTAPENKO, A., BRIDGLAND, A., MEYER, C., KOHL, S. A. A., BALLARD, A. J., COWIE, A., ROMERA-PAREDES, B., NIKOLOV, S., JAIN, R., ADLER, J., BACK, T., PETERSEN, S., REIMAN, D., CLANCY, E., ZIELINSKI, M., STEINEGGER, M., PACHOLSKA, M., BERGHAMMER, T., BODENSTEIN, S., SILVER, D., VINYALS, O., SENIOR, A. W., KAVUKCUOGLU, K., KOHLI, P. & HASSABIS, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature,* 596**,** 583--589.

JUNOP, M. S., MODESTI, M., GUARNÉ, A., GHIRLANDO, R., GELLERT, M. & YANG, W. 2000. Crystal structure of the Xrcc4 DNA repair protein and implications for end joining. *The EMBO Journal,* 19**,** 5962--5970.

KACHROO, A. H., LAURENT, J. M., YELLMAN, C. M., MEYER, A. G., WILKE, C. O. & MARCOTTE, E. M. 2015. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science,* 348**,** 921--925.

KADYROV, F. A., DZANTIEV, L., CONSTANTIN, N. & MODRICH, P. 2006. Endonucleolytic Function of MutLα in Human Mismatch Repair. *Cell,* 126**,** 297--308.

KADYROV, F. A., GENSCHEL, J., FANG, Y., PENLAND, E., EDELMANN, W. & MODRICH, P. 2009. A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. *Proceedings of the National Academy of Sciences,* 106**,** 8495--8500.

KADYROV, F. A., HOLMES, S. F., ARANA, M. E., LUKIANOVA, O. A., O'DONNELL, M., KUNKEL, T. A. & MODRICH, P. 2007. Saccharomyces cerevisiae MutLα Is a Mismatch Repair Endonuclease*. *Journal of Biological Chemistry,* 282**,** 37181--37190.

KANSIKAS, M., KARIOLA, R. & NYSTRÖM, M. 2011. Verification of the three-step model in assessing the pathogenicity of mismatch repair gene variants. *Human Mutation,* 32**,** 107--115.

KARCHIN, R., MONTEIRO, A. N. A., TAVTIGIAN, S. V., CARVALHO, M. A. & SALI, A. 2007. Functional Impact of Missense Variants in BRCA1 Predicted by Supervised Learning. *PLoS Computational Biology,* 3**,** e26.

KARCZEWSKI, K. J., FRANCIOLI, L. C., TIAO, G., CUMMINGS, B. B., ALFÖLDI, J., WANG, Q., COLLINS, R. L., LARICCHIA, K. M., GANNA, A., BIRNBAUM, D. P., GAUTHIER, L. D., BRAND, H., SOLOMONSON, M., WATTS, N. A., RHODES, D., SINGER-BERK, M., ENGLAND, E. M., SEABY, E. G., KOSMICKI, J. A., WALTERS, R. K., TASHMAN, K., FARJOUN, Y., BANKS, E., POTERBA, T., WANG, A., SEED, C., WHIFFIN, N., CHONG, J. X., SAMOCHA, K. E., PIERCE-HOFFMAN, E., ZAPPALA, Z., O'DONNELL-LURIA, A. H., MINIKEL, E. V., WEISBURD, B., LEK, M., WARE, J. S., VITTAL, C., ARMEAN, I. M., BERGELSON, L., CIBULSKIS, K., CONNOLLY, K. M., COVARRUBIAS, M., DONNELLY, S., FERRIERA, S., GABRIEL, S., GENTRY, J., GUPTA, N., JEANDET, T., KAPLAN, D., LLANWARNE, C., MUNSHI, R., NOVOD, S., PETRILLO, N., ROAZEN, D., RUANO-RUBIO, V., SALTZMAN, A., SCHLEICHER, M., SOTO, J., TIBBETTS, K., TOLONEN, C., WADE, G., TALKOWSKI, M. E., SALINAS, C. A. A., AHMAD, T., ALBERT, C. M., ARDISSINO, D., ATZMON, G., BARNARD, J., BEAUGERIE, L., BENJAMIN, E. J., BOEHNKE, M., BONNYCASTLE, L. L., BOTTINGER, E. P., BOWDEN, D. W., BOWN, M. J., CHAMBERS, J. C., CHAN, J. C., CHASMAN, D., CHO, J., CHUNG, M. K., COHEN, B., CORREA, A., DABELEA, D., DALY, M. J., DARBAR, D., DUGGIRALA, R., DUPUIS, J., ELLINOR, P. T., ELOSUA, R., ERDMANN, J., ESKO, T., FÄRKKILÄ, M., FLOREZ, J., FRANKE, A., GETZ, G., GLASER, B., GLATT, S. J., GOLDSTEIN, D., GONZALEZ, C., GROOP, L., et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature,* 581**,** 434--443.

KARNITZ, L. M., FLATTEN, K. S., WAGNER, J. M., LOEGERING, D., HACKBARTH, J. S., ARLANDER, S. J. H., VROMAN, B. T., THOMAS, M. B., BAEK, Y.-U., HOPKINS, K. M., LIEBERMAN, H. B., CHEN, J., CLIBY, W. A. & KAUFMANN, S. H. 2005. Gemcitabine-Induced Activation of Checkpoint Signaling Pathways That Affect Tumor Cell Survival. *Molecular Pharmacology,* 68**,** 1636--1644.

KARRAN, P. & MARINUS, M. G. 1982. Mismatch correction at O6-methylguanine residues in E. coli DNA. *Nature,* 296**,** 868--869.

KASTAN, M. B. & BARTEK, J. 2004. Cell-cycle checkpoints and cancer. *Nature,* 432**,** 316--323.

KATO, A., SATO, N., SUGAWARA, T., TAKAHASHI, K., KITO, M., MAKINO, K., SATO, T., SHIMIZU, D., SHIRASAWA, H., MIURA, H., SATO, W., KUMAZAWA, Y., SATO, A., KUMAGAI, J. & TERADA, Y. 2016. Isolated Loss of PMS2 Immunohistochemical Expression is Frequently Caused by Heterogenous MLH1 Promoter Hypermethylation in Lynch Syndrome Screening for Endometrial Cancer Patients. *The American Journal of Surgical Pathology,* 40**,** 770--776.

KAZAZIAN, H. H., BOEHM, C. D. & SELTZER, W. K. 2000. ACMG recommendations for standards for interpretation of sequence variations. *Genetics in Medicine,* 2**,** 302--303.

KCIUK, M., MARCINIAK, B., MOJZYCH, M. & KONTEK, R. 2020. Focus on UV-Induced DNA Damage and Repair—Disease Relevance and Protective Strategies. *International Journal of Molecular Sciences,* 21**,** 7264.

KELLAND, L. 2007. The resurgence of platinum-based cancer chemotherapy. *Nature Reviews Cancer,* 7**,** 573--584.

KELLY, M. A., CALESHU, C., MORALES, A., BUCHAN, J., WOLF, Z., HARRISON, S. M., COOK, S., DILLON, M. W., GARCIA, J., HAVERFIELD, E., JONGBLOED, J. D. H., MACAYA, D., MANRAI, A., ORLAND, K., RICHARD, G., SPOONAMORE, K., THOMAS, M., THOMSON, K., VINCENT, L. M., WALSH, R., WATKINS, H., WHIFFIN, N., INGLES, J., TINTELEN, J. P. V., SEMSARIAN, C., WARE, J. S., HERSHBERGER, R., FUNKE, B. & GROUP, F. T. C. C. D. W. 2018. Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genetics in Medicine,* 20**,** 351--359.

KEREM, B. S., ZIELENSKI, J., MARKIEWICZ, D., BOZON, D., GAZIT, E., YAHAV, J., KENNEDY, D., RIORDAN, J. R., COLLINS, F. S. & ROMMENS, J. M. 1990. Identification of mutations in regions corresponding to the two putative nucleotide (ATP)-binding folds of the cystic fibrosis gene. *Proceedings of the National Academy of Sciences,* 87**,** 8447--8451.

KERN, A. D. & HAHN, M. W. 2018. The Neutral Theory in Light of Natural Selection. *Molecular Biology and Evolution,* 35**,** 1366--1371.

KHACHIBI, M. E., DIAKITE, B., HAMZI, K., BADOU, A., SENHAJI, M. A., BAKHCHANE, A., JOUHADI, H., BARAKAT, A., BENIDER, A. & NADIFI, S. 2015. Screening of exon 11 of BRCA1 gene using the high resolution melting approach for diagnosis in Moroccan breast cancer patients. *BMC Cancer,* 15**,** 81.

KHERA, A. V., CHAFFIN, M., ARAGAM, K. G., HAAS, M. E., ROSELLI, C., CHOI, S. H., NATARAJAN, P., LANDER, E. S., LUBITZ, S. A., ELLINOR, P. T. & KATHIRESAN, S. 2018. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics,* 50**,** 1219--1224.

KIM, H., YANG, K., DEJSUPHONG, D. & D'ANDREA, A. D. 2012. Regulation of Rev1 by the Fanconi anemia core complex. *Nature Structural \& Molecular Biology,* 19**,** 164--170.

KIM, J. M., KEE, Y., GURTAN, A. & D'ANDREA, A. D. 2008. Cell cycle–dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24. *Blood,* 111**,** 5215--5222.

KIM, M., KIM, H.-S., D'SOUZA, A., GALLAGHER, K., JEONG, E., TOPOLSKA-WOŚ, A., MEUR, K. O. L., TSAI, C.-L., TSAI, M.-S., KEE, M., TAINER, J. A., YEO, J.-E., CHAZIN, W. J. & SCHÄRER, O. D. 2022. Two interaction surfaces between XPA and RPA organize the preincision complex in nucleotide excision repair. *Proceedings of the National Academy of Sciences,* 119**,** e2207408119.

KIM, W., BENNETT, ERIC J., HUTTLIN, EDWARD L., GUO, A., LI, J., POSSEMATO, A., SOWA, MATHEW E., RAD, R., RUSH, J., COMB, MICHAEL J., HARPER, J. W. & GYGI, STEVEN P. 2011. Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome. *Molecular Cell,* 44**,** 325--340.

KING, B. S., COOPER, K. L., LIU, K. J. & HUDSON, L. G. 2012. Poly(ADP-ribose) Contributes to an Association between Poly(ADP-ribose) Polymerase-1 and Xeroderma Pigmentosum Complementation Group A in Nucleotide Excision Repair*. *Journal of Biological Chemistry,* 287**,** 39824--39833.

KLECZKOWSKA, H. E., MARRA, G., LETTIERI, T. & JIRICNY, J. 2001. hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes \& Development,* 15**,** 724--736.

KLEIN, R. J., ZEISS, C., CHEW, E. Y., TSAI, J.-Y., SACKLER, R. S., HAYNES, C., HENNING, A. K., SANGIOVANNI, J. P., MANE, S. M., MAYNE, S. T., BRACKEN, M. B., FERRIS, F. L., OTT, J., BARNSTABLE, C. & HOH, J. 2005. Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science,* 308**,** 385--389.

KLEIN DOUWEL, D., BOONEN, RICK A. C. M., LONG, DAVID T., SZYPOWSKA, ANNA A., RÄSCHLE, M., WALTER, JOHANNES C. & KNIPSCHEER, P. 2014. XPF-ERCC1 Acts in Unhooking DNA Interstrand Crosslinks in Cooperation with FANCD2 and FANCP/SLX4. *Molecular Cell,* 54**,** 460--471.

KNEITZ, B., COHEN, P. E., AVDIEVICH, E., ZHU, L., KANE, M. F., HOU, H., KOLODNER, R. D., KUCHERLAPATI, R., POLLARD, J. W. & EDELMANN, W. 2000. MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. *Genes \& Development,* 14**,** 1085--1097.

KNIPSCHEER, P., RÄSCHLE, M., SMOGORZEWSKA, A., ENOIU, M., HO, T. V., SCHÄRER, O. D., ELLEDGE, S. J. & WALTER, J. C. 2009. The Fanconi Anemia Pathway Promotes Replication-Dependent DNA Interstrand Cross-Link Repair. *Science,* 326**,** 1698--1701.

KNUDSON, A. G. 1985. Hereditary cancer, oncogenes, and antioncogenes. *Cancer research,* 45**,** 1437--43.

KOBOLDT, D. C. 2020. Best practices for variant calling in clinical sequencing. *Genome Medicine,* 12**,** 91.

KOBOLDT, D. C., ZHANG, Q., LARSON, D. E., SHEN, D., MCLELLAN, M. D., LIN, L., MILLER, C. A., MARDIS, E. R., DING, L. & WILSON, R. K. 2012. VarScan 2: Somatic mutation and copy

number alteration discovery in cancer by exome sequencing. *Genome Research,* 22**,** 568--576.
KOCH, C. A., AGYEI, R., GALICIA, S., METALNIKOV, P., O'DONNELL, P., STAROSTINE, A., WEINFELD, M. & DUROCHER, D. 2004. Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *The EMBO Journal,* 23**,** 3874--3885.
KOIKE, M., YUTOKU, Y. & KOIKE, A. 2016. Cloning, localization and focus formation at DNA damage sites of canine XRCC4. *Journal of Veterinary Medical Science,* 78**,** 1865--1871.
KOIKE, M., YUTOKU, Y. & KOIKE, A. 2022. Feline XRCC4 undergoes rapid Ku-dependent recruitment to DNA damage sites. *FEBS Open Bio,* 12**,** 798--810.
KOLAS, N. K., CHAPMAN, J. R., NAKADA, S., YLANKO, J., CHAHWAN, R., SWEENEY, F. D. R. D., PANIER, S., MENDEZ, M., WILDENHAIN, J., THOMSON, T. M., PELLETIER, L., JACKSON, S. P. & DUROCHER, D. 2007. Orchestration of the DNA-Damage Response by the RNF8 Ubiquitin Ligase. *Science,* 318**,** 1637--1640.
KOLAS, N. K., SVETLANOV, A., LENZI, M. L., MACALUSO, F. P., LIPKIN, S. M., LISKAY, R. M., GREALLY, J., EDELMANN, W. & COHEN, P. E. 2005. Localization of MMR proteins on meiotic chromosomes in mice indicates distinct functions during prophase I. *The Journal of Cell Biology,* 171**,** 447--458.
KOMOR, A. C., BADRAN, A. H. & LIU, D. R. 2017. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell,* 168**,** 20--36.
KONDO, E., HORII, A. & FUKUSHIGE, S. 2001. The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. *Nucleic Acids Research,* 29**,** 1695--1702.
KONDO, E., SUZUKI, H., HORII, A. & FUKUSHIGE, S. 2003. A yeast two-hybrid assay provides a simple way to evaluate the vast majority of hMLH1 germ-line mutations. *Cancer research,* 63**,** 3302--8.
KOSINSKI, J., HINRICHSEN, I., BUJNICKI, J. M., FRIEDHOFF, P. & PLOTZ, G. 2010. Identification of Lynch syndrome mutations in the MLH1–PMS2 interface that disturb dimerization and mismatch repair. *Human Mutation,* 31**,** 975--982.
KOWALCZYKOWSKI, S. C. 2015. An Overview of the Molecular Mechanisms of Recombinational DNA Repair. *Cold Spring Harbor Perspectives in Biology,* 7**,** a016410.
KRAIS, J. J., WANG, Y., PATEL, P., BASU, J., BERNHARDY, A. J. & JOHNSON, N. 2021. RNF168-mediated localization of BARD1 recruits the BRCA1-PALB2 complex to DNA damage. *Nature Communications,* 12**,** 5016.
KRASTEV, D. B., LI, S., SUN, Y., WICKS, A. J., HOSLETT, G., WEEKES, D., BADDER, L. M., KNIGHT, E. G., MARLOW, R., PARDO, M. C., YU, L., TALELE, T. T., BARTEK, J., CHOUDHARY, J. S., POMMIER, Y., PETTITT, S. J., TUTT, A. N. J., RAMADAN, K. & LORD, C. J. 2022. The ubiquitin-dependent ATPase p97 removes cytotoxic trapped PARP1 from chromatin. *Nature Cell Biology,* 24**,** 62--73.
KRATZ, K., SCHÖPF, B., KADEN, S., SENDOEL, A., EBERHARD, R., LADEMANN, C., CANNAVÓ, E., SARTORI, A. A., HENGARTNER, M. O. & JIRICNY, J. 2010. Deficiency of FANCD2-Associated Nuclease KIAA1018/FAN1 Sensitizes Cells to Interstrand Crosslinking Agents. *Cell,* 142**,** 77--88.
KROKAN, H. E. & BJØRÅS, M. 2013. Base Excision Repair. *Cold Spring Harbor Perspectives in Biology,* 5**,** a012583.
KRUSCHE, P., TRIGG, L., BOUTROS, P. C., MASON, C. E., VEGA, F. M. D. L., MOORE, B. L., GONZALEZ-PORTA, M., EBERLE, M. A., TEZAK, Z., LABABIDI, S., TRUTY, R., ASIMENOS, G., FUNKE, B., FLEHARTY, M., CHAPMAN, B. A., SALIT, M. & ZOOK, J. M. A. 2019. Best practices for benchmarking germline small-variant calls in human genomes. *Nature Biotechnology,* 37**,** 555--560.

KRYSTON, T. B., GEORGIEV, A. B., PISSIS, P. & GEORGAKILAS, A. G. 2011. Role of oxidative stress and DNA damage in human carcinogenesis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 711**,** 193--201.

KUHLEN, M. & BORKHARDT, A. 2018. Trio sequencing in pediatric cancer and clinical implications. *EMBO Molecular Medicine,* 10**,** e8641.

KULKARNI, D. S., OWENS, S. N., HONDA, M., ITO, M., YANG, Y., CORRIGAN, M. W., CHEN, L., QUAN, A. L. & HUNTER, N. 2020. PCNA activates the MutLγ endonuclease to promote meiotic crossing over. *Nature,* 586**,** 623--627.

KUMAGAI, A., LEE, J., YOO, H. Y. & DUNPHY, W. G. 2006. TopBP1 Activates the ATR-ATRIP Complex. *Cell,* 124**,** 943--955.

KUMAR, D., ABDULOVIC, A. L., VIBERG, J., NILSSON, A. K., KUNKEL, T. A. & CHABES, A. 2011. Mechanisms of mutagenesis in vivo due to imbalanced dNTP pools. *Nucleic Acids Research,* 39**,** 1360--1371.

KUNKEL, T. A. & ERIE, D. A. 2005. DNA Mismatch Repair*. *Annual Review of Biochemistry,* 74**,** 681--710.

KURAOKA, I., MORITA, E. H., SAIJO, M., MATSUDA, T., MORIKAWA, K., SHIRAKAWA, M. & TANAKA, K. 1996. Identification of a damaged-DNA binding domain of the XPA protein. *Mutation Research/DNA Repair,* 362**,** 87--95.

KUSAKABE, M., ONISHI, Y., TADA, H., KURIHARA, F., KUSAO, K., FURUKAWA, M., IWAI, S., YOKOI, M., SAKAI, W. & SUGASAWA, K. 2019. Mechanism and regulation of DNA damage recognition in nucleotide excision repair. *Genes and Environment,* 41**,** 2.

KUZNETSOV, S. G., LIU, P. & SHARAN, S. K. 2008. Mouse embryonic stem cell–based functional assay to evaluate mutations in BRCA2. *Nature Medicine,* 14**,** 875--881.

KWEON, J., JANG, A.-H., SHIN, H. R., SEE, J.-E., LEE, W., LEE, J. W., CHANG, S., KIM, K. & KIM, Y. 2020. A CRISPR-based base-editing screen for the functional assessment of BRCA1 variants. *Oncogene,* 39**,** 30--35.

LACHAUD, C., CASTOR, D., HAIN, K., MUÑOZ, I., WILSON, J., MACARTNEY, T. J., SCHINDLER, D. & ROUSE, J. 2014. Distinct functional roles for the two SLX4 ubiquitin-binding UBZ domains mutated in Fanconi anemia. *Journal of Cell Science,* 127**,** 2811--2817.

LADDACH, A., NG, J. C.-F., CHUNG, S. S. & FRATERNALI, F. 2018. Genetic variants and protein–protein interactions: a multidimensional network-centric view. *Current Opinion in Structural Biology,* 50**,** 82--90.

LAHUE, R. S., AU, K. G. & MODRICH, P. 1989. DNA Mismatch Correction in a Defined System. *Science,* 245**,** 160--164.

LAMBERT, S., MASON, S. J., BARBER, L. J., HARTLEY, J. A., PEARCE, J. A., CARR, A. M. & MCHUGH, P. J. 2003. Schizosaccharomyces pombe Checkpoint Response to DNA Interstrand Cross-Links. *Molecular and Cellular Biology,* 23**,** 4728--4737.

LAN, H. & SUN, Y. 2021. Tumor Suppressor FBXW7 and Its Regulation of DNA Damage Response and Repair. *Frontiers in Cell and Developmental Biology,* 9**,** 751574.

LANDRUM, M. J. & KATTMAN, B. L. 2018. ClinVar at five years: Delivering on the promise. *Human Mutation,* 39**,** 1623--1630.

LANDRUM, M. J., LEE, J. M., BENSON, M., BROWN, G. R., CHAO, C., CHITIPIRALLA, S., GU, B., HART, J., HOFFMAN, D., JANG, W., KARAPETYAN, K., KATZ, K., LIU, C., MADDIPATLA, Z., MALHEIRO, A., MCDANIEL, K., OVETSKY, M., RILEY, G., ZHOU, G., HOLMES, J. B., KATTMAN, B. L. & MAGLOTT, D. R. 2017. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research,* 46**,** gkx1153--.

LANDRUM, M. J., LEE, J. M., RILEY, G. R., JANG, W., RUBINSTEIN, W. S., CHURCH, D. M. & MAGLOTT, D. R. 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research,* 42**,** D980--D985.

LAPPALAINEN, T. & MACARTHUR, D. G. 2021. From variant to function in human disease genetics. *Science,* 373**,** 1464--1468.

LE, D. T., DURHAM, J. N., SMITH, K. N., WANG, H., BARTLETT, B. R., AULAKH, L. K., LU, S., KEMBERLING, H., WILT, C., LUBER, B. S., WONG, F., AZAD, N. S., RUCKI, A. A., LAHERU, D., DONEHOWER, R., ZAHEER, A., FISHER, G. A., CROCENZI, T. S., LEE, J. J., GRETEN, T. F., DUFFY, A. G., CIOMBOR, K. K., EYRING, A. D., LAM, B. H., JOE, A., KANG, S. P., HOLDHOFF, M., DANILOVA, L., COPE, L., MEYER, C., ZHOU, S., GOLDBERG, R. M., ARMSTRONG, D. K., BEVER, K. M., FADER, A. N., TAUBE, J., HOUSSEAU, F., SPETZLER, D., XIAO, N., PARDOLL, D. M., PAPADOPOULOS, N., KINZLER, K. W., ESHLEMAN, J. R., VOGELSTEIN, B., ANDERS, R. A. & DIAZJR, L. A. 2017. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science,* 357**,** 409--413.

LE, D. T., URAM, J. N., WANG, H., BARTLETT, B. R., KEMBERLING, H., EYRING, A. D., SKORA, A. D., LUBER, B. S., AZAD, N. S., LAHERU, D., BIEDRZYCKI, B., DONEHOWER, R. C., ZAHEER, A., FISHER, G. A., CROCENZI, T. S., LEE, J. J., DUFFY, S. M., GOLDBERG, R. M., CHAPELLE, A. D. L., KOSHIJI, M., BHAIJEE, F., HUEBNER, T., HRUBAN, R. H., WOOD, L. D., CUKA, N., PARDOLL, D. M., PAPADOPOULOS, N., KINZLER, K. W., ZHOU, S., CORNISH, T. C., TAUBE, J. M., ANDERS, R. A., ESHLEMAN, J. R., VOGELSTEIN, B. & DIAZ, L. A. 2015. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *The New England Journal of Medicine,* 372**,** 2509--2520.

LEACH, F. S., NICOLAIDES, N. C., PAPADOPOULOS, N., LIU, B., JEN, J., PARSONS, R., PELTOMÄKI, P., SISTONEN, P., AALTONEN, L. A., NYSTRÖM-LAHTI, M., GUAN, X. Y., ZHANG, J., MELTZER, P. S., YU, J.-W., KAO, F.-T., CHEN, D. J., CEROSALETTI, K. M., FOURNIER, R. E. K., TODD, S., LEWIS, T., LEACH, R. J., NAYLOR, S. L., WEISSENBACH, J., MECKLIN, J.-P., JÄRVINEN, H., PETERSEN, G. M., HAMILTON, S. R., GREEN, J., JASS, J., WATSON, P., LYNCH, H. T., TRENT, J. M., CHAPELLE, A. D. L., KINZLER, K. W. & VOGELSTEIN, B. 1993. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell,* 75**,** 1215--1225.

LEBER, R., WISE, T. W., MIZUTA, R. & MEEK, K. 1998. The XRCC4 Gene Product Is a Target for and Interacts with the DNA-dependent Protein Kinase*. *Journal of Biological Chemistry,* 273**,** 1794--1801.

LEE, C., BANERJEE, T., GILLESPIE, J., CERAVOLO, A., PARVINSMITH, M. R., STARITA, L. M., FIELDS, S., TOLAND, A. E. & PARVIN, J. D. 2015a. Functional Analysis of BARD1 Missense Variants in Homology-Directed Repair of DNA Double Strand Breaks. *Human Mutation,* 36**,** 1205--1214.

LEE, J.-B., CHO, W.-K., PARK, J., JEON, Y., KIM, D., LEE, S. H. & FISHEL, R. 2014a. Single-molecule views of MutS on mismatched DNA. *DNA Repair,* 20**,** 82--93.

LEE, J.-H. & PAULL, T. T. 2005. ATM Activation by DNA Double-Strand Breaks Through the Mre11-Rad50-Nbs1 Complex. *Science,* 308**,** 551--554.

LEE, K.-J., DONG, X., WANG, J., TAKEDA, Y. & DYNAN, W. S. 2002. Identification of Human Autoantibodies to the DNA Ligase IV/XRCC4 Complex and Mapping of an Autoimmune Epitope to a Potential Regulatory Region. *The Journal of Immunology,* 169**,** 3413--3421.

LEE, K.-J., JOVANOVIC, M., UDAYAKUMAR, D., BLADEN, C. L. & DYNAN, W. S. 2004. Identification of DNA-PKcs phosphorylation sites in XRCC4 and effects of mutations at these sites on DNA end joining in a cell-free system. *DNA Repair,* 3**,** 267--276.

LEE, M.-S., YU, M., KIM, K.-Y., PARK, G.-H., KWACK, K. & KIM, K. P. 2015b. Functional Validation of Rare Human Genetic Variants Involved in Homologous Recombination Using Saccharomyces cerevisiae. *PLoS ONE,* 10**,** e0124152.

LEE, M. G. & 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. S., GREEN, R., MARSILLAC, S. M., COQUELLE, N., WILLIAMS, R. S., YEUNG, T., FOO, D., HAU, D. D., HUI, B., MONTEIRO, A. N. A. & GLOVER, J. N. M. 2010. Comprehensive Analysis of Missense Variations in the BRCT Domain of BRCA1 by Structural and Functional Assays. *Cancer Research,* 70**,** 4880--4890.

LEE, T. H., PARK, J. M., LEEM, S. H. & KANG, T. H. 2014b. Coordinated regulation of XPA stability by ATR and HERC2 during nucleotide excision repair. *Oncogene,* 33**,** 19--25.

LEES-MILLER, S. P., CHEN, Y. R. & 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. D. M., DAY, M., BLAAUWENDRAAD, J., BUNDSCHUH, D., BONHAM, S., FISCHER, R., FINK, D., KESSLER, B. M., OLIVER, A. W., PEARL, L. H., BLACKFORD, A. N. & STUCKI, M. 2019. MDC1 Interacts with TOPBP1 to Maintain Chromosomal Stability during Mitosis. *Molecular Cell,* 74**,** 571--583.e8.

LELLIS, L. D., ACETO, G. M., CURIA, M. C., CATALANO, T., MAMMARELLA, S., VESCHI, S., FANTINI, F., BATTISTA, P., STIGLIANO, V., MESSERINI, L., MARENI, C., SALA, P., BERTARIO, L., RADICE, P. & CAMA, A. 2013. Integrative Analysis of Hereditary Nonpolyposis Colorectal Cancer: the Contribution of Allele-Specific Expression and Other Assays to Diagnostic Algorithms. *PLoS ONE,* 8**,** e81194.

LEMONIDIS, K., ARKINSON, C., RENNIE, M. L. & WALDEN, H. 2022. Mechanism, specificity, and function of FANCD2-FANCI ubiquitination and deubiquitination. *The FEBS Journal,* 289**,** 4811--4829.

LENZI, M. L., SMITH, J., SNOWDEN, T., KIM, M., FISHEL, R., POULOS, B. K. & COHEN, P. E. 2005. Extreme Heterogeneity in the Molecular Events Leading to the Establishment of Chiasmata during Meiosis I in Human Oocytes. *The American Journal of Human Genetics,* 76**,** 112--127.

LEPPARD, J. B., DONG, Z., MACKEY, Z. B. & TOMKINSON, A. E. 2003. Physical and Functional Interaction between DNA Ligase IIIα and Poly(ADP-Ribose) Polymerase 1 in DNA Single-Strand Break Repair. *Molecular and Cellular Biology,* 23**,** 5919--5927.

LEVINSON, G. & GUTMAN, G. A. 1987. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12. *Nucleic Acids Research,* 15**,** 5323--5338.

LEY, T. J., MARDIS, E. R., DING, L., FULTON, B., MCLELLAN, M. D., CHEN, K., DOOLING, D., DUNFORD-SHORE, B. H., MCGRATH, S., HICKENBOTHAM, M., COOK, L., ABBOTT, R., LARSON, D. E., KOBOLDT, D. C., POHL, C., SMITH, S., HAWKINS, A., ABBOTT, S., LOCKE, D., HILLIER, L. W., MINER, T., FULTON, L., MAGRINI, V., WYLIE, T., GLASSCOCK, J., CONYERS, J., SANDER, N., SHI, X., OSBORNE, J. R., MINX, P., GORDON, D., CHINWALLA, A., ZHAO, Y., RIES, R. E., PAYTON, J. E., WESTERVELT, P., TOMASSON, M. H., WATSON, M., BATY, J., IVANOVICH, J., HEATH, S., SHANNON, W. D., NAGARAJAN, R., WALTER, M. J., LINK, D. C., GRAUBERT, T. A., DIPERSIO, J. F. & WILSON, R. K. 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature,* 456**,** 66--72.

LI, G., ALT, F. W., CHENG, H.-L., BRUSH, J. W., GOFF, P. H., MURPHY, M. M., FRANCO, S., ZHANG, Y. & ZHA, S. 2008. Lymphocyte-Specific Compensation for XLF/Cernunnos End-Joining Functions in V(D)J Recombination. *Molecular Cell,* 31**,** 631--640.

LI, G. M. & MODRICH, P. 1995. Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. *Proceedings of the National Academy of Sciences,* 92**,** 1950--1954.

LI, L., ELLEDGE, S. J., PETERSON, C. A., BALES, E. S. & LEGERSKI, R. J. 1994. Specific association between the human DNA repair proteins XPA and ERCC1. *Proceedings of the National Academy of Sciences,* 91**,** 5012--5016.

LI, L., PETERSON, C. A., LU, X. & LEGERSKI, R. J. 1995a. Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair. *Molecular and Cellular Biology,* 15**,** 1993--1998.

LI, Z., OTEVREL, T., GAO, Y., CHENG, H.-L., SEED, B., STAMATO, T. D., TACCIOLI, G. E. & ALT, F. W. 1995b. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell,* 83**,** 1079--1089.

LICATA, L., VIA, A., TURINA, P., BABBI, G., BENEVENUTA, S., CARTA, C., CASADIO, R., CICCONARDI, A., FACCHIANO, A., FARISELLI, P., GIORDANO, D., ISIDORI, F., MARABOTTI, A., MARTELLI, P. L., PASCARELLA, S., PINELLI, M., PIPPUCCI, T., RUSSO, R., SAVOJARDO, C., SCAFURI, B.,

VALERIANI, L. & CAPRIOTTI, E. 2023. Resources and tools for rare disease variant interpretation. *Frontiers in Molecular Biosciences,* 10**,** 1169109.
LIM, K., LEE, M.-K., DUONG, P. T. M., LIU, D., SUNG, S. & CHOI, B.-S. 2015. Biophysical characterization of the interaction between FAAP20-UBZ4 domain and Rev1-BRCT domain. *FEBS Letters,* 589**,** 3037--3043.
LIN, K. K., HARRELL, M. I., OZA, A. M., OAKNIN, A., RAY-COQUARD, I., TINKER, A. V., HELMAN, E., RADKE, M. R., SAY, C., VO, L.-T., MANN, E., ISAACSON, J. D., MALONEY, L., O'MALLEY, D. M., CHAMBERS, S. K., KAUFMANN, S. H., SCOTT, C. L., KONECNY, G. E., COLEMAN, R. L., SUN, J. X., GIORDANO, H., BRENTON, J. D., HARDING, T. C., MCNEISH, I. A. & SWISHER, E. M. 2018. BRCA Reversion Mutations in Circulating Tumor DNA Predict Primary and Acquired Resistance to the PARP Inhibitor Rucaparib in High-Grade Ovarian Carcinoma. *Cancer Discovery,* 9**,** CD--18-0715.
LIN, Y.-L., SHIVJI, M. K. K., CHEN, C., KOLODNER, R., WOOD, R. D. & DUTTA, A. 1998. The Evolutionarily Conserved Zinc Finger Motif in the Largest Subunit of Human Replication Protein A Is Required for DNA Replication and Mismatch Repair but Not for Nucleotide Excision Repair*. *Journal of Biological Chemistry,* 273**,** 1453--1461.
LIN, Z., NEI, M. & MA, H. 2007. The origins and early evolution of DNA mismatch repair genes—multiple horizontal gene transfers and co-evolution. *Nucleic Acids Research,* 35**,** 7591--7603.
LIPKIN, S. M., MOENS, P. B., WANG, V., LENZI, M., SHANMUGARAJAH, D., GILGEOUS, A., THOMAS, J., CHENG, J., TOUCHMAN, J. W., GREEN, E. D., SCHWARTZBERG, P., COLLINS, F. S. & COHEN, P. E. 2002. Meiotic arrest and aneuploidy in MLH3-deficient mice. *Nature Genetics,* 31**,** 385--390.
LIU, F., STANTON, J. J., WU, Z. & PIWNICA-WORMS, H. 1997. The human Myt1 kinase preferentially phosphorylates Cdc2 on threonine 14 and localizes to the endoplasmic reticulum and Golgi complex. *Molecular and Cellular Biology,* 17**,** 571--583.
LIU, H., RUDOLF, J., JOHNSON, K. A., MCMAHON, S. A., OKE, M., CARTER, L., MCROBBIE, A.-M., BROWN, S. E., NAISMITH, J. H. & WHITE, M. F. 2008. Structure of the DNA Repair Helicase XPD. *Cell,* 133**,** 801--812.
LIU, J., DOTY, T., GIBSON, B. & HEYER, W.-D. 2010a. 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. & FISHEL, R. 2016. Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. *Nature,* 539**,** 583--587.
LIU, Q., GUNTUKU, S., CUI, X.-S., MATSUOKA, S., CORTEZ, D., TAMAI, K., LUO, G., CARATTINI-RIVERA, S., DEMAYO, F., BRADLEY, A., DONEHOWER, L. A. & ELLEDGE, S. J. 2000. Chk1 is an essential kinase that is regulated by Atr and required for the G2/M DNA damage checkpoint. *Genes \& Development,* 14**,** 1448--1459.
LIU, S. & KONG, D. 2021. End resection: a key step in homologous recombination and DNA double-strand break repair. *Genome Instability \& Disease,* 2**,** 39--50.
LIU, T., GHOSAL, G., YUAN, J., CHEN, J. & HUANG, J. 2010b. FAN1 Acts with FANCI-FANCD2 to Promote DNA Interstrand Cross-Link Repair. *Science,* 329**,** 693--696.
LIU, T., YAN, H., KUISMANEN, S., PERCESEPE, A., BISGAARD, M. L., PEDRONI, M., BENATTI, P., KINZLER, K. W., VOGELSTEIN, B., LEON, M. P. D., PELTOMÄKI, P. & LINDBLOM, A. 2001. The role of hPMS1 and hPMS2 in predisposing to colorectal cancer. *Cancer research,* 61**,** 7798--802.
LIVESEY, B. J. & MARSH, J. A. 2022. Interpreting protein variant effects with computational predictors and deep mutational scanning. *Disease Models \& Mechanisms,* 15**,** dmm049510.
LOEB, L. A. 2016. Human Cancers Express a Mutator Phenotype: Hypothesis, Origin, and Consequences. *Cancer Research,* 76**,** 2057--2059.

LOEB, L. A. & MONNAT, R. J. 2008. DNA polymerases and human disease. *Nature Reviews Genetics,* 9**,** 594--604.
LOEB, L. A., SPRINGGATE, C. F. & BATTULA, N. 1974. Errors in DNA replication as a basis of malignant changes. *Cancer research,* 34**,** 2311--21.
LOGSDON, G. A., VOLLGER, M. R. & EICHLER, E. E. 2020. Long-read human genome sequencing and its applications. *Nature Reviews Genetics,* 21**,** 597--614.
LONG, X.-D., ZHAO, D., WANG, C., HUANG, X.-Y., YAO, J.-G., MA, Y., WEI, Z.-H., LIU, M., ZENG, L.-X., MO, X.-Q., ZHANG, J.-J., XUE, F., ZHAI, B. & XIA, Q. 2013. Genetic Polymorphisms in DNA Repair Genes XRCC4 and XRCC5 and Aflatoxin B1–related Hepatocellular Carcinoma. *Epidemiology,* 24**,** 671--681.
LONGERICH, S., FILIPPO, J. S., LIU, D. & SUNG, P. 2009. FANCI Binds Branched DNA and Is Monoubiquitinated by UBE2T-FANCL*. *Journal of Biological Chemistry,* 284**,** 23182--23186.
LONGLEY, D. B., HARKIN, D. P. & JOHNSTON, P. G. 2003. 5-Fluorouracil: mechanisms of action and clinical strategies. *Nature Reviews Cancer,* 3**,** 330--338.
LONGLEY, M. J., PIERCE, A. J. & MODRICH, P. 1997. DNA Polymerase δ Is Required for Human Mismatch Repair in Vitro *. *Journal of Biological Chemistry,* 272**,** 10917--10921.
LOOIJ, M. V. D., PAPP, J., SZTÁN, M., PULAY, T., ELFADIL, I., BESZNYAK, I., TÓTH, J., DEVILEE, P. & OLÁH, E. 2001. Allelic imbalance and microsatellite instability in BRCA1 associated breast and ovarian tumors. *International journal of oncology,* 18**,** 775--80.
LORD, J. & BARALLE, D. 2021. Splicing in the Diagnosis of Rare Disease: Advances and Challenges. *Frontiers in Genetics,* 12**,** 689892.
LU, A. L., CLARK, S. & MODRICH, P. 1983. Methyl-directed repair of DNA base-pair mismatches in vitro. *Proceedings of the National Academy of Sciences,* 80**,** 4639--4643.
LU, H., ZHOU, Q., HE, J., JIANG, Z., PENG, C., TONG, R. & SHI, J. 2020. Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Signal Transduction and Targeted Therapy,* 5**,** 213.
LUENGO, A., GUI, D. Y. & HEIDEN, M. G. V. 2017. Targeting Metabolism for Cancer Therapy. *Cell Chemical Biology,* 24**,** 1161--1180.
LUMPKIN, R. J., GU, H., ZHU, Y., LEONARD, M., AHMAD, A. S., CLAUSER, K. R., MEYER, J. G., BENNETT, E. J. & KOMIVES, E. A. 2017. Site-specific identification and quantitation of endogenous SUMO modifications under native conditions. *Nature Communications,* 8**,** 1171.
LYNCH, H. T., SHAW, M. W., MAGNUSON, C. W., LARSEN, A. L. & KRUSH, A. J. 1966. Hereditary Factors in Cancer: Study of Two Large Midwestern Kindreds. *Archives of Internal Medicine,* 117**,** 206--212.
LYNCH, H. T., SNYDER, C. L., SHAW, T. G., HEINEN, C. D. & HITCHINS, M. P. 2015. Milestones of Lynch syndrome: 1895–2015. *Nature Reviews Cancer,* 15**,** 181--194.
MA, S. L., VEGA-WARNER, V., GILLIES, C., SAMPSON, M. G., KHER, V., SETHI, S. K. & OTTO, E. A. 2015. Whole Exome Sequencing Reveals Novel PHEX Splice Site Mutations in Patients with Hypophosphatemic Rickets. *PLoS ONE,* 10**,** e0130729.
MA, Y., PANNICKE, U., SCHWARZ, K. & 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.
MACHERET, M. & HALAZONETIS, T. D. 2015. DNA Replication Stress as a Hallmark of Cancer. *Annual Review of Pathology: Mechanisms of Disease,* 10**,** 425--448.
MACHIDA, Y. J., MACHIDA, Y., CHEN, Y., GURTAN, A. M., KUPFER, G. M., D'ANDREA, A. D. & DUTTA, A. 2006. UBE2T Is the E2 in the Fanconi Anemia Pathway and Undergoes Negative Autoregulation. *Molecular Cell,* 23**,** 589--596.
MACKAY, C., DÉCLAIS, A.-C., LUNDIN, C., AGOSTINHO, A., DEANS, A. J., MACARTNEY, T. J., HOFMANN, K., GARTNER, A., WEST, S. C., HELLEDAY, T., LILLEY, D. M. J. & ROUSE, J. 2010. Identification of KIAA1018/FAN1, a DNA Repair Nuclease Recruited to DNA Damage by Monoubiquitinated FANCD2. *Cell,* 142**,** 65--76.

MAHMOOD, T. & YANG, P.-C. 2012. Western Blot: Technique, Theory, and Trouble Shooting. *North American Journal of Medical Sciences,* 4**,** 429--434.

MAILAND, N., BEKKER-JENSEN, S., FAUSTRUP, H., MELANDER, F., BARTEK, J., LUKAS, C. & LUKAS, J. 2007. RNF8 Ubiquitylates Histones at DNA Double-Strand Breaks and Promotes Assembly of Repair Proteins. *Cell,* 131**,** 887--900.

MAILAND, N., FALCK, J., LUKAS, C., SYLJUÅSEN, R. G., WELCKER, M., BARTEK, J. & LUKAS, J. 2000. Rapid Destruction of Human Cdc25A in Response to DNA Damage. *Science,* 288**,** 1425--1429.

MALIVERT, L., ROPARS, V., NUNEZ, M., DREVET, P., MIRON, S., FAURE, G., GUEROIS, R., MORNON, J.-P., REVY, P., CHARBONNIER, J.-B., CALLEBAUT, I. & VILLARTAY, J.-P. D. 2010. Delineation of the Xrcc4-interacting Region in the Globular Head Domain of Cernunnos/XLF*. *Journal of Biological Chemistry,* 285**,** 26475--26483.

MANHART, C. M. & ALANI, E. 2016. Roles for mismatch repair family proteins in promoting meiotic crossing over. *DNA Repair,* 38**,** 84--93.

MANOLIO, T. A., FOWLER, D. M., STARITA, L. M., HAENDEL, M. A., MACARTHUR, D. G., BIESECKER, L. G., WORTHEY, E., CHISHOLM, R. L., GREEN, E. D., JACOB, H. J., MCLEOD, H. L., RODEN, D., RODRIGUEZ, L. L., WILLIAMS, M. S., COOPER, G. M., COX, N. J., HERMAN, G. E., KINGSMORE, S., LO, C., LUTZ, C., MACRAE, C. A., NUSSBAUM, R. L., ORDOVAS, J. M., RAMOS, E. M., ROBINSON, P. N., RUBINSTEIN, W. S., SEIDMAN, C., STRANGER, B. E., WANG, H., WESTERFIELD, M. & BULT, C. 2017. Bedside Back to Bench: Building Bridges between Basic and Clinical Genomic Research. *Cell,* 169**,** 6--12.

MANTHEI, K. A. & KECK, J. L. 2013. The BLM dissolvasome in DNA replication and repair. *Cellular and Molecular Life Sciences,* 70**,** 4067--4084.

MAREK, L. R. & BALE, A. E. 2006. Drosophila homologs of FANCD2 and FANCL function in DNA repair. *DNA Repair,* 5**,** 1317--1326.

MARSHALL, C. R., CHOWDHURY, S., TAFT, R. J., LEBO, M. S., BUCHAN, J. G., HARRISON, S. M., ROWSEY, R., KLEE, E. W., LIU, P., WORTHEY, E. A., JOBANPUTRA, V., DIMMOCK, D., KEARNEY, H. M., BICK, D., KULKARNI, S., TAYLOR, S. L., BELMONT, J. W., STAVROPOULOS, D. J., LENNON, N. J. & INITIATIVE, M. G. 2020. Best practices for the analytical validation of clinical whole-genome sequencing intended for the diagnosis of germline disease. *npj Genomic Medicine,* 5**,** 47.

MARSTON, A. L. & AMON, A. 2004. Meiosis: cell-cycle controls shuffle and deal. *Nature Reviews Molecular Cell Biology,* 5**,** 983--997.

MARTEIJN, J. A., LANS, H., VERMEULEN, W. & HOEIJMAKERS, J. H. J. 2014. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature Reviews Molecular Cell Biology,* 15**,** 465--481.

MARTIN, H. C., KIM, G. E., PAGNAMENTA, A. T., MURAKAMI, Y., CARVILL, G. L., MEYER, E., COPLEY, R. R., RIMMER, A., BARCIA, G., FLEMING, M. R., KRONENGOLD, J., BROWN, M. R., HUDSPITH, K. A., BROXHOLME, J., KANAPIN, A., CAZIER, J. B., KINOSHITA, T., NABBOUT, R., CONSORTIUM, T. W. G. S., BENTLEY, D., MCVEAN, G., HEAVIN, S., ZAIWALLA, Z., MCSHANE, T., MEFFORD, H. C., SHEARS, D., STEWART, H., KURIAN, M. A., SCHEFFER, I. E., BLAIR, E., DONNELLY, P., KACZMAREK, L. K. & TAYLOR, J. C. 2014. Clinical whole-genome sequencing in severe early-onset epilepsy reveals new genes and improves molecular diagnosis. *Human Molecular Genetics,* 23**,** 3200--3211.

MASUTANI, C., SUGASAWA, K., YANAGISAWA, J., SONOYAMA, T., UI, M., ENOMOTO, T., TAKIO, K., TANAKA, K., SPEK, P. J. & BOOTSMA, D. 1994. Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *The EMBO Journal,* 13**,** 1831--1843.

MATEOS-GOMEZ, P. A., GONG, F., NAIR, N., MILLER, K. M., LAZZERINI-DENCHI, E. & SFEIR, A. 2015. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature,* 518**,** 254--257.

MATHIEU, N., KACZMAREK, N., RÜTHEMANN, P., LUCH, A. & NAEGELI, H. 2013. DNA Quality Control by a Lesion Sensor Pocket of the Xeroderma Pigmentosum Group D Helicase Subunit of TFIIH. *Current Biology,* 23**,** 204--212.

MATREYEK, K. A., STARITA, L. M., STEPHANY, J. J., MARTIN, B., CHIASSON, M. A., GRAY, V. E., KIRCHER, M., KHECHADURI, A., DINES, J. N., HAUSE, R. J., BHATIA, S., EVANS, W. E., RELLING, M. V., YANG, W., SHENDURE, J. & FOWLER, D. M. 2018. Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nature Genetics,* 50**,** 874--882.

MATSUDA, T., SAIJO, M., KURAOKA, I., KOBAYASHI, T., NAKATSU, Y., NAGAI, A., ENJOJI, T., MASUTANI, C., SUGASAWA, K., HANAOKA, F., YASUI, A. & TANAKA, K. 1995. DNA Repair Protein XPA Binds Replication Protein A (RPA) (∗). *Journal of Biological Chemistry,* 270**,** 4152--4157.

MATSUMOTO, Y., SUZUKI, N., NAMBA, N., UMEDA, N., MA, X.-J., MORITA, A., TOMITA, M., ENOMOTO, A., SERIZAWA, S., HIRANO, K., SAKAI, K., YASUDA, H. & HOSOI, Y. 2000. Cleavage and phosphorylation of XRCC4 protein induced by X-irradiation. *FEBS Letters,* 478**,** 67--71.

MATSUOKA, S., BALLIF, B. A., SMOGORZEWSKA, A., III, E. R. M., HUROV, K. E., LUO, J., BAKALARSKI, C. E., ZHAO, Z., SOLIMINI, N., LERENTHAL, Y., SHILOH, Y., GYGI, S. P. & ELLEDGE, S. J. 2007. ATM and ATR Substrate Analysis Reveals Extensive Protein Networks Responsive to DNA Damage. *Science,* 316**,** 1160--1166.

MATSUOKA, S., HUANG, M. & ELLEDGE, S. J. 1998. Linkage of ATM to Cell Cycle Regulation by the Chk2 Protein Kinase. *Science,* 282**,** 1893--1897.

MATSUOKA, S., ROTMAN, G., OGAWA, A., SHILOH, Y., TAMAI, K. & ELLEDGE, S. J. 2000. Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. *Proceedings of the National Academy of Sciences,* 97**,** 10389--10394.

MATTHEWS, H. K., BERTOLI, C. & BRUIN, R. A. M. D. 2022. Cell cycle control in cancer. *Nature Reviews Molecular Cell Biology,* 23**,** 74--88.

MATTIROLI, F., VISSERS, JOSEPH H. A., VAN DIJK, WILLEM J., IKPA, P., CITTERIO, E., VERMEULEN, W., MARTEIJN, JURGEN A. & SIXMA, TITIA K. 2012. RNF168 Ubiquitinates K13-15 on H2A/H2AX to Drive DNA Damage Signaling. *Cell,* 150**,** 1182--1195.

MAVRAGANI, I. V., NIKITAKI, Z., KALOSPYROS, S. A. & GEORGAKILAS, A. G. 2019. Ionizing Radiation and Complex DNA Damage: From Prediction to Detection Challenges and Biological Significance. *Cancers,* 11**,** 1789.

MAXAM, A. M. & GILBERT, W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences,* 74**,** 560--564.

MAZUR, D. J., MENDILLO, M. L. & KOLODNER, R. D. 2006. Inhibition of Msh6 ATPase Activity by Mispaired DNA Induces a Msh2(ATP)-Msh6(ATP) State Capable of Hydrolysis-Independent Movement along DNA. *Molecular Cell,* 22**,** 39--49.

MCHUGH, P. J. 2020. XPF–ERCC1: Linchpin of DNA crosslink repair. *PLoS Genetics,* 16**,** e1008616.

MCKENZIE, R., SMALE, D., BODEKER, G. & CLAUDE, H. 2003. Ozone profile differences between Europe and New Zealand: Effects on surface UV irradiance and its estimation from satellite sensors. *Journal of Geophysical Research: Atmospheres,* 108.

MCKUSICK, V. A. 2007. Mendelian Inheritance in Man and Its Online Version, OMIM. *The American Journal of Human Genetics,* 80**,** 588--604.

MCVEY, M. 2010. Strategies for DNA interstrand crosslink repair: insights from worms, flies, frogs, and slime molds. *Environmental and molecular mutagenesis,* 51**,** 646--58.

MEETEI, A. R., MEDHURST, A. L., LING, C., XUE, Y., SINGH, T. R., BIER, P., STELTENPOOL, J., STONE, S., DOKAL, I., MATHEW, C. G., HOATLIN, M., JOENJE, H., WINTER, J. P. D. & WANG, W. 2005. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nature Genetics,* 37**,** 958--963.

MEETEI, A. R., WINTER, J. P. D., MEDHURST, A. L., WALLISCH, M., WAISFISZ, Q., VRUGT, H. J. V. D., OOSTRA, A. B., YAN, Z., LING, C., BISHOP, C. E., HOATLIN, M. E., JOENJE, H. & WANG, W. 2003. A novel ubiquitin ligase is deficient in Fanconi anemia. *Nature Genetics,* 35**,** 165--170.

MEETEI, A. R., YAN, Z. & WANG, W. 2004. FANCL Replaces BRCA1 as the Likely Ubiquitin Ligase Responsible for FANCD2 Monoubiquitination. *Cell Cycle,* 3**,** 174--176.
MEI, L., ZHANG, J., HE, K. & ZHANG, J. 2019. Ataxia telangiectasia and Rad3-related inhibitors and cancer therapy: where we stand. *Journal of Hematology \& Oncology,* 12**,** 43.
MEIER, B., BARBER, L. J., LIU, Y., SHTESSEL, L., BOULTON, S. J., GARTNER, A. & AHMED, S. 2009. The MRT-1 nuclease is required for DNA crosslink repair and telomerase activity in vivo in Caenorhabditis elegans. *The EMBO Journal,* 28**,** 3549--3563.
MELANDER, F., BEKKER-JENSEN, S., FALCK, J., BARTEK, J., MAILAND, N. & 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--26.
MELLON, I., HOCK, T., REID, R., PORTER, P. C. & STATES, J. C. 2002. Polymorphisms in the human xeroderma pigmentosum group A gene and their impact on cell survival and nucleotide excision repair. *DNA Repair,* 1**,** 531--546.
MENDEZ-DORANTES, C., BHARGAVA, R. & 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.
MENDILLO, M. L., HARGREAVES, V. V., JAMISON, J. W., MO, A. O., LI, S., PUTNAM, C. D., WOODS, V. L. & KOLODNER, R. D. 2009. A conserved MutS homolog connector domain interface interacts with MutL homologs. *Proceedings of the National Academy of Sciences,* 106**,** 22223--22228.
MENG, F., LIANG, Z., ZHAO, K. & LUO, C. 2021. Drug design targeting active posttranslational modification protein isoforms. *Medicinal Research Reviews,* 41**,** 1701--1750.
MENGWASSER, K. E., ADEYEMI, R. O., LENG, Y., CHOI, M. Y., CLAIRMONT, C., D'ANDREA, A. D. & ELLEDGE, S. J. 2019. Genetic Screens Reveal FEN1 and APEX2 as BRCA2 Synthetic Lethal Targets. *Molecular Cell,* 73**,** 885--899.e6.
MER, G., BOCHKAREV, A., GUPTA, R., BOCHKAREVA, E., FRAPPIER, L., INGLES, C. J., EDWARDS, A. M. & CHAZIN, W. J. 2000. Structural Basis for the Recognition of DNA Repair Proteins UNG2, XPA, and RAD52 by Replication Factor RPA. *Cell,* 103**,** 449--456.
MERTINS, P., QIAO, J. W., PATEL, J., UDESHI, N. D., CLAUSER, K. R., MANI, D. R., BURGESS, M. W., GILLETTE, M. A., JAFFE, J. D. & CARR, S. A. 2013. Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nature Methods,* 10**,** 634--637.
MESMAN, R. L. S., CALLÉJA, F. M. G. R., HENDRIKS, G., MOROLLI, B., MISOVIC, B., DEVILEE, P., ASPEREN, C. J. V., VRIELING, H. & VREESWIJK, M. P. G. 2019. The functional impact of variants of uncertain significance in BRCA2. *Genetics in Medicine,* 21**,** 293--302.
MESTER, J. L., GHOSH, R., PESARAN, T., HUETHER, R., KARAM, R., HRUSKA, K. S., COSTA, H. A., LACHLAN, K., NGEOW, J., BARNHOLTZ-SLOAN, J., SESOCK, K., HERNANDEZ, F., ZHANG, L., MILKO, L., PLON, S. E., HEGDE, M. & ENG, C. 2018. Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. *Human Mutation,* 39**,** 1581--1592.
MEYERS, M., WAGNER, M. W., MAZUREK, A., SCHMUTTE, C., FISHEL, R. & BOOTHMAN, D. A. 2005. DNA Mismatch Repair-dependent Response to Fluoropyrimidine-generated Damage*. *Journal of Biological Chemistry,* 280**,** 5516--5526.
MILES, J. A., FROST, M. G., CARROLL, E., ROWE, M. L., HOWARD, M. J., SIDHU, A., CHAUGULE, V. K., ALPI, A. F. & WALDEN, H. 2015. The Fanconi Anemia DNA Repair Pathway Is Regulated by an Interaction between Ubiquitin and the E2-like Fold Domain of FANCL. *The Journal of biological chemistry,* 290**,** 20995--1006.
MIMITOU, E. P. & SYMINGTON, L. S. 2008. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature,* 455**,** 770--774.
MIMORI, T. & HARDIN, J. A. 1986. Mechanism of interaction between Ku protein and DNA. *The Journal of biological chemistry,* 261**,** 10375--9.

MIRMAN, Z., SASI, N. K., KING, A., CHAPMAN, J. R. & LANGE, T. D. 2022. 53BP1–shieldin-dependent DSB processing in BRCA1-deficient cells requires CST–Polα–primase fill-in synthesis. *Nature Cell Biology,* 24**,** 51--61.

MITTAL, R. D., GANGWAR, R., MANDAL, R. K., SRIVASTAVA, P. & AHIRWAR, D. K. 2012. Gene variants of XRCC4 and XRCC3 and their association with risk for urothelial bladder cancer. *Molecular Biology Reports,* 39**,** 1667--1675.

MIYAKI, M., KONISHI, M., TANAKA, K., KIKUCHI-YANOSHITA, R., MURAOKA, M., YASUNO, M., IGARI, T., KOIKE, M., CHIBA, M. & MORI, T. 1997. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nature Genetics,* 17**,** 271--272.

MIYAMOTO, I., MIURA, N., NIWA, H., MIYAZAKI, J. & TANAKA, K. 1992. Mutational analysis of the structure and function of the xeroderma pigmentosum group A complementing protein. Identification of essential domains for nuclear localization and DNA excision repair. *The Journal of biological chemistry,* 267**,** 12182--7.

MIZUTA, R., CHENG, H. L., GAO, Y. & ALT, F. W. 1997. Molecular genetic characterization of XRCC4 function. *International Immunology,* 9**,** 1607--1613.

MODESTI, M., HESSE, J. E. & GELLERT, M. 1999. DNA binding of Xrcc4 protein is associated with V(D)J recombination but not with stimulation of DNA ligase IV activity. *The EMBO Journal,* 18**,** 2008--2018.

MODESTI, M., JUNOP, M. S., GHIRLANDO, R., RAKT, M. V. D., GELLERT, M., YANG, W. & KANAAR, R. 2003. Tetramerization and DNA Ligase IV Interaction of the DNA Double-strand Break Repair Protein XRCC4 are Mutually Exclusive. *Journal of Molecular Biology,* 334**,** 215--228.

MODRICH, P. 1987. DNA Mismatch Correction. *Annual Review of Biochemistry,* 56**,** 435--466.

MODRICH, P. 2016. Mechanisms in E. coli and Human Mismatch Repair (Nobel Lecture). *Angewandte Chemie International Edition,* 55**,** 8490--8501.

MOENS, P. B., KOLAS, N. K., TARSOUNAS, M., MARCON, E., COHEN, P. E. & SPYROPOULOS, B. 2002. The time course and chromosomal localization of recombination-related proteins at meiosis in the mouse are compatible with models that can resolve the early DNA-DNA interactions without reciprocal recombination. *Journal of Cell Science,* 115**,** 1611--1622.

MOGHANI, A. R. A., SHARMA, M. K. & MATSUMOTO, Y. 2018. In cellulo phosphorylation of DNA double-strand break repair protein XRCC4 on Ser260 by DNA-PK. *Journal of Radiation Research,* 59**,** rry072.

MOHAMMADI, S., SABERIDOKHT, B., SUBRAMANIAM, S. & GRAMA, A. 2015. Scope and limitations of yeast as a model organism for studying human tissue-specific pathways. *BMC Systems Biology,* 9**,** 96.

MOHD, A. B., PALAMA, B., NELSON, S. E., TOMER, G., NGUYEN, M., HUO, X. & BUERMEYER, A. B. 2006. Truncation of the C-terminus of human MLH1 blocks intracellular stabilization of PMS2 and disrupts DNA mismatch repair. *DNA Repair,* 5**,** 347--361.

MOL, C. D., IZUMI, T., MITRA, S. & TAINER, J. A. 2000. DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination. *Nature,* 403**,** 451--456.

MØLLER, P., SEPPÄLÄ, T. T., BERNSTEIN, I., HOLINSKI-FEDER, E., SALA, P., EVANS, D. G., LINDBLOM, A., MACRAE, F., BLANCO, I., SIJMONS, R. H., JEFFRIES, J., VASEN, H. F. A., BURN, J., NAKKEN, S., HOVIG, E., RØDLAND, E. A., THARMARATNAM, K., CAPPEL, W. H. D. V. T. N., HILL, J., WIJNEN, J. T., JENKINS, M. A., GREEN, K., LALLOO, F., SUNDE, L., MINTS, M., BERTARIO, L., PINEDA, M., NAVARRO, M., MORAK, M., RENKONEN-SINISALO, L., VALENTIN, M. D., FRAYLING, I. M., PLAZZER, J.-P., PYLVANAINEN, K., GENUARDI, M., MECKLIN, J.-P., MOESLEIN, G., SAMPSON, J. R., CAPELLA, G. & GROUP, M. 2018. Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database. *Gut,* 67**,** 1306.

MONTECUCCO, A., ZANETTA, F. & BIAMONTI, G. 2015. Molecular mechanisms of etoposide. *EXCLI Journal,* 14**,** 95--108.

MOORE, S., STANLEY, F. K. T. & GOODARZI, A. A. 2014. The repair of environmentally relevant DNA double strand breaks caused by high linear energy transfer irradiation – No simple task. *DNA Repair,* 17**,** 64--73.

MORENO, O. M., PAREDES, A. C., SUAREZ-OBANDO, F. & ROJAS, A. 2021. An update on Fanconi anemia: Clinical, cytogenetic and molecular approaches (Review). *Biomedical Reports,* 15**,** 74.

MORITA, E. H., OHKUBO, T., KURAOKA, I., SHIRAKAWA, M., TANAKA, K. & MORIKAWA, K. 1996. Implications of the zinc-finger motif found in the DNA-binding domain of the human XPA protein. *Genes to Cells,* 1**,** 437--442.

MORRIS, J. R., PANGON, L., BOUTELL, C., KATAGIRI, T., KEEP, N. H. & SOLOMON, E. 2006. Genetic analysis of BRCA1 ubiquitin ligase activity and its relationship to breast cancer susceptibility. *Human Molecular Genetics,* 15**,** 599--606.

MOTYCKA, T. A., BESSHO, T., POST, S. M., SUNG, P. & TOMKINSON, A. E. 2004. Physical and Functional Interaction between the XPF/ERCC1 Endonuclease and hRad52*. *Journal of Biological Chemistry,* 279**,** 13634--13639.

MUÑOZ, I. M., HAIN, K., DÉCLAIS, A.-C., GARDINER, M., TOH, G. W., SANCHEZ-PULIDO, L., HEUCKMANN, J. M., TOTH, R., MACARTNEY, T., EPPINK, B., KANAAR, R., PONTING, C. P., LILLEY, D. M. J. & ROUSE, J. 2009. Coordination of Structure-Specific Nucleases by Human SLX4/BTBD12 Is Required for DNA Repair. *Molecular Cell,* 35**,** 116--127.

MUNSHI, P. N., LUBIN, M. & BERTINO, J. R. 2014. 6-Thioguanine: A Drug With Unrealized Potential for Cancer Therapy. *The Oncologist,* 19**,** 760--765.

MURAI, J., HUANG, S.-Y. N., DAS, B. B., RENAUD, A., ZHANG, Y., DOROSHOW, J. H., JI, J., TAKEDA, S. & POMMIER, Y. 2012. Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. *Cancer Research,* 72**,** 5588--5599.

MURRAY, JENNIE E., VAN DER BURG, M., IJSPEERT, H., CARROLL, P., WU, Q., OCHI, T., LEITCH, A., MILLER, EDWARD S., KYSELA, B., JAWAD, A., BOTTANI, A., BRANCATI, F., CAPPA, M., CORMIER-DAIRE, V., DESHPANDE, C., FAQEIH, EISSA A., GRAHAM, GAIL E., RANZA, E., BLUNDELL, TOM L., JACKSON, ANDREW P., STEWART, GRANT S. & BICKNELL, LOUISE S. 2015. Mutations in the NHEJ Component XRCC4 Cause Primordial Dwarfism. *The American Journal of Human Genetics,* 96**,** 412--424.

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., CHAPMAN, J. R., BARTKE, T. & GROTH, A. 2019. H4K20me0 recognition by BRCA1–BARD1 directs homologous recombination to sister chromatids. *Nature Cell Biology,* 21**,** 311--318.

NAKANE, H., TAKEUCHI, S., YUBA, S., SAIJO, M., NAKATSU, Y., MURAI, H., NAKATSURU, Y., ISHIKAWA, T., HIROTA, S., KITAMURA, Y., KATO, Y., TSUNODA, Y., MIYAUCHI, H., HORIO, T., TOKUNAGA, T., MATSUNAGA, T., NIKAIDO, O., NISHIMUNE, Y., OKADA, Y. & TANAKA, K. 1995. High incidence of ultraviolet-B-or chemical-carcinogen-induced skin tumours in mice lacking the xeroderma pigmentosum group A gene. *Nature,* 377**,** 165--168.

NANGALIA, J. & CAMPBELL, P. J. 2019. Genome Sequencing during a Patient's Journey through Cancer. *New England Journal of Medicine,* 381**,** 2145--2156.

NASSIF, N., PENNEY, J., PAL, S., ENGELS, W. R. & 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.

NATARAJAN, N., WATSON, P., SILVA-LOPEZ, E. & LYNCH, H. T. 2010. Comparison of Extended Colectomy and Limited Resection in Patients With Lynch Syndrome. *Diseases of the Colon \& Rectum,* 53**,** 77--82.

NEVELING, K., ENDT, D., HOEHN, H. & SCHINDLER, D. 2009. Genotype–phenotype correlations in Fanconi anemia. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 668**,** 73--91.

NEVERS, P. & SPATZ, H.-C. 1975. Escherichia coli mutants uvr D and uvr E deficient in gene conversion of λ-heteroduplexes. *Molecular and General Genetics MGG,* 139**,** 233--243.
NG, P. C. & HENIKOFF, S. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research,* 31**,** 3812--3814.
NGHIEM, P., PARK, P. K., KIM, Y.-S., VAZIRI, C. & SCHREIBER, S. L. 2001. ATR inhibition selectively sensitizes G1 checkpoint-deficient cells to lethal premature chromatin condensation. *Proceedings of the National Academy of Sciences,* 98**,** 9092--9097.
NGUYEN, H. H., PARK, J., KANG, S. & KIM, M. 2015. Surface Plasmon Resonance: A Versatile Technique for Biosensor Applications. *Sensors,* 15**,** 10481--10510.
NGUYEN, T.-A., SLATTERY, S. D., MOON, S.-H., DARLINGTON, Y. F., LU, X. & DONEHOWER, L. A. 2010. The oncogenic phosphatase WIP1 negatively regulates nucleotide excision repair. *DNA Repair,* 9**,** 813--823.
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., ADAMS, M. D., VENTER, J. C., DUNLOP, M. G., HAMILTON, S. R., PETERSEN, G. M., CHAPELLE, A. D. L., VOGELSTEIN, B. & KINZLER, K. W. 1994. Mutations of two P/WS homologues in hereditary nonpolyposis colon cancer. *Nature,* 371**,** 75--80.
NIEDERNHOFER, L. J., ODIJK, H., BUDZOWSKA, M., DRUNEN, E. V., MAAS, A., THEIL, A. F., WIT, J. D., JASPERS, N. G. J., BEVERLOO, H. B., HOEIJMAKERS, J. H. J. & KANAAR, R. 2004. The Structure-Specific Endonuclease Ercc1-Xpf Is Required To Resolve DNA Interstrand Cross-Link-Induced Double-Strand Breaks. *Molecular and Cellular Biology,* 24**,** 5776--5787.
NIELSEN, F. C., JÄGER, A. C., LÜTZEN, A., BUNDGAARD, J. R. & RASMUSSEN, L. J. 2004. Characterization of human exonuclease 1 in complex with mismatch repair proteins, subcellular localization and association with PCNA. *Oncogene,* 23**,** 1457--1468.
NIELSEN, S. V., STEIN, A., DINITZEN, A. B., PAPALEO, E., TATHAM, M. H., POULSEN, E. G., KASSEM, M. M., RASMUSSEN, L. J., LINDORFF-LARSEN, K. & HARTMANN-PETERSEN, R. 2017. Predicting the impact of Lynch syndrome-causing missense mutations from structural calculations. *PLoS Genetics,* 13**,** e1006739.
NIESSEN, R. C., BERENDS, M. J. W., WU, Y., SIJMONS, R. H., HOLLEMA, H., LIGTENBERG, M. J. L., WALLE, H. E. K. D., VRIES, E. G. E. D., KARRENBELD, A., BUYS, C. H. C. M., ZEE, A. G. J. V. D., HOFSTRA, R. M. W. & KLEIBEUKER, J. H. 2006. Identification of mismatch repair gene mutations in young patients with colorectal cancer and in patients with multiple tumours associated with hereditary non-polyposis colorectal cancer. *Gut,* 55**,** 1781.
NIJMAN, S. M. B., HUANG, T. T., DIRAC, A. M. G., BRUMMELKAMP, T. R., KERKHOVEN, R. M., D'ANDREA, A. D. & BERNARDS, R. 2005. The Deubiquitinating Enzyme USP1 Regulates the Fanconi Anemia Pathway. *Molecular Cell,* 17**,** 331--339.
NIMONKAR, A. V. & 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.
NIRAJ, J., FÄRKKILÄ, A. & D'ANDREA, A. D. 2019. The Fanconi Anemia Pathway in Cancer. *Annual Review of Cancer Biology,* 3**,** 457--478.
NIROULA, A. & VIHINEN, M. 2016. Variation Interpretation Predictors: Principles, Types, Performance, and Choice. *Human Mutation,* 37**,** 579--597.
NISHANT, K. T., PLYS, A. J. & ALANI, E. 2008. A Mutation in the Putative MLH3 Endonuclease Domain Confers a Defect in Both Mismatch Repair and Meiosis in Saccharomyces cerevisiae. *Genetics,* 179**,** 747--755.
NISHI, R., OKUDA, Y., WATANABE, E., MORI, T., IWAI, S., MASUTANI, C., SUGASAWA, K. & HANAOKA, F. 2005. Centrin 2 Stimulates Nucleotide Excision Repair by Interacting with Xeroderma Pigmentosum Group C Protein. *Molecular and Cellular Biology,* 25**,** 5664--5674.
NOCENTINI, S., COIN, F., SAIJO, M., TANAKA, K. & EGLY, J.-M. 1997. DNA Damage Recognition by XPA Protein Promotes Efficient Recruitment of Transcription Factor II H*. *Journal of Biological Chemistry,* 272**,** 22991--22994.

NOORDERMEER, S. M., ADAM, S., SETIAPUTRA, D., BARAZAS, M., PETTITT, S. J., LING, A. K., OLIVIERI, M., ÁLVAREZ-QUILÓN, A., MOATTI, N., ZIMMERMANN, M., ANNUNZIATO, S., KRASTEV, D. B., SONG, F., BRANDSMA, I., FRANKUM, J., BROUGH, R., SHERKER, A., LANDRY, S., SZILARD, R. K., MUNRO, M. M., MCEWAN, A., RUGY, T. G. D., LIN, Z.-Y., HART, T., MOFFAT, J., GINGRAS, A.-C., MARTIN, A., ATTIKUM, H. V., JONKERS, J., LORD, C. J., ROTTENBERG, S. & DUROCHER, D. 2018. The shieldin complex mediates 53BP1-dependent DNA repair. *Nature,* 560**,** 117--121.

NOORDERMEER, S. M. & ATTIKUM, H. V. 2019. PARP Inhibitor Resistance: A Tug-of-War in BRCA-Mutated Cells. *Trends in Cell Biology,* 29**,** 820--834.

NORMANNO, D., NÉGREL, A., MELO, A. J. D., BETZI, S., MEEK, K. & MODESTI, M. 2017. Mutational phospho-mimicry reveals a regulatory role for the XRCC4 and XLF C-terminal tails in modulating DNA bridging during classical non-homologous end joining. *eLife,* 6**,** e22900.

NOTREDAME, C., HIGGINS, D. G. & HERINGA, J. 2000. T-coffee: a novel method for fast and accurate multiple sequence alignment11Edited by J. Thornton. *Journal of Molecular Biology,* 302**,** 205--217.

NOVAK, J. E., ROSS-MACDONALD, P. B. & ROEDER, G. S. 2001. The Budding Yeast Msh4 Protein Functions in Chromosome Synapsis and the Regulation of Crossover Distribution. *Genetics,* 158**,** 1013--1025.

NUYTEMANS, K., THEUNS, J., CRUTS, M. & BROECKHOVEN, C. V. 2010. Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update. *Human Mutation,* 31**,** 763--780.

NYKAMP, K., ANDERSON, M., POWERS, M., GARCIA, J., HERRERA, B., HO, Y.-Y., KOBAYASHI, Y., PATIL, N., THUSBERG, J., WESTBROOK, M., GROUP, T. I. C. G. & TOPPER, S. 2017. Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria. *Genetics in Medicine,* 19**,** 1105--1117.

NYSTRÖM-LAHTI, M., PERRERA, C., RÄSCHLE, M., PANYUSHKINA-SEILER, E., MARRA, G., CURCI, A., QUARESIMA, B., COSTANZO, F., D'URSO, M., VENUTA, S. & JIRICNY, J. 2002. Functional analysis of MLH1 mutations linked to hereditary nonpolyposis colon cancer. *Genes, Chromosomes and Cancer,* 33**,** 160--167.

O'DONOVAN, A., DAVIES, A. A., MOGGS, J. G., WEST, S. C. & WOOD, R. D. 1994. XPG endonuclease makes the 3′ incision in human DNA nucleotide excision repair. *Nature,* 371**,** 432--435.

O'NEIL, K. T. & DEGRADO, W. F. 1990. A Thermodynamic Scale for the Helix-Forming Tendencies of the Commonly Occurring Amino Acids. *Science,* 250**,** 646--651.

O'CONNOR, MARK J. 2015. Targeting the DNA Damage Response in Cancer. *Molecular Cell,* 60**,** 547--560.

OCHS, F., KAREMORE, G., MIRON, E., BROWN, J., SEDLACKOVA, H., RASK, M.-B., LAMPE, M., BUCKLE, V., SCHERMELLEH, L., LUKAS, J. & LUKAS, C. 2019. Stabilization of chromatin topology safeguards genome integrity. *Nature,* 574**,** 571--574.

OERS, J. M. M. V., ROA, S., WERLING, U., LIU, Y., GENSCHEL, J., HOU, H., SELLERS, R. S., MODRICH, P., SCHARFF, M. D. & EDELMANN, W. 2010. PMS2 endonuclease activity has distinct biological functions and is essential for genome maintenance. *Proceedings of the National Academy of Sciences,* 107**,** 13384--13389.

OGI, T., LIMSIRICHAIKUL, S., OVERMEER, R. M., VOLKER, M., TAKENAKA, K., CLONEY, R., NAKAZAWA, Y., NIIMI, A., MIKI, Y., JASPERS, N. G., MULLENDERS, L. H. F., YAMASHITA, S., FOUSTERI, M. I. & LEHMANN, A. R. 2010. Three DNA Polymerases, Recruited by Different Mechanisms, Carry Out NER Repair Synthesis in Human Cells. *Molecular Cell,* 37**,** 714--727.

OSTROW, K. L., MCGUIRE, V., WHITTEMORE, A. S. & DICIOCCIO, R. A. 2004. The effects of BRCA1 missense variants V1804D and M1628T on transcriptional activity. *Cancer Genetics and Cytogenetics,* 153**,** 177--180.

OZA, A. M., DISTEFANO, M. T., HEMPHILL, S. E., CUSHMAN, B. J., GRANT, A. R., SIEGERT, R. K., SHEN, J., CHAPIN, A., BOCZEK, N. J., SCHIMMENTI, L. A., MURRY, J. B., HASADSRI, L., NARA, K., KENNA, M., BOOTH, K. T., AZAIEZ, H., GRIFFITH, A., AVRAHAM, K. B., KREMER, H., REHM, H.

L., AMR, S. S., TAYOUN, A. N. A. & GROUP, C. H. L. C. D. W. 2018. Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Human Mutation,* 39**,** 1593--1613.

PAK, M. A., MARKHIEVA, K. A., NOVIKOVA, M. S., PETROV, D. S., VOROBYEV, I. S., MAKSIMOVA, E. S., KONDRASHOV, F. A. & IVANKOV, D. N. 2021. Using AlphaFold to predict the impact of single mutations on protein stability and function. *bioRxiv***,** 2021.09.19.460937.

PALOMBO, F., IACCARINO, I., NAKAJIMA, E., IKEJIMA, M., SHIMADA, T. & JIRICNY, J. 1996. hMutSβ, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Current Biology,* 6**,** 1181--1184.

PANG, Q., PROLLA, T. A. & LISKAY, R. M. 1997. Functional domains of the Saccharomyces cerevisiae Mlh1p and Pms1p DNA mismatch repair proteins and their relevance to human hereditary nonpolyposis colorectal cancer-associated mutations. *Molecular and Cellular Biology,* 17**,** 4465--4473.

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., VENTER, J. C., HAMILTON, S. R., PETERSEN, G. M., WATSON, P., LYNCH, H. T., PELTOMÄKI, P., MECKLIN, J.-P., CHAPELLE, A. D. L., KINZLER, K. W. & VOGELSTEIN, B. 1994. Mutation of a mutL Homolog in Hereditary Colon Cancer. *Science,* 263**,** 1625--1629.

PAQUET, D., KWART, D., CHEN, A., SPROUL, A., JACOB, S., TEO, S., OLSEN, K. M., GREGG, A., NOGGLE, S. & TESSIER-LAVIGNE, M. 2016. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature,* 533**,** 125--129.

PARK, C.-H., MU, D., REARDON, J. T. & SANCAR, A. 1995. The General Transcription-Repair Factor TFIIH Is Recruited to the Excision Repair Complex by the XPA Protein Independent of the TFIIE Transcription Factor (∗). *Journal of Biological Chemistry,* 270**,** 4896--4902.

PARK, J.-M. & KANG, T.-H. 2016. Transcriptional and Posttranslational Regulation of Nucleotide Excision Repair: The Guardian of the Genome against Ultraviolet Radiation. *International Journal of Molecular Sciences,* 17**,** 1840.

PARSONS, R., LI, G.-M., LONGLEY, M. J., FANG, W.-H., PAPADOPOULOS, N., JEN, J., CHAPELLE, A. D. L., KINZLER, K. W., VOGELSTEIN, B. & MODRICH, P. 1993. Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell,* 75**,** 1227--1236.

PEARLMAN, R., HARALDSDOTTIR, S., CHAPELLE, A. D. L., JONASSON, J. G., LIYANARACHCHI, S., FRANKEL, W. L., RAFNAR, T., STEFANSSON, K., PRITCHARD, C. C. & HAMPEL, H. 2019. Clinical characteristics of patients with colorectal cancer with double somatic mismatch repair mutations compared with Lynch syndrome. *Journal of Medical Genetics,* 56**,** 462.

PEARS, C. J. & LAKIN, N. D. 2014. Emerging models for DNA repair: Dictyostelium discoideum as a model for nonhomologous end-joining. *DNA Repair,* 17**,** 121--131.

PELLEGRINO, S., MICHELENA, J., TELONI, F., IMHOF, R. & ALTMEYER, M. 2017. Replication-Coupled Dilution of H4K20me2 Guides 53BP1 to Pre-replicative Chromatin. *Cell Reports,* 19**,** 1819--1831.

PELTOMÄKI, P., GAO, X. & MECKLIN, J.-P. 2001. Genotype and phenotype in hereditary nonpolyposis colon cancer: a study of families with different vs. shared predisposing mutations. *Familial Cancer,* 1**,** 9--15.

PELTOMÄKI, P. & VASEN, H. 2004. Mutations Associated with HNPCC Predisposition \&\#x2014; Update of ICG-HNPCC/INSiGHT Mutation Database. *Disease Markers,* 20**,** 269--276.

PELTOMAKI, P. & VASEN, H. F. 1997. Mutations predisposing to hereditary nonpolyposis colorectal cancer: Database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. *Gastroenterology,* 113**,** 1146--1158.

PERGOLA, F., ZDZIENICKA, M. Z. & LIEBER, M. R. 1993. V(D)J recombination in mammalian cell mutants defective in DNA double-strand break repair. *Molecular and Cellular Biology,* 13**,** 3464--3471.

PERUTZ, M. F., BAUER, C., GROS, G., LECLERCQ, F., VANDECASSERIE, C., SCHNEK, A. G., BRAUNITZER, G., FRIDAY, A. E. & JOYSEY, K. A. 1981. Allosteric regulation of crocodilian haemoglobin. *Nature,* 291**,** 682--684.

PETERSON, T. A., DOUGHTY, E. & KANN, M. G. 2013. Towards Precision Medicine: Advances in Computational Approaches for the Analysis of Human Variants. *Journal of Molecular Biology,* 425**,** 4047--4063.

PETTITT, S. J., FRANKUM, J. R., PUNTA, M., LISE, S., ALEXANDER, J., CHEN, Y., YAP, T. A., HAIDER, S., TUTT, A. N. J. & LORD, C. J. 2020. Clinical BRCA1/2 Reversion Analysis Identifies Hotspot Mutations and Predicted Neoantigens Associated with Therapy Resistance. *Cancer Discovery,* 10**,** 1475--1488.

PFISTER, S. X., MARKKANEN, E., JIANG, Y., SARKAR, S., WOODCOCK, M., ORLANDO, G., MAVROMMATI, I., PAI, C.-C., ZALMAS, L.-P., DROBNITZKY, N., DIANOV, G. L., VERRILL, C., MACAULAY, V. M., YING, S., THANGUE, N. B. L., D'ANGIOLELLA, V., RYAN, A. J. & HUMPHREY, T. C. 2015. Inhibiting WEE1 Selectively Kills Histone H3K36me3-Deficient Cancers by dNTP Starvation. *Cancer Cell,* 28**,** 557--568.

PHELAN, C. M., DAPIC, V., TICE, B., FAVIS, R., KWAN, E., BARANY, F., MANOUKIAN, S., RADICE, P., LUIJT, R. B. V. D., NESSELROOIJ, B. P. M. V., CHENEVIX-TRENCH, G., KCONFAB, CALDES, T., HOYA, M. D. L., LINDQUIST, S., TAVTIGIAN, S. V., GOLDGAR, D., BORG, A., NAROD, S. A. & MONTEIRO, A. N. A. 2005. Classification of BRCA1 missense variants of unknown clinical significance. *Journal of Medical Genetics,* 42**,** 138.

PICCIONE, M., FORTINA, A. B., FERRI, G., ANDOLINA, G., BERETTA, L., CIVIDINI, A., MARNI, E. D., CAROPPO, F., CITERNESI, U. & LIDDO, R. D. 2021. Xeroderma Pigmentosum: General Aspects and Management. *Journal of Personalized Medicine,* 11**,** 1146.

PIERCE, A. J., JOHNSON, R. D., THOMPSON, L. H. & JASIN, M. 1999. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes \& Development,* 13**,** 2633--2638.

PIIRILÄ, H., VÄLIAHO, J. & VIHINEN, M. 2006. Immunodeficiency mutation databases (IDbases). *Human Mutation,* 27**,** 1200--1208.

PIR, M. S., BILGIN, H. I., SAYICI, A., COŞKUN, F., TORUN, F. M., ZHAO, P., KANG, Y., CEVIK, S. & KAPLAN, OKTAY I. 2021. ConVarT: a search engine for matching human genetic variants with variants from non-human species. *Nucleic Acids Research*.

PLEASANCE, E. D., CHEETHAM, R. K., STEPHENS, P. J., MCBRIDE, D. J., HUMPHRAY, S. J., GREENMAN, C. D., VARELA, I., LIN, M.-L., ORDÓÑEZ, G. R., BIGNELL, G. R., YE, K., ALIPAZ, J., BAUER, M. J., BEARE, D., BUTLER, A., CARTER, R. J., CHEN, L., COX, A. J., EDKINS, S., KOKKO-GONZALES, P. I., GORMLEY, N. A., GROCOCK, R. J., HAUDENSCHILD, C. D., HIMS, M. M., JAMES, T., JIA, M., KINGSBURY, Z., LEROY, C., MARSHALL, J., MENZIES, A., MUDIE, L. J., NING, Z., ROYCE, T., SCHULZ-TRIEGLAFF, O. B., SPIRIDOU, A., STEBBINGS, L. A., SZAJKOWSKI, L., TEAGUE, J., WILLIAMSON, D., CHIN, L., ROSS, M. T., CAMPBELL, P. J., BENTLEY, D. R., FUTREAL, P. A. & STRATTON, M. R. 2010a. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature,* 463**,** 191--196.

PLEASANCE, E. D., STEPHENS, P. J., O'MEARA, S., MCBRIDE, D. J., MEYNERT, A., JONES, D., LIN, M.-L., BEARE, D., LAU, K. W., GREENMAN, C., VARELA, I., NIK-ZAINAL, S., DAVIES, H. R., ORDOÑEZ, G. R., MUDIE, L. J., LATIMER, C., EDKINS, S., STEBBINGS, L., CHEN, L., JIA, M., LEROY, C., MARSHALL, J., MENZIES, A., BUTLER, A., TEAGUE, J. W., MANGION, J., SUN, Y. A., MCLAUGHLIN, S. F., PECKHAM, H. E., TSUNG, E. F., COSTA, G. L., LEE, C. C., MINNA, J. D., GAZDAR, A., BIRNEY, E., RHODES, M. D., MCKERNAN, K. J., STRATTON, M. R., FUTREAL, P. A. & CAMPBELL, P. J. 2010b. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature,* 463**,** 184--190.

PLOTZ, G., RAEDLE, J., BRIEGER, A., TROJAN, J. & ZEUZEM, S. 2002. hMutSα forms an ATP-dependent complex with hMutLα and hMutLβ on DNA. *Nucleic Acids Research,* 30**,** 711--718.

PLOTZ, G., RAEDLE, J., BRIEGER, A., TROJAN, J. & ZEUZEM, S. 2003. N-terminus of hMLH1 confers interaction of hMutLα and hMutLβ with hMutSα. *Nucleic Acids Research,* 31**,** 3217--3226.

PLOTZ, G., WELSCH, C., GIRON-MONZON, L., FRIEDHOFF, P., ALBRECHT, M., PIIPER, A., BIONDI, R. M., LENGAUER, T., ZEUZEM, S. & RAEDLE, J. 2006. Mutations in the MutSα interaction interface of MLH1 can abolish DNA mismatch repair. *Nucleic Acids Research,* 34**,** 6574--6586.

PLUCIENNIK, A., DZANTIEV, L., IYER, R. R., CONSTANTIN, N., KADYROV, F. A. & MODRICH, P. 2010. PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. *Proceedings of the National Academy of Sciences,* 107**,** 16066--16071.

PORTER, P. C., MELLON, I. & STATES, J. C. 2005. XP-A cells complemented with Arg228Gln and Val234Leu polymorphic XPA alleles repair BPDE-induced DNA damage better than cells complemented with the wild type allele. *DNA Repair,* 4**,** 341--349.

POTENSKI, C. J. & KLEIN, H. L. 2014. How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells. *Nucleic Acids Research,* 42**,** 10226--10234.

PRADHAN, S., DAS, P. & MATTAPARTHI, V. S. K. 2018. Characterizing the Binding Interactions between DNA-Binding Proteins, XPA and XPE: A Molecular Dynamics Approach. *ACS Omega,* 3**,** 15442--15454.

PRAKASH, R., ZHANG, Y., FENG, W. & JASIN, M. 2015. Homologous Recombination and Human Health: The Roles of BRCA1, BRCA2, and Associated Proteins. *Cold Spring Harbor Perspectives in Biology,* 7**,** a016600.

PRELICH, G. 2012. Gene Overexpression: Uses, Mechanisms, and Interpretation. *Genetics,* 190**,** 841--854.

PRINDLE, M. J. & LOEB, L. A. 2012. DNA polymerase delta in dna replication and genome maintenance. *Environmental and Molecular Mutagenesis,* 53**,** 666--682.

PRITCHARD, C. C., SALIPANTE, S. J., KOEHLER, K., SMITH, C., SCROGGINS, S., WOOD, B., WU, D., LEE, M. K., DINTZIS, S., ADEY, A., LIU, Y., EATON, K. D., MARTINS, R., STRICKER, K., MARGOLIN, K. A., HOFFMAN, N., CHURPEK, J. E., TAIT, J. F., KING, M.-C. & WALSH, T. 2014. Validation and Implementation of Targeted Capture and Sequencing for the Detection of Actionable Mutation, Copy Number Variation, and Gene Rearrangement in Clinical Cancer Specimens. *The Journal of Molecular Diagnostics,* 16**,** 56--67.

PRIYA, B., RAVI, S. & KIRUBAKARAN, S. 2023. Targeting ATM and ATR for cancer therapeutics: Inhibitors in clinic. *Drug Discovery Today,* 28**,** 103662.

PROKOP, J. W., MAY, T., STRONG, K., BILINOVICH, S. M., BUPP, C., RAJASEKARAN, S., WORTHEY, E. A. & LAZAR, J. 2018. Genome sequencing in the clinic: the past, present, and future of genomic medicine. *Physiological Genomics,* 50**,** 563--579.

PROLLA, T. A., BAKER, S. M., HARRIS, A. C., TSAO, J.-L., YAO, X., BRONNER, C. E., ZHENG, B., GORDON, M., RENEKER, J., ARNHEIM, N., SHIBATA, D., BRADLEY, A. & LISKAY, R. M. 1998. Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DMA mismatch repair. *Nature Genetics,* 18**,** 276--279.

PULZOVÁ, L. B., WARD, T. A. & CHOVANEC, M. 2020. XPA: DNA Repair Protein of Significant Clinical Importance. *International Journal of Molecular Sciences,* 21**,** 2182.

RADEMAKERS, S., VOLKER, M., HOOGSTRATEN, D., NIGG, A. L., MONÉ, M. J., ZEELAND, A. A. V., HOEIJMAKERS, J. H. J., HOUTSMULLER, A. B. & VERMEULEN, W. 2003. Xeroderma Pigmentosum Group A Protein Loads as a Separate Factor onto DNA Lesions. *Molecular and Cellular Biology,* 23**,** 5755--5767.

RADERSCHALL, E., GOLUB, E. I. & 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.

RAEVAARA, T. E., KORHONEN, M. K., LOHI, H., HAMPEL, H., LYNCH, E., LÖNNQVIST, K. E., HOLINSKI-FEDER, E., SUTTER, C., MCKINNON, W., DURAISAMY, S., GERDES, A.-M., PELTOMÄKI, P., KOHONEN-CCORISH, M., MANGOLD, E., MACRAE, F., GREENBLATT, M., CHAPELLE, A. D. L. & NYSTRÖM, M. 2005. Functional Significance and Clinical Phenotype of Nontruncating Mismatch Repair Variants of MLH1. *Gastroenterology,* 129**,** 537--549.

RAFNAR, T., GUDBJARTSSON, D. F., SULEM, P., JONASDOTTIR, A., SIGURDSSON, A., JONASDOTTIR, A., BESENBACHER, S., LUNDIN, P., STACEY, S. N., GUDMUNDSSON, J., MAGNUSSON, O. T., ROUX, L. L., ORLYGSDOTTIR, G., HELGADOTTIR, H. T., JOHANNSDOTTIR, H., GYLFASON, A., TRYGGVADOTTIR, L., JONASSON, J. G., JUAN, A. D., ORTEGA, E., RAMON-CAJAL, J. M., GARCÍA-PRATS, M. D., MAYORDOMO, C., PANADERO, A., RIVERA, F., ABEN, K. K. H., ALTENA, A. M. V., MASSUGER, L. F. A. G., AAVIKKO, M., KUJALA, P. M., STAFF, S., AALTONEN, L. A., OLAFSDOTTIR, K., BJORNSSON, J., KONG, A., SALVARSDOTTIR, A., SAEMUNDSSON, H., OLAFSSON, K., BENEDIKTSDOTTIR, K. R., GULCHER, J., MASSON, G., KIEMENEY, L. A., MAYORDOMO, J. I., THORSTEINSDOTTIR, U. & STEFANSSON, K. 2011. Mutations in BRIP1 confer high risk of ovarian cancer. *Nature Genetics,* 43**,** 1104--1107.

RAGLAND, R. L., PATEL, S., RIVARD, R. S., SMITH, K., PETERS, A. A., BIELINSKY, A.-K. & BROWN, E. J. 2013. RNF4 and PLK1 are required for replication fork collapse in ATR-deficient cells. *Genes \& Development,* 27**,** 2259--2273.

RAJENDRA, E., OESTERGAARD, VIBE H., LANGEVIN, F., WANG, M., DORNAN, GILLIAN L., PATEL, KETAN J. & PASSMORE, LORI A. 2014. The Genetic and Biochemical Basis of FANCD2 Monoubiquitination. *Molecular Cell,* 54**,** 858--869.

RAMILO, C., GU, L., GUO, S., ZHANG, X., PATRICK, S. M., TURCHI, J. J. & LI, G.-M. 2002. Partial Reconstitution of Human DNA Mismatch Repair In Vitro: Characterization of the Role of Human Replication Protein A. *Molecular and Cellular Biology,* 22**,** 2037--2046.

RANJHA, L., ANAND, R. & CEJKA, P. 2014. The Saccharomyces cerevisiae Mlh1-Mlh3 Heterodimer Is an Endonuclease That Preferentially Binds to Holliday Junctions*. *Journal of Biological Chemistry,* 289**,** 5674--5686.

RÄSCHLE, M., DUFNER, P., MARRA, G. & JIRICNY, J. 2002. Mutations within the hMLH1 and hPMS2 Subunits of the Human MutLα Mismatch Repair Factor Affect Its ATPase Activity, but Not Its Ability to Interact with hMutSα*. *Journal of Biological Chemistry,* 277**,** 21810--21820.

RÄSCHLE, M., KNIPSCHEER, P., KNIPSHEER, P., ENOIU, M., ANGELOV, T., SUN, J., GRIFFITH, J. D., ELLENBERGER, T. E., SCHÄRER, O. D. & WALTER, J. C. 2008. Mechanism of Replication-Coupled DNA Interstrand Crosslink Repair. *Cell,* 134**,** 969--980.

RÄSCHLE, M., MARRA, G., NYSTRÖM-LAHTI, M., SCHÄR, P. & JIRICNY, J. 1999. Identification of hMutLβ, a Heterodimer of hMLH1 and hPMS1*. *Journal of Biological Chemistry,* 274**,** 32368--32375.

RASTOGI, R. P., RICHA, KUMAR, A., TYAGI, M. B. & SINHA, R. P. 2010. Molecular Mechanisms of Ultraviolet Radiation-Induced DNA Damage and Repair. *Journal of Nucleic Acids,* 2010**,** 592980.

RATH, A., MISHRA, A., FERREIRA, V. D., HU, C., OMERZA, G., KELLY, K., HESSE, A., REDDI, H. V., GRADY, J. P. & HEINEN, C. D. 2019. Functional interrogation of Lynch syndrome-associated MSH2 missense variants via CRISPR-Cas9 gene editing in human embryonic stem cells. *Human Mutation,* 40**,** 2044--2056.

RATH, A., RADECKI, A. A., RAHMAN, K., GILMORE, R. B., HUDSON, J. R., CENCI, M., TAVTIGIAN, S. V., GRADY, J. P. & HEINEN, C. D. 2022. A calibrated cell-based functional assay to aid classification of MLH1 DNA mismatch repair gene variants. *Human Mutation*.

RATZ, M., TESTA, I., HELL, S. W. & JAKOBS, S. 2015. CRISPR/Cas9-mediated endogenous protein tagging for RESOLFT super-resolution microscopy of living human cells. *Scientific Reports,* 5**,** 9592.

RAYNER, E., TIERSMA, Y., FORTUNO, C., HEES-STUIVENBERG, S., DROST, M., THOMPSON, B., SPURDLE, A. B. & WIND, N. 2022. Predictive functional assay-based classification of PMS2 variants in Lynch syndrome. *Human Mutation,* 43**,** 1249--1258.

RECUERO-CHECA, M. A., DORÉ, A. S., ARIAS-PALOMO, E., RIVERA-CALZADA, A., SCHERES, S. H. W., MAMAN, J. D., PEARL, L. H. & LLORCA, O. 2009. Electron microscopy of Xrcc4 and the DNA ligase IV–Xrcc4 DNA repair complex. *DNA Repair,* 8**,** 1380--1389.

REHM, H. L., BERG, J. S., BROOKS, L. D., BUSTAMANTE, C. D., EVANS, J. P., LANDRUM, M. J., LEDBETTER, D. H., MAGLOTT, D. R., MARTIN, C. L., NUSSBAUM, R. L., PLON, S. E., RAMOS, E. M., SHERRY, S. T., WATSON, M. S. & CLINGEN 2015. ClinGen — The Clinical Genome Resource. *New England Journal of Medicine,* 372**,** 2235--2242.

REID, D. A., KEEGAN, S., LEO-MACIAS, A., WATANABE, G., STRANDE, N. T., CHANG, H. H., OKSUZ, B. A., FENYO, D., LIEBER, M. R., RAMSDEN, D. A. & ROTHENBERG, E. 2015. Organization and dynamics of the nonhomologous end-joining machinery during DNA double-strand break repair. *Proceedings of the National Academy of Sciences,* 112**,** E2575--E2584.

REINHOLD, W. C., SUNSHINE, M., LIU, H., VARMA, S., KOHN, K. W., MORRIS, J., DOROSHOW, J. & POMMIER, Y. 2012. CellMiner: A Web-Based Suite of Genomic and Pharmacologic Tools to Explore Transcript and Drug Patterns in the NCI-60 Cell Line Set. *Cancer Research,* 72**,** 3499--3511.

RENNIE, M. L., LEMONIDIS, K., ARKINSON, C., CHAUGULE, V. K., CLARKE, M., STREETLEY, J., SPAGNOLO, L. & WALDEN, H. 2020. Differential functions of FANCI and FANCD2 ubiquitination stabilize ID2 complex on DNA. *EMBO reports,* 21**,** e50133.

RIABINSKA, A., DAHEIM, M., HERTER-SPRIE, G. S., WINKLER, J., FRITZ, C., HALLEK, M., THOMAS, R. K., KREUZER, K.-A., FRENZEL, L. P., MONFARED, P., MARTINS-BOUCAS, J., CHEN, S. & REINHARDT, H. C. 2013. Therapeutic Targeting of a Robust Non-Oncogene Addiction to PRKDC in ATM-Defective Tumors. *Science Translational Medicine,* 5**,** 189ra78.

RIBALLO, E., WOODBINE, L., STIFF, T., WALKER, S. A., GOODARZI, A. A. & JEGGO, P. A. 2009. XLF-Cernunnos promotes DNA ligase IV–XRCC4 re-adenylation following ligation. *Nucleic Acids Research,* 37**,** 482--492.

RIBIC, C. M., SARGENT, D. J., MOORE, M. J., THIBODEAU, S. N., FRENCH, A. J., GOLDBERG, R. M., HAMILTON, S. R., LAURENT-PUIG, P., GRYFE, R., SHEPHERD, L. E., TU, D., REDSTON, M. & GALLINGER, S. 2003. Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. *The New England Journal of Medicine,* 349**,** 247--257.

RICHARDS, C. S., BALE, S., BELLISSIMO, D. B., DAS, S., GRODY, W. W., HEGDE, M. R., LYON, E., WARD, B. E. & COMMITTEE, T. M. S. O. T. A. L. Q. A. 2008. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genetics in Medicine,* 10**,** 294--300.

RICHARDS, S., AZIZ, N., BALE, S., BICK, D., DAS, S., GASTIER-FOSTER, J., GRODY, W. W., HEGDE, M., LYON, E., SPECTOR, E., VOELKERDING, K., REHM, H. L. & COMMITTEE, A. L. Q. A. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine,* 17**,** 405--423.

RICHARDSON, M. E., HU, C., LEE, K. Y., LADUCA, H., FULK, K., DURDA, K. M., DECKMAN, A. M., GOLDGAR, D. E., MONTEIRO, A. N. A., GNANAOLIVU, R., HART, S. N., POLLEY, E. C., CHAO, E., PESARAN, T. & COUCH, F. J. 2021. Strong functional data for pathogenicity or neutrality classify BRCA2 DNA-binding-domain variants of uncertain significance. *The American Journal of Human Genetics*.

RICHTER-PECHAŃSKA, P., KUNZ, J. B., HOF, J., ZIMMERMANN, M., RAUSCH, T., BANDAPALLI, O. R., ORLOVA, E., SCAPINELLO, G., SAGI, J. C., STANULLA, M., SCHRAPPE, M., CARIO, G., KIRSCHNER-SCHWABE, R., ECKERT, C., BENES, V., KORBEL, J. O., MUCKENTHALER, M. U. & KULOZIK, A. E. 2017. Identification of a genetically defined ultra-high-risk group in relapsed pediatric T-lymphoblastic leukemia. *Blood Cancer Journal,* 7**,** e523--e523.

RICKMAN, KIMBERLY A., LACH, FRANCIS P., ABHYANKAR, A., DONOVAN, FRANK X., SANBORN, ERICA M., KENNEDY, JENNIFER A., SOUGNEZ, C., GABRIEL, STACEY B., ELEMENTO, O., CHANDRASEKHARAPPA, SETTARA C., SCHINDLER, D., AUERBACH, ARLEEN D. & SMOGORZEWSKA, A. 2015. Deficiency of UBE2T, the E2 Ubiquitin Ligase Necessary for

FANCD2 and FANCI Ubiquitination, Causes FA-T Subtype of Fanconi Anemia. *Cell Reports,* 12**,** 35--41.

ROBINSON, P. S., COORENS, T. H. H., PALLES, C., MITCHELL, E., ABASCAL, F., OLAFSSON, S., LEE, B. C. H., LAWSON, A. R. J., LEE-SIX, H., MOORE, L., SANDERS, M. A., HEWINSON, J., MARTIN, L., PINNA, C. M. A., GALAVOTTI, S., RAHBARI, R., CAMPBELL, P. J., MARTINCORENA, I., TOMLINSON, I. & STRATTON, M. R. 2021. Increased somatic mutation burdens in normal human cells due to defective DNA polymerases. *Nature Genetics,* 53**,** 1434--1442.

ROCH, B., ABRAMOWSKI, V., ETIENNE, O., MUSILLI, S., DAVID, P., CHARBONNIER, J.-B., CALLEBAUT, I., BOUSSIN, F. D. & VILLARTAY, J.-P. D. 2021. An XRCC4 mutant mouse, a model for human X4 syndrome, reveals interplays with Xlf, PAXX, and ATM in lymphoid development. *eLife,* 10**,** e69353.

RODRÍGUEZ–SOLER, M., PÉREZ–CARBONELL, L., GUARINOS, C., ZAPATER, P., CASTILLEJO, A., BARBERÁ, V. M., JUÁREZ, M., BESSA, X., XICOLA, R. M., CLOFENT, J., BUJANDA, L., BALAGUER, F., REÑÉ, J. M., DE–CASTRO, L., MARÍN–GABRIEL, J. C., LANAS, A., CUBIELLA, J., NICOLÁS–PÉREZ, D., BREA–FERNÁNDEZ, A., CASTELLVÍ–BEL, S., ALENDA, C., RUIZ–PONTE, C., CARRACEDO, A., CASTELLS, A., ANDREU, M., LLOR, X., SOTO, J. L., PAYÁ, A. & JOVER, R. 2013. Risk of Cancer in Cases of Suspected Lynch Syndrome Without Germline Mutation. *Gastroenterology,* 144**,** 926--932.e1.

ROGAKOU, E. P., PILCH, D. R., ORR, A. H., IVANOVA, V. S. & BONNER, W. M. 1998. DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139*. *Journal of Biological Chemistry,* 273**,** 5858--5868.

ROMEO, F., FALBO, L., SANZO, M. D., MISAGGI, R., FANIELLO, M. C., VIGLIETTO, G., CUDA, G., COSTANZO, F. & QUARESIMA, B. 2011. BRCA1 is required for hMLH1 stabilization following doxorubicin-induced DNA damage. *The International Journal of Biochemistry \& Cell Biology,* 43**,** 1754--1763.

ROPARS, V., DREVET, P., LEGRAND, P., BACONNAIS, S., AMRAM, J., FAURE, G., MARQUEZ, J. A., PIETREMENT, O., GUEROIS, R., CALLEBAUT, I., CAM, E. L., REVY, P., VILLARTAY, J. P. D. & CHARBONNIER, J. B. 2011. Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining. *Proceedings of the National Academy of Sciences,* 108**,** 12663--12668.

ROSE, M., BURGESS, J. T., O'BYRNE, K., RICHARD, D. J. & BOLDERSON, E. 2020. PARP Inhibitors: Clinical Relevance, Mechanisms of Action and Tumor Resistance. *Frontiers in Cell and Developmental Biology,* 8**,** 564601.

ROSIN, N., ELCIOGLU, N. H., BELEGGIA, F., ISGÜVEN, P., ALTMÜLLER, J., THIELE, H., STEINDL, K., JOSET, P., RAUCH, A., NÜRNBERG, P., WOLLNIK, B. & YIGIT, G. 2015. Mutations in XRCC4 cause primary microcephaly, short stature and increased genomic instability. *Human Molecular Genetics,* 24**,** 3708--3717.

ROSS-MACDONALD, P. & ROEDER, G. S. 1994. Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. *Cell,* 79**,** 1069--1080.

ROTH, D. B. 2014. V(D)J Recombination: Mechanism, Errors, and Fidelity. *Microbiology Spectrum,* 2.

ROUDKO, V., BOZKUS, C. C., GREENBAUM, B., LUCAS, A., SAMSTEIN, R. & BHARDWAJ, N. 2021. Lynch Syndrome and MSI-H Cancers: From Mechanisms to "Off-The-Shelf" Cancer Vaccines. *Frontiers in Immunology,* 12**,** 757804.

ROWLING, P. J. E., COOK, R. & ITZHAKI, L. S. 2010. Toward Classification of BRCA1 Missense Variants Using a Biophysical Approach. *Journal of Biological Chemistry,* 285**,** 20080--20087.

ROY, R., CHUN, J. & POWELL, S. N. 2012. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nature Reviews Cancer,* 12**,** 68--78.

ROY, S., MELO, A. J. D., XU, Y., TADI, S. K., NÉGREL, A., HENDRICKSON, E., MODESTI, M. & MEEK, K. 2015. XRCC4/XLF Interaction Is Variably Required for DNA Repair and Is Not Required for Ligase IV Stimulation. *Molecular and Cellular Biology,* 35**,** 3017--3028.

RUIZ-PESINI, E., LOTT, M. T., PROCACCIO, V., POOLE, J. C., BRANDON, M. C., MISHMAR, D., YI, C., KREUZIGER, J., BALDI, P. & WALLACE, D. C. 2007. An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Research,* 35**,** D823--D828.

RYDBERG, B. 1978. Bromouracil mutagenesis and mismatch repair in mutator strains of Escherichia coli. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 52**,** 11--24.

SACCO, F., HUMPHREY, S. J., COX, J., MISCHNIK, M., SCHULTE, A., KLABUNDE, T., SCHÄFER, M. & MANN, M. 2016. Glucose-regulated and drug-perturbed phosphoproteome reveals molecular mechanisms controlling insulin secretion. *Nature Communications,* 7**,** 13250.

SAIJO, M., KURAOKA, I., MASUTANI, C., HANAOKA, F. & TANAKA, K. 1996. Sequential Binding of DNA Repair Proteins RPA and ERCC1 to XPA in vitro. *Nucleic Acids Research,* 24**,** 4719--4724.

SAIJO, M., MATSUDA, T., KURAOKA, I. & TANAKA, K. 2004. Inhibition of nucleotide excision repair by anti-XPA monoclonal antibodies which interfere with binding to RPA, ERCC1, and TFIIH. *Biochemical and Biophysical Research Communications,* 321**,** 815--822.

SAIJO, M., TAKEDACHI, A. & TANAKA, K. 2011. Nucleotide Excision Repair by Mutant Xeroderma Pigmentosum Group A (XPA) Proteins with Deficiency in Interaction with RPA*. *Journal of Biological Chemistry,* 286**,** 5476--5483.

SAINI, N., GIACOBONE, C. K., KLIMCZAK, L. J., PAPAS, B. N., BURKHOLDER, A. B., LI, J.-L., FARGO, D. C., BAI, R., GERRISH, K., INNES, C. L., SCHURMAN, S. H. & GORDENIN, D. A. 2021. UV-exposure, endogenous DNA damage, and DNA replication errors shape the spectra of genome changes in human skin. *PLoS Genetics,* 17**,** e1009302.

SAITO, S., KUROSAWA, A. & ADACHI, N. 2016. Mutations in XRCC4 cause primordial dwarfism without causing immunodeficiency. *Journal of Human Genetics,* 61**,** 679--685.

SALLMYR, A. & TOMKINSON, A. E. 2018. Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. *Journal of Biological Chemistry,* 293**,** 10536--10546.

SANAVIA, T., BIROLO, G., MONTANUCCI, L., TURINA, P., CAPRIOTTI, E. & FARISELLI, P. 2020. Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine. *Computational and Structural Biotechnology Journal,* 18**,** 1968--1979.

SANCHEZ, Y., WONG, C., THOMA, R. S., RICHMAN, R., WU, Z., PIWNICA-WORMS, H. & 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.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences,* 74**,** 5463--5467.

SANTAMARIA, A., WANG, B., ELOWE, S., MALIK, R., ZHANG, F., BAUER, M., SCHMIDT, A., SILLJÉ, H. H. W., KÖRNER, R. & NIGG, E. A. 2011. The Plk1-dependent Phosphoproteome of the Early Mitotic Spindle*. *Molecular \& Cellular Proteomics,* 10**,** M110.004457.

SANTOS-PEREIRA, J. M. & AGUILERA, A. 2015. R loops: new modulators of genome dynamics and function. *Nature Reviews Genetics,* 16**,** 583--597.

SANTOS, T. D. L., HUNTER, N., LEE, C., LARKIN, B., LOIDL, J. & HOLLINGSWORTH, N. M. 2003. The Mus81/Mms4 Endonuclease Acts Independently of Double-Holliday Junction Resolution to Promote a Distinct Subset of Crossovers During Meiosis in Budding Yeast. *Genetics,* 164**,** 81--94.

SARGENT, D. J., MARSONI, S., MONGES, G., THIBODEAU, S. N., LABIANCA, R., HAMILTON, S. R., FRENCH, A. J., KABAT, B., FOSTER, N. R., TORRI, V., RIBIC, C., GROTHEY, A., MOORE, M., ZANIBONI, A., SEITZ, J.-F., SINICROPE, F. & GALLINGER, S. 2010. Defective Mismatch Repair As a Predictive Marker for Lack of Efficacy of Fluorouracil-Based Adjuvant Therapy in Colon Cancer. *Journal of Clinical Oncology,* 28**,** 3219--3226.

SARTORI, A. A., LUKAS, C., COATES, J., MISTRIK, M., FU, S., BARTEK, J., BAER, R., LUKAS, J. & JACKSON, S. P. 2007. Human CtIP promotes DNA end resection. *Nature,* 450**,** 509--514.

SATOKATA, I., TANAKA, K., MIURA, N., NARITA, M., MIMAKI, T., SATOH, Y., KONDO, S. & OKADA, Y. 1992. Three nonsense mutations responsible for group A xeroderma pigmentosum. *Mutation Research/DNA Repair,* 273**,** 193--202.

SAUNDERS, C. T., WONG, W. S. W., SWAMY, S., BECQ, J., MURRAY, L. J. & CHEETHAM, R. K. 2012. Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics,* 28**,** 1811--1817.

SAVITSKY, K., BAR-SHIRA, A., GILAD, S., ROTMAN, G., ZIV, Y., VANAGAITE, L., TAGLE, D. A., SMITH, S., UZIEL, T., SFEZ, S., ASHKENAZI, M., PECKER, I., FRYDMAN, M., HARNIK, R., PATANJALI, S. R., SIMMONS, A., CLINES, G. A., SARTIEL, A., GATTI, R. A., CHESSA, L., SANAL, O., LAVIN, M. F., JASPERS, N. G. J., TAYLOR, A. M. R., ARLETT, C. F., MIKI, T., WEISSMAN, S. M., LOVETT, M., COLLINS, F. S. & SHILOH, Y. 1995. A Single Ataxia Telangiectasia Gene with a Product Similar to PI-3 Kinase. *Science,* 268**,** 1749--1753.

SCALZITTI, N., KRESS, A., ORHAND, R., WEBER, T., MOULINIER, L., JEANNIN-GIRARDON, A., COLLET, P., POCH, O. & THOMPSON, J. D. 2021. Spliceator: multi-species splice site prediction using convolutional neural networks. *BMC Bioinformatics,* 22**,** 561.

SCHÄRER, O. D. 2013. Nucleotide excision repair in eukaryotes. *Cold Spring Harbor perspectives in biology,* 5**,** a012609.

SCHMUTTE, C., SADOFF, M. M., SHIM, K.-S., ACHARYA, S. & FISHEL, R. 2001. The Interaction of DNA Mismatch Repair Proteins with Human Exonuclease I*. *Journal of Biological Chemistry,* 276**,** 33011--33018.

SCHNEIDER, C. A., RASBAND, W. S. & ELICEIRI, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods,* 9**,** 671--675.

SCHREIBER, V., DANTZER, F., AME, J.-C. & MURCIA, G. D. 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. & 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. *American journal of human genetics,* 60**,** 1031--40.

SCHÜRER, K. A., RUDOLPH, C., ULRICH, H. D. & KRAMER, W. 2004. Yeast MPH1 Gene Functions in an Error-Free DNA Damage Bypass Pathway That Requires Genes From Homologous Recombination, but Not From Postreplicative Repair. *Genetics,* 166**,** 1673--1686.

SCHYMKOWITZ, J., BORG, J., STRICHER, F., NYS, R., ROUSSEAU, F. & SERRANO, L. 2005. The FoldX web server: an online force field. *Nucleic Acids Research,* 33**,** W382--W388.

SCRIMA, A., KONÍČKOVÁ, R., CZYZEWSKI, B. K., KAWASAKI, Y., JEFFREY, P. D., GROISMAN, R., NAKATANI, Y., IWAI, S., PAVLETICH, N. P. & THOMÄ, N. H. 2008. Structural Basis of UV DNA-Damage Recognition by the DDB1–DDB2 Complex. *Cell,* 135**,** 1213--1223.

SCULLY, R., GANESAN, S., VLASAKOVA, K., CHEN, J., SOCOLOVSKY, M. & LIVINGSTON, D. M. 1999. Genetic Analysis of BRCA1 Function in a Defined Tumor Cell Line. *Molecular Cell,* 4**,** 1093--1099.

SCULLY, R., PANDAY, A., ELANGO, R. & WILLIS, N. A. 2019. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews Molecular Cell Biology,* 20**,** 698--714.

SEMENENKO, V. A. & STEWART, R. D. 2004. A Fast Monte Carlo Algorithm to Simulate the Spectrum of DNA Damages Formed by Ionizing Radiation. *Radiation Research,* 161**,** 451--457.

SEMLOW, D. R., ZHANG, J., BUDZOWSKA, M., DROHAT, A. C. & WALTER, J. C. 2016. Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. *Cell,* 167**,** 498--511.e14.

SETHY, C. & KUNDU, C. N. 2021. 5-Fluorouracil (5-FU) resistance and the new strategy to enhance the sensitivity against cancer: Implication of DNA repair inhibition. *Biomedicine \& Pharmacotherapy,* 137**,** 111285.

SETTON, J., ZINDA, M., RIAZ, N., DUROCHER, D., ZIMMERMANN, M., KOEHLER, M., REIS-FILHO, J. S. & POWELL, S. N. 2021. Synthetic Lethality in Cancer Therapeutics: The Next Generation. *Cancer Discovery,* 11**,** 1626--1635.

SFEIR, A. & SYMINGTON, L. S. 2015. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends in Biochemical Sciences,* 40**,** 701--714.

SHAHEEN, R., FAQEIH, E., ANSARI, S., ABDEL-SALAM, G., AL-HASSNAN, Z. N., AL-SHIDI, T., ALOMAR, R., SOGATY, S. & ALKURAYA, F. S. 2014. Genomic analysis of primordial dwarfism reveals novel disease genes. *Genome Research,* 24**,** 291--299.

SHAKEEL, S., RAJENDRA, E., ALCÓN, P., O'REILLY, F., CHOREV, D. S., MASLEN, S., DEGLIESPOSTI, G., RUSSO, C. J., HE, S., HILL, C. H., SKEHEL, J. M., SCHERES, S. H. W., PATEL, K. J., RAPPSILBER, J., ROBINSON, C. V. & PASSMORE, L. A. 2019. Structure of the Fanconi anaemia monoubiquitin ligase complex. *Nature***,** 1--4.

SHANG, Y. & MENG, F.-L. 2021. Repair of programmed DNA lesions in antibody class switch recombination: common and unique features. *Genome Instability \& Disease,* 2**,** 115--125.

SHARAN, S. K., PYLE, A., COPPOLA, V., BABUS, J., SWAMINATHAN, S., BENEDICT, J., SWING, D., MARTIN, B. K., TESSAROLLO, L., EVANS, J. P., FLAWS, J. A. & HANDEL, M. A. 2004. BRCA2 deficiency in mice leads to meiotic impairment and infertility. *Development,* 131**,** 131--142.

SHARMA, K., D'SOUZA, ROCHELLE C. J., TYANOVA, S., SCHAAB, C., WIŚNIEWSKI, JACEK R., COX, J. & MANN, M. 2014. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Reports,* 8**,** 1583--1594.

SHARMA, S. & CANMAN, C. E. 2012. REV1 and DNA polymerase zeta in DNA interstrand crosslink repair. *Environmental and Molecular Mutagenesis,* 53**,** 725--740.

SHCHERBAKOVA, P. V. & KUNKEL, T. A. 1999. Mutator Phenotypes Conferred by MLH1 Overexpression and by Heterozygosity for mlh1 Mutations. *Molecular and Cellular Biology,* 19**,** 3177--3183.

SHELL, S. M., LI, Z., SHKRIABAI, N., KVARATSKHELIA, M., BROSEY, C., SERRANO, M. A., CHAZIN, W. J., MUSICH, P. R. & ZOU, Y. 2009. Checkpoint Kinase ATR Promotes Nucleotide Excision Repair of UV-induced DNA Damage via Physical Interaction with Xeroderma Pigmentosum Group A*. *Journal of Biological Chemistry,* 284**,** 24213--24222.

SHENDURE, J., BALASUBRAMANIAN, S., CHURCH, G. M., GILBERT, W., ROGERS, J., SCHLOSS, J. A. & WATERSTON, R. H. 2017. DNA sequencing at 40: past, present and future. *Nature,* 550**,** 345--353.

SHENDURE, J., PORRECA, G. J., REPPAS, N. B., LIN, X., MCCUTCHEON, J. P., ROSENBAUM, A. M., WANG, M. D., ZHANG, K., MITRA, R. D. & CHURCH, G. M. 2005. Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science,* 309**,** 1728--1732.

SHERRY, S. T., WARD, M. H., KHOLODOV, M., BAKER, J., PHAN, L., SMIGIELSKI, E. M. & SIROTKIN, K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research,* 29**,** 308--311.

SHIBATA, A., CONRAD, S., BIRRAUX, J., GEUTING, V., BARTON, O., ISMAIL, A., KAKAROUGKAS, A., MEEK, K., TAUCHER-SCHOLZ, G., LÖBRICH, M. & JEGGO, P. A. 2011. Factors determining DNA double-strand break repair pathway choice in G2 phase. *The EMBO Journal,* 30**,** 1079--1092.

SHIBATA, A. & JEGGO, P. A. 2021. ATM's Role in the Repair of DNA Double-Strand Breaks. *Genes,* 12**,** 1370.

SHILOH, Y. & LEDERMAN, H. M. 2017. Ataxia-telangiectasia (A-T): An emerging dimension of premature ageing. *Ageing Research Reviews,* 33**,** 76--88.

SHIMA, N., MUNROE, R. J. & SCHIMENTI, J. C. 2004. The Mouse Genomic Instability Mutation chaos1 Is an Allele of Polq That Exhibits Genetic Interaction with Atm. *Molecular and Cellular Biology,* 24**,** 10381--10389.

SHIMAMOTO, T., TANIMURA, T., YONEDA, Y., KOBAYAKAWA, Y., SUGASAWA, K., HANAOKA, F., OKA, M., OKADA, Y., TANAKA, K. & KOHNO, K. 1995. Expression and Functional Analyses of the Dxpa Gene, the Drosophila Homolog of the Human Excision Repair Gene XPA. *Journal of Biological Chemistry,* 270**,** 22452--22459.

SHIMODAIRA, H., FILOSI, N., SHIBATA, H., SUZUKI, T., RADICE, P., KANAMARU, R., FRIEND, S. H., KOLODNER, R. D. & ISHIOKA, C. 1998. Functional analysis of human MLH1 mutations in Saccharomyces cerevisiae. *Nature Genetics,* 19**,** 384--389.

SHIVJI, M. K. K., PODUST, V. N., HUEBSCHER, U. & WOOD, R. D. 1995. Nucleotide Excision Repair DNA Synthesis by DNA Polymerase .epsilon. in the Presence of PCNA, RFC, and RPA. *Biochemistry,* 34**,** 5011--5017.

SIBANDA, B. L., CRITCHLOW, S. E., BEGUN, J., PEI, X. Y., JACKSON, S. P., BLUNDELL, T. L. & PELLEGRINI, L. 2001. Crystal structure of an Xrcc4–DNA ligase IV complex. *Nature Structural Biology,* 8**,** 1015--1019.

SIGURDSSON, S., DIRAC-SVEJSTRUP, A. B. & SVEJSTRUP, J. Q. 2010. Evidence that Transcript Cleavage Is Essential for RNA Polymerase II Transcription and Cell Viability. *Molecular Cell,* 38**,** 202--210.

SIJACKI, T., ALCÓN, P., CHEN, Z. A., MCLAUGHLIN, S. H., SHAKEEL, S., RAPPSILBER, J. & PASSMORE, L. A. 2022. The DNA-damage kinase ATR activates the FANCD2-FANCI clamp by priming it for ubiquitination. *Nature Structural \& Molecular Biology,* 29**,** 881--890.

SILVA, I. U. D., MCHUGH, P. J., CLINGEN, P. H. & HARTLEY, J. A. 2000. Defining the Roles of Nucleotide Excision Repair and Recombination in the Repair of DNA Interstrand Cross-Links in Mammalian Cells. *Molecular and Cellular Biology,* 20**,** 7980--7990.

SIMONEAU, A. & ZOU, L. 2021. An extending ATR–CHK1 circuitry: the replication stress response and beyond. *Current Opinion in Genetics \& Development,* 71**,** 92--98.

SIMONETTA, M., KRIJGER, I. D., SERRAT, J., MOATTI, N., FORTUNATO, D., HOEKMAN, L., BLEIJERVELD, O. B., ALTELAAR, A. F. M. & 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, P., FRAGOZA, R., BLENGINI, C. S., TRAN, T. N., PANNAFINO, G., AL-SWEEL, N., SCHIMENTI, K. J., SCHINDLER, K., ALANI, E. A., YU, H. & SCHIMENTI, J. C. 2021. Human MLH1/3 variants causing aneuploidy, pregnancy loss, and premature reproductive aging. *Nature Communications,* 12**,** 5005.

SINGH, T. R., ALI, A. M., BUSYGINA, V., RAYNARD, S., FAN, Q., DU, C.-H., ANDREASSEN, P. R., SUNG, P. & 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.

SINGH, T. R., SARO, D., ALI, A. M., ZHENG, X.-F., DU, C.-H., KILLEN, M. W., SACHPATZIDIS, A., WAHENGBAM, K., PIERCE, A. J., XIONG, Y., SUNG, P. & MEETEI, A. R. 2010. MHF1-MHF2, a Histone-Fold-Containing Protein Complex, Participates in the Fanconi Anemia Pathway via FANCM. *Molecular Cell,* 37**,** 879--886.

SINHA, R. P. & HÄDER, D.-P. 2002. UV-induced DNA damage and repair: a review. *Photochemical \& Photobiological Sciences,* 1**,** 225--236.

SMITH, L. M., SANDERS, J. Z., KAISER, R. J., HUGHES, P., DODD, C., CONNELL, C. R., HEINER, C., KENT, S. B. H. & HOOD, L. E. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature,* 321**,** 674--679.

SMOGORZEWSKA, A., DESETTY, R., SAITO, T. T., SCHLABACH, M., LACH, F. P., SOWA, M. E., CLARK, A. B., KUNKEL, T. A., HARPER, J. W., COLAIÁCOVO, M. P. & ELLEDGE, S. J. 2010. A Genetic Screen Identifies FAN1, a Fanconi Anemia-Associated Nuclease Necessary for DNA Interstrand Crosslink Repair. *Molecular Cell,* 39**,** 36--47.

SMOGORZEWSKA, A., MATSUOKA, S., VINCIGUERRA, P., MCDONALD, E. R., HUROV, K. E., LUO, J., BALLIF, B. A., GYGI, S. P., HOFMANN, K., D'ANDREA, A. D. & ELLEDGE, S. J. 2007. Identification of the FANCI Protein, a Monoubiquitinated FANCD2 Paralog Required for DNA Repair. *Cell,* 129**,** 289--301.

SOBOCIŃSKA, J., KOLENDA, T., TERESIAK, A., BADZIĄG-LEŚNIAK, N., KOPCZYŃSKA, M., GUGLAS, K., PRZYBYŁA, A., FILAS, V., BOGAJEWSKA-RYŁKO, E., LAMPERSKA, K. & MACKIEWICZ, A. 2020.

Diagnostics of Mutations in MMR/EPCAM Genes and Their Role in the Treatment and Care of Patients with Lynch Syndrome. *Diagnostics,* 10**,** 786.

SOKOL, E. S., JIN, D. X., FINE, A., TRABUCCO, S. E., MAUND, S., FRAMPTON, G., MOLINERO, L. & ANTONARAKIS, E. S. 2022. PARP Inhibitor Insensitivity to BRCA1/2 Monoallelic Mutations in Microsatellite Instability-High Cancers. *JCO Precision Oncology,* 6**,** e2100531.

SPAMPINATO, C. & MODRICH, P. 2000. The MutL ATPase Is Required for Mismatch Repair*. *Journal of Biological Chemistry,* 275**,** 9863--9869.

SPIVAK, G. 2015. Nucleotide excision repair in humans. *DNA repair,* 36**,** 13--8.

SPUGNESI, L., BALIA, C., COLLAVOLI, A., FALASCHI, E., QUERCIOLI, V., CALIGO, M. A. & GALLI, A. 2013. Effect of the expression of BRCA2 on spontaneous homologous recombination and DNA damage-induced nuclear foci in Saccharomyces cerevisiae. *Mutagenesis,* 28**,** 187--195.

SPURDLE, A. B., HEALEY, S., DEVEREAU, A., HOGERVORST, F. B. L., MONTEIRO, A. N. A., NATHANSON, K. L., RADICE, P., STOPPA-LYONNET, D., TAVTIGIAN, S., WAPPENSCHMIDT, B., COUCH, F. J., GOLDGAR, D. E. & ENIGMA 2012. ENIGMA—Evidence-based network for the interpretation of germline mutant alleles: An international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Human Mutation,* 33**,** 2--7.

SPYCHER, C., MILLER, E. S., TOWNSEND, K., PAVIC, L., MORRICE, N. A., JANSCAK, P., STEWART, G. S. & 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.

SRIVATSAN, A., BOWEN, N. & KOLODNER, R. D. 2014. Mispair-specific Recruitment of the Mlh1-Pms1 Complex Identifies Repair Substrates of the Saccharomyces cerevisiae Msh2-Msh3 Complex*. *Journal of Biological Chemistry,* 289**,** 9352--9364.

ST-DENIS, N., GABRIEL, M., TUROWEC, J. P., GLOOR, G. B., LI, S. S. C., GINGRAS, A.-C. & LITCHFIELD, D. W. 2015. Systematic investigation of hierarchical phosphorylation by protein kinase CK2. *Journal of Proteomics,* 118**,** 49--62.

STARESINCIC, L., FAGBEMI, A. F., ENZLIN, J. H., GOURDIN, A. M., WIJGERS, N., DUNAND-SAUTHIER, I., GIGLIA-MARI, G., CLARKSON, S. G., VERMEULEN, W. & SCHÄRER, O. D. 2009. Coordination of dual incision and repair synthesis in human nucleotide excision repair. *The EMBO Journal,* 28**,** 1111--1120.

STARITA, L. M., AHITUV, N., DUNHAM, M. J., KITZMAN, J. O., ROTH, F. P., SEELIG, G., SHENDURE, J. & FOWLER, D. M. 2017. Variant Interpretation: Functional Assays to the Rescue. *The American Journal of Human Genetics,* 101**,** 315--325.

STARITA, L. M., ISLAM, M. M., BANERJEE, T., ADAMOVICH, A. I., GULLINGSRUD, J., FIELDS, S., SHENDURE, J. & PARVIN, J. D. 2018. A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. *The American Journal of Human Genetics,* 103**,** 498--508.

STATES, J. C., MCDUFFIE, E. R., MYRAND, S. P., MCDOWELL, M. & CLEAVER, J. E. 1998. Distribution of mutations in the human xeroderma pigmentosum group A gene and their relationships to the functional regions of the DNA damage recognition protein. *Human Mutation,* 12**,** 103--113.

STAVNEZER, J., GUIKEMA, J. E. J. & SCHRADER, C. E. 2008. Mechanism and Regulation of Class Switch Recombination. *Annual Review of Immunology,* 26**,** 261--292.

STEFL, S., NISHI, H., PETUKH, M., PANCHENKO, A. R. & ALEXOV, E. 2013. Molecular Mechanisms of Disease-Causing Missense Mutations. *Journal of Molecular Biology,* 425**,** 3919--3936.

STEIN, A., FOWLER, D. M., HARTMANN-PETERSEN, R. & LINDORFF-LARSEN, K. 2019. Biophysical and Mechanistic Models for Disease-Causing Protein Variants. *Trends in Biochemical Sciences,* 44**,** 575--588.

STENSON, P. D., BALL, E. V., MORT, M., PHILLIPS, A. D., SHAW, K. & COOPER, D. N. 2012. The Human Gene Mutation Database (HGMD) and Its Exploitation in the Fields of Personalized Genomics and Molecular Evolution. *Current Protocols in Bioinformatics,* 39**,** 1.13.1--1.13.20.

STOFFEL, E. M. & CHITTENDEN, A. 2010. Genetic Testing for Hereditary Colorectal Cancer: Challenges in Identifying, Counseling, and Managing High-Risk Patients. *Gastroenterology,* 139**,** 1436--1441.e1.

STONE, E. A. & SIDOW, A. 2005. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Research,* 15**,** 978--986.

STUCKI, M., CLAPPERTON, J. A., MOHAMMAD, D., YAFFE, M. B., SMERDON, S. J. & JACKSON, S. P. 2005. MDC1 Directly Binds Phosphorylated Histone H2AX to Regulate Cellular Responses to DNA Double-Strand Breaks. *Cell,* 123**,** 1213--1226.

SU, S. S. & MODRICH, P. 1986. Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. *Proceedings of the National Academy of Sciences,* 83**,** 5057--5061.

SUGASAWA, K., NG, J. M. Y., MASUTANI, C., IWAI, S., SPEK, P. J. V. D., EKER, A. P. M., HANAOKA, F., BOOTSMA, D. & HOEIJMAKERS, J. H. J. 1998. Xeroderma Pigmentosum Group C Protein Complex Is the Initiator of Global Genome Nucleotide Excision Repair. *Molecular Cell,* 2**,** 223--232.

SUGITANI, N., SIVLEY, R. M., PERRY, K. E., CAPRA, J. A. & CHAZIN, W. J. 2016. XPA: A key scaffold for human nucleotide excision repair. *DNA Repair,* 44**,** 123--135.

SUGITANI, N., VOEHLER, M. W., ROH, M. S., TOPOLSKA-WOŚ, A. M. & CHAZIN, W. J. 2017. Analysis of DNA binding by human factor xeroderma pigmentosum complementation group A (XPA) provides insight into its interactions with nucleotide excision repair substrates. *Journal of Biological Chemistry,* 292**,** 16847--16857.

SUN, J., LEE, K.-J., DAVIS, A. J. & CHEN, D. J. 2012. Human Ku70/80 Protein Blocks Exonuclease 1-mediated DNA Resection in the Presence of Human Mre11 or Mre11/Rad50 Protein Complex*. *Journal of Biological Chemistry,* 287**,** 4936--4945.

SUN, W., NANDI, S., OSMAN, F., AHN, J. S., JAKOVLESKA, J., LORENZ, A. & WHITBY, M. C. 2008. The FANCM Ortholog Fml1 Promotes Recombination at Stalled Replication Forks and Limits Crossing Over during DNA Double-Strand Break Repair. *Molecular Cell,* 32**,** 118--128.

SWANN, P. F., WATERS, T. R., MOULTON, D. C., XU, Y.-Z., ZHENG, Q., EDWARDS, M. & MACE, R. 1996. Role of Postreplicative DNA Mismatch Repair in the Cytotoxic Action of Thioguanine. *Science,* 273**,** 1109--1111.

SWUEC, P., RENAULT, L., BORG, A., SHAH, F., MURPHY, V. J., TWEST, S. V., SNIJDERS, A. P., DEANS, A. J. & COSTA, A. 2017. The FA Core Complex Contains a Homo-dimeric Catalytic Module for the Symmetric Mono-ubiquitination of FANCI-FANCD2. *Cell Reports,* 18**,** 611--623.

SZANKASI, P. & SMITH, G. R. 1995. A Role for Exonuclease I from S. pombe in Mutation Avoidance and Mismatch Correction. *Science,* 267**,** 1166--1169.

TAKAHASHI, M., SHIMODAIRA, H., ANDREUTTI-ZAUGG, C., IGGO, R., KOLODNER, R. D. & ISHIOKA, C. 2007. Functional Analysis of Human MLH1 Variants Using Yeast and In vitro Mismatch Repair Assays. *Cancer Research,* 67**,** 4595--4604.

TAKAHASHI, Y., ENDO, Y., SUGIYAMA, Y., INOUE, S., IIJIMA, M., TOMITA, Y., KURU, S., TAKIGAWA, M. & MORIWAKI, S. 2010. XPA Gene Mutations Resulting in Subtle Truncation of Protein in Xeroderma Pigmentosum Group A Patients with Mild Skin Symptoms. *Journal of Investigative Dermatology,* 130**,** 2481--2488.

TAM, V., PATEL, N., TURCOTTE, M., BOSSÉ, Y., PARÉ, G. & MEYRE, D. 2019. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics,* 20**,** 467--484.

TAN, T. Y., LUNKE, S., CHONG, B., PHELAN, D., FANJUL-FERNANDEZ, M., MARUM, J. E., KUMAR, V. S., STARK, Z., YEUNG, A., BROWN, N. J., STUTTERD, C., DELATYCKI, M. B., SADEDIN, S., MARTYN, M., GORANITIS, I., THORNE, N., GAFF, C. L. & WHITE, S. M. 2019. A head-to-head evaluation of the diagnostic efficacy and costs of trio versus singleton exome sequencing analysis. *European Journal of Human Genetics,* 27**,** 1791--1799.

TAN, W., TWEST, S. V., LEIS, A., BYTHELL-DOUGLAS, R., MURPHY, V. J., SHARP, M., PARKER, M. W., CRISMANI, W. & DEANS, A. J. 2020. Monoubiquitination by the human Fanconi anemia core complex clamps FANCI:FANCD2 on DNA in filamentous arrays. *eLife,* 9**,** e54128.

TANAKA, K., MIURA, N., SATOKATA, I., MIYAMOTO, I., YOSHIDA, M. C., SATOH, Y., KONDO, S., YASUI, A., OKAYAMA, H. & OKADA, Y. 1990. Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zinc-finger domain. *Nature,* 348**,** 73--76.

TANAKA, K., SATOKATA, I., OGITA, Z., UCHIDA, T. & OKADA, Y. 1989. Molecular cloning of a mouse DNA repair gene that complements the defect of group-A xeroderma pigmentosum. *Proceedings of the National Academy of Sciences,* 86**,** 5512--5516.

TAPIAS, A., AURIOL, J., FORGET, D., ENZLIN, J. H., SCHÄRER, O. D., COIN, F., COULOMBE, B. & EGLY, J.-M. 2004. Ordered Conformational Changes in Damaged DNA Induced by Nucleotide Excision Repair Factors*. *Journal of Biological Chemistry,* 279**,** 19074--19083.

TARSOUNAS, M. & SUNG, P. 2020. The antitumorigenic roles of BRCA1–BARD1 in DNA repair and replication. *Nature Reviews Molecular Cell Biology,* 21**,** 284--299.

TAVTIGIAN, S. V., DEFFENBAUGH, A. M., YIN, L., JUDKINS, T., SCHOLL, T., SAMOLLOW, P. B., SILVA, D. D., ZHARKIKH, A. & THOMAS, A. 2006. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *Journal of Medical Genetics,* 43**,** 295.

TAYLOR, E. R. & MCGOWAN, C. H. 2008. Cleavage mechanism of human Mus81–Eme1 acting on Holliday-junction structures. *Proceedings of the National Academy of Sciences,* 105**,** 3757--3762.

TEJPAR, S., SARIDAKI, Z., DELORENZI, M., BOSMAN, F. & ROTH, A. D. 2011. Microsatellite Instability, Prognosis and Drug Sensitivity of Stage II and III Colorectal Cancer: More Complexity to the Puzzle. *JNCI: Journal of the National Cancer Institute,* 103**,** 841--844.

TERASAWA, M., SHINOHARA, A. & SHINOHARA, M. 2014. Canonical Non-Homologous End Joining in Mitosis Induces Genome Instability and Is Suppressed by M-phase-Specific Phosphorylation of XRCC4. *PLoS Genetics,* 10**,** e1004563.

THOMPSON, B. A., GREENBLATT, M. S., VALLEE, M. P., HERKERT, J. C., TESSEREAU, C., YOUNG, E. L., ADZHUBEY, I. A., LI, B., BELL, R., FENG, B., MOONEY, S. D., RADIVOJAC, P., SUNYAEV, S. R., FREBOURG, T., HOFSTRA, R. M. W., SIJMONS, R. H., BOUCHER, K., THOMAS, A., GOLDGAR, D. E., SPURDLE, A. B. & TAVTIGIAN, S. V. 2013. Calibration of Multiple In Silico Tools for Predicting Pathogenicity of Mismatch Repair Gene Missense Substitutions. *Human Mutation,* 34**,** 255--265.

THOMPSON, B. A., SPURDLE, A. B., PLAZZER, J.-P., GREENBLATT, M. S., AKAGI, K., AL-MULLA, F., BAPAT, B., BERNSTEIN, I., CAPELLÁ, G., DUNNEN, J. T. D., SART, D. D., FABRE, A., FARRELL, M. P., FARRINGTON, S. M., FRAYLING, I. M., FREBOURG, T., GOLDGAR, D. E., HEINEN, C. D., HOLINSKI-FEDER, E., KOHONEN-CORISH, M., ROBINSON, K. L., LEUNG, S. Y., MARTINS, A., MOLLER, P., MORAK, M., NYSTROM, M., PELTOMAKI, P., PINEDA, M., QI, M., RAMESAR, R., RASMUSSEN, L. J., ROYER-POKORA, B., SCOTT, R. J., SIJMONS, R., TAVTIGIAN, S. V., TOPS, C. M., WEBER, T., WIJNEN, J., WOODS, M. O., MACRAE, F., GENUARDI, M., CASTILLEJO, A., SEXTON, A., CHAN, A. K. W., VIEL, A., BLANCO, A., FRENCH, A., LANER, A., WAGNER, A., OUWELAND, A. V. D., MENSENKAMP, A., PAYÁ, A., BETZ, B., REDEKER, B., SMITH, B., ESPENSCHIED, C., CUMMINGS, C., ENGEL, C., FORNES, C., VALENZUELA, C., ALENDA, C., BUCHANAN, D., BARANA, D., KONSTANTINOVA, D., CAIRNS, D., GLASER, E., SILVA, F., LALLOO, F., CRUCIANELLI, F., HOGERVORST, F., CASEY, G., TOMLINSON, I., BLANCO, I., VILLAR, I. L., GARCIA-PLANELLS, J., BIGLER, J., SHIA, J., MARTINEZ-LOPEZ, J., GILLE, J. J. P., HOPPER, J., POTTER, J., SOTO, J. L., KANTELINEN, J., ELLIS, K., MANN, K., VARESCO, L., ZHANG, L., MARCHAND, L. L., MARAFIE, M. J., NORDLING, M., TIBILETTI, M. G., KAHAN, M. A., LIGTENBERG, M., CLENDENNING, M., JENKINS, M., SPEEVAK, M., DIGWEED, M., KLOOR, M., HITCHINS, M., MYERS, M., et al. 2014. Application of a 5-tiered scheme for standardized

classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nature Genetics,* 46**,** 107--115.

THORN, C. F., OSHIRO, C., MARSH, S., HERNANDEZ-BOUSSARD, T., MCLEOD, H., KLEIN, T. E. & ALTMAN, R. B. 2011. Doxorubicin pathways. *Pharmacogenetics and Genomics,* 21**,** 440--446.

THORSLUND, T., MCILWRAITH, M. J., COMPTON, S. A., LEKOMTSEV, S., PETRONCZKI, M., GRIFFITH, J. D. & 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. & MAILAND, N. 2015. Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage. *Nature,* 527**,** 389--393.

THOUVENOT, P., YAMIN, B. B., FOURRIÈRE, L., LESCURE, A., BOUDIER, T., NERY, E. D., CHAUCHEREAU, A., GOLDGAR, D. E., HOUDAYER, C., STOPPA-LYONNET, D., NICOLAS, A. & MILLOT, G. A. 2016. Functional Assessment of Genetic Variants with Outcomes Adapted to Clinical Decision-Making. *PLoS Genetics,* 12**,** e1006096.

TISHKOFF, D. X., BOERGER, A. L., BERTRAND, P., FILOSI, N., GAIDA, G. M., KANE, M. F. & KOLODNER, R. D. 1997. Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2. *Proceedings of the National Academy of Sciences,* 94**,** 7487--7492.

TITUS, T. A., SELVIG, D. R., QIN, B., WILSON, C., STARKS, A. M., ROE, B. A. & POSTLETHWAIT, J. H. 2006. The Fanconi anemia gene network is conserved from zebrafish to human. *Gene,* 371**,** 211--223.

TIWARI, V. & WILSON, D. M. 2019. DNA Damage and Associated DNA Repair Defects in Disease and Premature Aging. *The American Journal of Human Genetics,* 105**,** 237--257.

TOLEDO, L., NEELSEN, K. J. & LUKAS, J. 2017. Replication Catastrophe: When a Checkpoint Fails because of Exhaustion. *Molecular Cell,* 66**,** 735--749.

TOLEDO, LUIS I., ALTMEYER, M., RASK, M.-B., LUKAS, C., LARSEN, DORTHE H., POVLSEN, LOU K., BEKKER-JENSEN, S., MAILAND, N., BARTEK, J. & LUKAS, J. 2013. ATR Prohibits Replication Catastrophe by Preventing Global Exhaustion of RPA. *Cell,* 155**,** 1088--1103.

TOMER, G., BUERMEYER, A. B., NGUYEN, M. M. & LISKAY, R. M. 2002. Contribution of Human Mlh1 and Pms2 ATPase Activities to DNA Mismatch Repair*. *Journal of Biological Chemistry,* 277**,** 21801--21809.

TOMIDA, J., TAKATA, K. I., BHETAWAL, S., PERSON, M. D., CHAO, H. P., TANG, D. G. & WOOD, R. D. 2018. FAM35A associates with REV7 and modulates DNA damage responses of normal and BRCA1-defective cells. *The EMBO Journal,* 37**,** e99543.

TOMINAGA, Y., LI, C., WANG, R.-H. & DENG, C.-X. 2006. Murine Wee1 Plays a Critical Role in Cell Cycle Regulation and Pre-Implantation Stages of Embryonic Development. *International Journal of Biological Sciences,* 2**,** 161--170.

TOPATANA, W., JUENGPANICH, S., LI, S., CAO, J., HU, J., LEE, J., SULIYANTO, K., MA, D., ZHANG, B., CHEN, M. & CAI, X. 2020. Advances in synthetic lethality for cancer therapy: cellular mechanism and clinical translation. *Journal of Hematology \& Oncology,* 13**,** 118.

TOPOLSKA-WOŚ, A. M., SUGITANI, N., CORDOBA, J. J., LE MEUR, K. V., LE MEUR, R. A., KIM, H. S., YEO, J.-E., ROSENBERG, D., HAMMEL, M., SCHÄRER, O. D. & CHAZIN, W. J. 2020. A key interaction with RPA orients XPA in NER complexes. *Nucleic Acids Research,* 48**,** 2173--2188.

TOURNIER, I., VEZAIN, M., MARTINS, A., CHARBONNIER, F., BAERT-DESURMONT, S., OLSCHWANG, S., WANG, Q., BUISINE, M. P., SORET, J., TAZI, J., FRÉBOURG, T. & TOSI, M. 2008. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Human Mutation,* 29**,** 1412--1424.

TRAN, H. T., GORDENIN, D. A. & RESNICK, M. A. 1999. The 3'→5' Exonucleases of DNA Polymerases δ and ε and the 5'→3' Exonuclease Exo1 Have Major Roles in Postreplication Mutation Avoidance in Saccharomyces cerevisiae. *Molecular and Cellular Biology,* 19**,** 2000--2007.

TRAN, P. T., ERDENIZ, N., SYMINGTON, L. S. & LISKAY, R. M. 2004. EXO1-A multi-tasking eukaryotic nuclease. *DNA Repair,* 3**,** 1549--1559.

TROJAN, J., ZEUZEM, S., RANDOLPH, A., HEMMERLE, C., BRIEGER, A., RAEDLE, J., PLOTZ, G., JIRICNY, J. & MARRA, G. 2002. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. *Gastroenterology,* 122**,** 211--219.

TSAI, C. J., KIM, S. A. & CHU, G. 2007. Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proceedings of the National Academy of Sciences,* 104**,** 7851--7856.

TSENG, H.-C., TSAI, M.-H., CHIU, C.-F., WANG, C.-H., CHANG, N.-W., HUANG, C.-Y., TSAI, C.-W., LIANG, S.-Y., WANG, C.-L. & BAU, D.-T. 2008. Association of XRCC4 codon 247 polymorphism with oral cancer susceptibility in Taiwan. *Anticancer research,* 28**,** 1687--91.

TSODIKOV, O. V., IVANOV, D., ORELLI, B., STARESINCIC, L., SHOSHANI, I., OBERMAN, R., SCHÄRER, O. D., WAGNER, G. & ELLENBERGER, T. 2007. Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *The EMBO Journal,* 26**,** 4768--4776.

TSUGE, K. & SHIMAMOTO, A. 2022. Research on Werner Syndrome: Trends from Past to Present and Future Prospects. *Genes,* 13**,** 1802.

TU, X., KAHILA, M. M., ZHOU, Q., YU, J., KALARI, K. R., WANG, L., HARMSEN, W. S., YUAN, J., BOUGHEY, J. C., GOETZ, M. P., SARKARIA, J. N., LOU, Z. & MUTTER, R. W. 2018. ATR inhibition is a promising radiosensitizing strategy for triple negative breast cancer. *Molecular Cancer Therapeutics,* 17**,** molcanther.0470.2018.

TUBBS, A. & NUSSENZWEIG, A. 2017. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell,* 168**,** 644--656.

TURNBULL, C., SCOTT, R. H., THOMAS, E., JONES, L., MURUGAESU, N., PRETTY, F. B., HALAI, D., BAPLE, E., CRAIG, C., HAMBLIN, A., HENDERSON, S., PATCH, C., O'NEILL, A., DEVEREAU, A., SMITH, K., MARTIN, A. R., SOSINSKY, A., MCDONAGH, E. M., SULTANA, R., MUELLER, M., SMEDLEY, D., TOMS, A., DINH, L., FOWLER, T., BALE, M., HUBBARD, T., RENDON, A., HILL, S., CAULFIELD, M. J. & PROJECT, G. 2018. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *Bmj,* 361**,** k1687.

TWEST, S. V., MURPHY, V. J., HODSON, C., TAN, W., SWUEC, P., O'ROURKE, J. J., HEIERHORST, J., CRISMANI, W. & DEANS, A. J. 2017. Mechanism of Ubiquitination and Deubiquitination in the Fanconi Anemia Pathway. *Molecular Cell,* 65**,** 247--259.

UDESHI, N. D., SVINKINA, T., MERTINS, P., KUHN, E., MANI, D. R., QIAO, J. W. & CARR, S. A. 2013. Refined Preparation and Use of Anti-diglycine Remnant (K-ε-GG) Antibody Enables Routine Quantification of 10,000s of Ubiquitination Sites in Single Proteomics Experiments*. *Molecular \& Cellular Proteomics,* 12**,** 825--831.

UFFELMANN, E., HUANG, Q. Q., MUNUNG, N. S., VRIES, J. D., OKADA, Y., MARTIN, A. R., MARTIN, H. C., LAPPALAINEN, T. & POSTHUMA, D. 2021. Genome-wide association studies. *Nature Reviews Methods Primers,* 1**,** 59.

UMAR, A., BOLAND, C. R., TERDIMAN, J. P., SYNGAL, S., CHAPELLE, A. D. L., RÜSCHOFF, J., FISHEL, R., LINDOR, N. M., BURGART, L. J., HAMELIN, R., HAMILTON, S. R., HIATT, R. A., JASS, J., LINDBLOM, A., LYNCH, H. T., PELTOMAKI, P., RAMSEY, S. D., RODRIGUEZ-BIGAS, M. A., VASEN, H. F. A., HAWK, E. T., BARRETT, J. C., FREEDMAN, A. N. & SRIVASTAVA, S. 2004. Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. *JNCI: Journal of the National Cancer Institute,* 96**,** 261--268.

UMAR, A., BUERMEYER, A. B., SIMON, J. A., THOMAS, D. C., CLARK, A. B., LISKAY, R. M. & KUNKEL, T. A. 1996. Requirement for PCNA in DNA Mismatch Repair at a Step Preceding DNA Resynthesis. *Cell,* 87**,** 65--73.

VAITSIANKOVA, A., BURDOVA, K., SOBOL, M., GAUTAM, A., BENADA, O., HANZLIKOVA, H. & CALDECOTT, K. W. 2022. PARP inhibition impedes the maturation of nascent DNA strands during DNA replication. *Nature Structural \& Molecular Biology*, 1--10.

VALLON-CHRISTERSSON, J., CAYANAN, C., HARALDSSON, K., LOMAN, N., BERGTHORSSON, JON T., BRØNDUM-NIELSEN, K., GERDES, A.-M., MØLLER, P. L., KRISTOFFERSSON, U., OLSSON, H. K., BORG, A. K. & MONTEIRO, A. N. A. 2001. Functional analysis of BRCA1 C-terminal missense mutations identified in breast and ovarian cancer families. *Human Molecular Genetics,* 10**,** 353--360.

VAZ, F., HANENBERG, H., SCHUSTER, B., BARKER, K., WIEK, C., ERVEN, V., NEVELING, K., ENDT, D., KESTERTON, I., AUTORE, F., FRATERNALI, F., FREUND, M., HARTMANN, L., GRIMWADE, D., ROBERTS, R. G., SCHAAL, H., MOHAMMED, S., RAHMAN, N., SCHINDLER, D. & MATHEW, C. G. 2010. Mutation of the RAD51C gene in a Fanconi anemia–like disorder. *Nature Genetics,* 42**,** 406--409.

VEIGL, M. L., KASTURI, L., OLECHNOWICZ, J., MA, A., LUTTERBAUGH, J. D., PERIYASAMY, S., LI, G.-M., DRUMMOND, J., MODRICH, P. L., SEDWICK, W. D. & MARKOWITZ, S. D. 1998. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proceedings of the National Academy of Sciences,* 95**,** 8698--8702.

VERMA, P., ZHOU, Y., CAO, Z., DERASKA, P. V., DEB, M., ARAI, E., LI, W., SHAO, Y., PUENTES, L., LI, Y., PATANKAR, S., MACH, R. H., FARYABI, R. B., SHI, J. & GREENBERG, R. A. 2021. ALC1 links chromatin accessibility to PARP inhibitor response in homologous recombination-deficient cells. *Nature cell biology,* 23**,** 160--171.

VERMEULEN, W. & FOUSTERI, M. 2013. Mammalian Transcription-Coupled Excision Repair. *Cold Spring Harbor Perspectives in Biology,* 5**,** a012625.

VERNOLE, P., MUZI, A., VOLPI, A., TERRINONI, A., DORIO, A. S., TENTORI, L., SHAH, G. M. & GRAZIANI, G. 2011. DNA damage induced by endogenous aldehydes: Current state of knowledge. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 711**,** 13--27.

VETRO, A., IASCONE, M., LIMONGELLI, I., AMEZIANE, N., GANA, S., MINA, E. D., GIUSSANI, U., CICCONE, R., FORLINO, A., PEZZOLI, L., ROOIMANS, M. A., ESSEN, A. J., MESSA, J., RIZZUTI, T., BIANCHI, P., DORSMAN, J., WINTER, J. P., LALATTA, F. & ZUFFARDI, O. 2015. Loss-of-Function FANCL Mutations Associate with Severe Fanconi Anemia Overlapping the VACTERL Association. *Human Mutation,* 36**,** 562--568.

VIHINEN, M. 2021. Functional effects of protein variants. *Biochimie,* 180**,** 104--120.

VIRGILIO, M. D., CALLEN, E., YAMANE, A., ZHANG, W., JANKOVIC, M., GITLIN, A. D., FELDHAHN, N., RESCH, W., OLIVEIRA, T. Y., CHAIT, B. T., NUSSENZWEIG, A., CASELLAS, R., ROBBIANI, D. F. & NUSSENZWEIG, M. C. 2013. Rif1 Prevents Resection of DNA Breaks and Promotes Immunoglobulin Class Switching. *Science,* 339**,** 711--715.

VIRTS, E. L., JANKOWSKA, A., MACKAY, C., GLAAS, M. F., WIEK, C., KELICH, S. L., LOTTMANN, N., KENNEDY, F. M., MARCHAL, C., LEHNERT, E., SCHARF, R. E., DUFOUR, C., LANCIOTTI, M., FARRUGGIA, P., SANTORO, A., SAVASAN, S., SCHECKENBACH, K., SCHIPPER, J., WAGENMANN, M., LEWIS, T., LEFFAK, M., FARLOW, J. L., FOROUD, T. M., HONISCH, E., NIEDERACHER, D., CHAKRABORTY, S. C., VANCE, G. H., PRUSS, D., TIMMS, K. M., LANCHBURY, J. S., ALPI, A. F. & HANENBERG, H. 2015. AluY-mediated germline deletion, duplication and somatic stem cell reversion in UBE2T defines a new subtype of Fanconi anemia. *Human Molecular Genetics,* 24**,** 5093--5108.

VISSCHER, P. M., WRAY, N. R., ZHANG, Q., SKLAR, P., MCCARTHY, M. I., BROWN, M. A. & YANG, J. 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics,* 101**,** 5--22.

VISWANATHAN, M., BURDETT, V., BAITINGER, C., MODRICH, P. & LOVETT, S. T. 2001. Redundant Exonuclease Involvement in Escherichia coli Methyl-directed Mismatch Repair*. *Journal of Biological Chemistry,* 276**,** 31053--31058.

VOGELSANG, M., COMINO, A., ZUPANEC, N., HUDLER, P. & KOMEL, R. 2009. Assessing pathogenicity of MLH1 variants by co-expression of human MLH1 and PMS2genes in yeast. *BMC Cancer,* 9**,** 382.

VOLKER, M., MONÉ, M. J., KARMAKAR, P., HOFFEN, A. V., SCHUL, W., VERMEULEN, W., HOEIJMAKERS, J. H. J., DRIEL, R. V., ZEELAND, A. A. V. & MULLENDERS, L. H. F. 2001. Sequential Assembly of the Nucleotide Excision Repair Factors In Vivo. *Molecular Cell,* 8**,** 213--224.

WAGLE, N., BERGER, M. F., DAVIS, M. J., BLUMENSTIEL, B., DEFELICE, M., POCHANARD, P., DUCAR, M., HUMMELEN, P. V., MACCONAILL, L. E., HAHN, W. C., MEYERSON, M., GABRIEL, S. B. & GARRAWAY, L. A. 2012. High-Throughput Detection of Actionable Genomic Alterations in Clinical Tumor Samples by Targeted, Massively Parallel Sequencing. *Cancer Discovery,* 2**,** 82--93.

WAGNER, R. & MESELSON, M. 1976. Repair tracts in mismatched DNA heteroduplexes. *Proceedings of the National Academy of Sciences,* 73**,** 4135--4139.

WAKASUGI, M., KASASHIMA, H., FUKASE, Y., IMURA, M., IMAI, R., YAMADA, S., CLEAVER, J. E. & MATSUNAGA, T. 2009. Physical and functional interaction between DDB and XPA in nucleotide excision repair. *Nucleic Acids Research,* 37**,** 516--525.

WAKASUGI, M., SHIMIZU, M., MORIOKA, H., LINN, S., NIKAIDO, O. & MATSUNAGA, T. 2001. Damaged DNA-binding Protein DDB Stimulates the Excision of Cyclobutane Pyrimidine Dimers in Vitro in Concert with XPA and Replication Protein A*. *Journal of Biological Chemistry,* 276**,** 15434--15440.

WALKER, J. R., CORPINA, R. A. & GOLDBERG, J. 2001. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature,* 412**,** 607--614.

WALKER, R., MAHMOOD, K., JOO, J. E., CLENDENNING, M., GEORGESON, P., COMO, J., JOSELAND, S., PRESTON, S. G., ANTILL, Y., AUSTIN, R., BOUSSIOUTAS, A., BOWMAN, M., BURKE, J., CAMPBELL, A., DANESHVAR, S., EDWARDS, E., GLEESON, M., GOODWIN, A., HARRIS, M. T., HENDERSON, A., HIGGINS, M., HOPPER, J. L., HUTCHINSON, R. A., IP, E., ISBISTER, J., KASEM, K., MARFAN, H., MILNES, D., NG, A., NICHOLS, C., O'CONNELL, S., PACHTER, N., POPE, B. J., POPLAWSKI, N., RAGUNATHAN, A., SMYTH, C., SPIGELMAN, A., STOREY, K., SUSMAN, R., TAYLOR, J. A., WARWICK, L., WILDING, M., WILLIAMS, R., WIN, A. K., WALSH, M. D., MACRAE, F. A., JENKINS, M. A., ROSTY, C., WINSHIP, I. M., BUCHANAN, D. D. & AUSTRALIA, F. T. F. C. C. O. 2023. A tumor focused approach to resolving the etiology of DNA mismatch repair deficient tumors classified as suspected Lynch syndrome. *Journal of Translational Medicine,* 21**,** 282.

WANAT, J. J., SINGH, N. & ALANI, E. 2007. The effect of genetic background on the function of Saccharomyces cerevisiae mlh1 alleles that correspond to HNPCC missense mutations. *Human Molecular Genetics,* 16**,** 445--452.

WANG, D., LI, J., WANG, Y. & WANG, E. 2022a. A comparison on predicting functional impact of genomic variants. *NAR Genomics and Bioinformatics,* 4**,** lqab122.

WANG, H., SHI, L. Z., WONG, C. C. L., HAN, X., HWANG, P. Y.-H., TRUONG, L. N., ZHU, Q., SHAO, Z., CHEN, D. J., BERNS, M. W., YATES, J. R., CHEN, L. & WU, X. 2013. The Interaction of CtIP and Nbs1 Connects CDK and ATM to Regulate HR–Mediated Double-Strand Break Repair. *PLoS Genetics,* 9**,** e1003277.

WANG, H., ZHAO, Y., WANG, F., ZHU, X., LUO, N., SUN, T., QI, C. & LI, X. 2022b. Pathogenic somatic alterations of DDR genes in lung cancer are significantly different from germline mutations and are associated with more unstable genomes. *Journal of Translational Medicine,* 20**,** 408.

WANG, M., CHEN, S. & AO, D. 2021. Targeting DNA repair pathway in cancer: Mechanisms and clinical application. *MedComm,* 2**,** 654--691.

WANG, R., WANG, S., DHAR, A., PERALTA, C. & PAVLETICH, N. P. 2020a. DNA clamp function of the monoubiquitinated Fanconi anaemia ID complex. *Nature,* 580**,** 278--282.

WANG, S., MIN, Z., JI, Q., GENG, L., SU, Y., LIU, Z., HU, H., WANG, L., ZHANG, W., SUZUIKI, K., HUANG, Y., ZHANG, P., TANG, T.-S., QU, J., YU, Y., LIU, G.-H. & QIAO, J. 2020b. Rescue of premature aging defects in Cockayne syndrome stem cells by CRISPR/Cas9-mediated gene correction. *Protein \& Cell,* 11**,** 1--22.

WANG, X., WEI, X., THIJSSEN, B., DAS, J., LIPKIN, S. M. & YU, H. 2012. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nature Biotechnology,* 30**,** 159--164.

WATERS, T. R. & SWANN, P. F. 1997. Cytotoxic Mechanism of 6-Thioguanine: hMutSα, the Human Mismatch Binding Heterodimer, Binds to DNA Containing S 6-Methylthioguanine †. *Biochemistry,* 36**,** 2501--2506.

WEEGEN, Y. V. D., GOLAN-BERMAN, H., MEVISSEN, T. E. T., APELT, K., GONZÁLEZ-PRIETO, R., GOEDHART, J., HEILBRUN, E. E., VERTEGAAL, A. C. O., HEUVEL, D. V. D., WALTER, J. C., ADAR, S. & LUIJSTERBURG, M. S. 2020. The cooperative action of CSB, CSA, and UVSSA target TFIIH to DNA damage-stalled RNA polymerase II. *Nature Communications,* 11**,** 2104.

WEI, K., CLARK, A. B., WONG, E., KANE, M. F., MAZUR, D. J., PARRIS, T., KOLAS, N. K., RUSSELL, R., HOU, H., KNEITZ, B., YANG, G., KUNKEL, T. A., KOLODNER, R. D., COHEN, P. E. & EDELMANN, W. 2003. Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility. *Genes \& Development,* 17**,** 603--614.

WELSH, K. M., LU, A. L., CLARK, S. & MODRICH, P. 1987. Isolation and characterization of the Escherichia coli mutH gene product. *Journal of Biological Chemistry,* 262**,** 15624--15629.

WIETMARSCHEN, N. V., SRIDHARAN, S., NATHAN, W. J., TUBBS, A., CHAN, E. M., CALLEN, E., WU, W., BELINKY, F., TRIPATHI, V., WONG, N., FOSTER, K., NOORBAKHSH, J., GARIMELLA, K., CRUZ-MIGONI, A., SOMMERS, J. A., HUANG, Y., BORAH, A. A., SMITH, J. T., KALFON, J., KESTEN, N., FUGGER, K., WALKER, R. L., DOLZHENKO, E., EBERLE, M. A., HAYWARD, B. E., USDIN, K., FREUDENREICH, C. H., BROSH, R. M., WEST, S. C., MCHUGH, P. J., MELTZER, P. S., BASS, A. J. & NUSSENZWEIG, A. 2020. Repeat expansions confer WRN dependence in microsatellite-unstable cancers. *Nature,* 586**,** 292--298.

WILDENBERG, J. & MESELSON, M. 1975. Mismatch repair in heteroduplex DNA. *Proceedings of the National Academy of Sciences,* 72**,** 2202--2206.

WILKINSON, KEVIN A. & HENLEY, JEREMY M. 2010. Mechanisms, regulation and consequences of protein SUMOylation. *Biochemical Journal,* 428**,** 133--145.

WILLIAMS, R. S., CHASMAN, D. I., HAU, D. D., HUI, B., LAU, A. Y. & GLOVER, J. N. M. 2003. Detection of Protein Folding Defects Caused by BRCA1-BRCT Truncation and Missense Mutations*. *Journal of Biological Chemistry,* 278**,** 53007--53016.

WILTSHIRE, T., DUCY, M., FOO, T. K., HU, C., LEE, K. Y., NAGARAJ, A. B., RODRIGUE, A., GOMES, T. T., SIMARD, J., MONTEIRO, A. N. A., XIA, B., CARVALHO, M. A., MASSON, J.-Y. & COUCH, F. J. 2020. Functional characterization of 84 PALB2 variants of uncertain significance. *Genetics in Medicine,* 22**,** 622--632.

WIMMER, K., KRATZ, C. P., VASEN, H. F. A., CARON, O., COLAS, C., ENTZ-WERLE, N., GERDES, A.-M., GOLDBERG, Y., ILENCIKOVA, D., MULERIS, M., DUVAL, A., LAVOINE, N., RUIZ-PONTE, C., SLAVC, I., BURKHARDT, B., BRUGIERES, L. & CMMRD, E. U. C. C. F. 2014. Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'Care for CMMRD' (C4CMMRD). *Journal of Medical Genetics,* 51**,** 355.

WINKLER, G. S., ARAÚJO, S. J., FIEDLER, U., VERMEULEN, W., COIN, F., EGLY, J.-M., HOEIJMAKERS, J. H. J., WOOD, R. D., TIMMERS, H. T. M. & WEEDA, G. 2000. TFIIH with Inactive XPD Helicase Functions in Transcription Initiation but Is Defective in DNA Repair*. *Journal of Biological Chemistry,* 275**,** 4258--4266.

WINTERBOURN, C. C. 2008. Reconciling the chemistry and biology of reactive oxygen species. *Nature Chemical Biology,* 4**,** 278--286.

WONG, A. K. C., ORMONDE, P. A., PERO, R., CHEN, Y., LIAN, L., SALADA, G., BERRY, S., LAWRENCE, Q., DAYANANTH, P., HA, P., TAVTIGIAN, S. V., TENG, D. H. F. & BARTEL, P. L. 1998. Characterization of a carboxy-terminal BRCA1 interacting protein. *Oncogene,* 17**,** 2279--2285.

WOODBINE, L., GENNERY, A. R. & JEGGO, P. A. 2014. The clinical impact of deficiency in DNA non-homologous end-joining. *DNA Repair,* 16**,** 84--96.

WRIGHT, W. D., SHAH, S. S. & HEYER, W.-D. 2018. Homologous recombination and the repair of DNA double-strand breaks. *Journal of Biological Chemistry,* 293**,** 10524--10535.

WU, H., ZENG, H., LAM, R., TEMPEL, W., KERR, I. D. & MIN, J. 2015. Structure of the human MLH1 N-terminus: implications for predisposition to Lynch syndrome. *Acta Crystallographica Section F: Structural Biology Communications,* 71**,** 981--985.

WU, L., BACHRATI, C. Z., OU, J., XU, C., YIN, J., CHANG, M., WANG, W., LI, L., BROWN, G. W. & HICKSON, I. D. 2006a. BLAP75/RMI1 promotes the BLM-dependent dissolution of homologous recombination intermediates. *Proceedings of the National Academy of Sciences,* 103**,** 4068--4073.

WU, L. & HICKSON, I. D. 2003. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature,* 426**,** 870--874.

WU, P.-Y., FRIT, P., MEESALA, S., DAUVILLIER, S. P., MODESTI, M., ANDRES, S. N., HUANG, Y., SEKIGUCHI, J., CALSOU, P., SALLES, B. & JUNOP, M. S. 2009. Structural and Functional Interaction between the Human DNA Repair Proteins DNA Ligase IV and XRCC4. *Molecular and Cellular Biology,* 29**,** 3163--3172.

WU, Q., OCHI, T., MATAK-VINKOVIC, D., ROBINSON, CAROL V., CHIRGADZE, DIMITRI Y. & BLUNDELL, TOM L. 2011. Non-homologous end-joining partners in a helical dance: structural studies of XLF–XRCC4 interactions. *Biochemical Society Transactions,* 39**,** 1387--1392.

WU, R. & KAISER, A. D. 1968. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *Journal of Molecular Biology,* 35**,** 523--537.

WU, R. A., SEMLOW, D. R., KAMIMAE-LANNING, A. N., KOCHENOVA, O. V., CHISTOL, G., HODSKINSON, M. R., AMUNUGAMA, R., SPARKS, J. L., WANG, M., DENG, L., MIMOSO, C. A., LOW, E., PATEL, K. J. & WALTER, J. C. 2019. TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature,* 567**,** 267--272.

WU, W., LIU, Y., ZHOU, Q., WANG, Q., LUO, F., XU, Z., GENG, Q., LI, P., ZHANG, H. Z. & XIE, J. 2017. Novel homozygous FANCL mutation and somatic heterozygous SETBP1 mutation in a Chinese girl with Fanconi Anemia. *European Journal of Medical Genetics,* 60**,** 369--373.

WU, X., PLATT, J. L. & CASCALHO, M. 2003. Dimerization of MLH1 and PMS2 Limits Nuclear Localization of MutLα. *Molecular and Cellular Biology,* 23**,** 3320--3328.

WU, X., SHELL, S. M., YANG, Z. & ZOU, Y. 2006b. Phosphorylation of Nucleotide Excision Repair Factor Xeroderma Pigmentosum Group A by Ataxia Telangiectasia Mutated and Rad3-Related–Dependent Checkpoint Pathway Promotes Cell Survival in Response to UV Irradiation. *Cancer Research,* 66**,** 2997--3005.

WYATT, HALEY D. M., SARBAJNA, S., MATOS, J. & WEST, STEPHEN C. 2013. Coordinated Actions of SLX1-SLX4 and MUS81-EME1 for Holliday Junction Resolution in Human Cells. *Molecular Cell,* 52**,** 234--247.

XIE, J., GUILLEMETTE, S., PENG, M., GILBERT, C., BUERMEYER, A. & CANTOR, S. B. 2010. An MLH1 Mutation Links BACH1/FANCJ to Colon Cancer, Signaling, and Insight toward Directed Therapy. *Cancer Prevention Research,* 3**,** 1409--1416.

XU, C. 2018. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Computational and Structural Biotechnology Journal,* 16**,** 15--24.

XU, G., CHAPMAN, J. R., BRANDSMA, I., YUAN, J., MISTRIK, M., BOUWMAN, P., BARTKOVA, J., GOGOLA, E., WARMERDAM, D., BARAZAS, M., JASPERS, J. E., WATANABE, K., PIETERSE, M., KERSBERGEN, A., SOL, W., CELIE, P. H. N., SCHOUTEN, P. C., BROEK, B. V. D., SALMAN, A., NIEUWLAND, M., RINK, I. D., RONDE, J. D., JALINK, K., BOULTON, S. J., CHEN, J., GENT, D. C.

V., BARTEK, J., JONKERS, J., BORST, P. & ROTTENBERG, S. 2015. REV7 counteracts DNA double-strand break resection and affects PARP inhibition. *Nature,* 521**,** 541--544.
XU, Z., ZAN, H., PONE, E. J., MAI, T. & CASALI, P. 2012. Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nature Reviews Immunology,* 12**,** 517--531.
YAMAGUCHI, M., DAO, V. & MODRICH, P. 1998. MutS and MutL Activate DNA Helicase II in a Mismatch-dependent Manner*. *Journal of Biological Chemistry,* 273**,** 9197--9201.
YAMAMOTO, H. & IMAI, K. 2015. Microsatellite instability: an update. *Archives of Toxicology,* 89**,** 899--921.
YAMAMOTO, K. N., KOBAYASHI, S., TSUDA, M., KURUMIZAKA, H., TAKATA, M., KONO, K., JIRICNY, J., TAKEDA, S. & HIROTA, K. 2011. Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. *Proceedings of the National Academy of Sciences,* 108**,** 6492--6496.
YAMANE, K., TAYLOR, K. & KINSELLA, T. J. 2004. Mismatch repair-mediated G2/M arrest by 6-thioguanine involves the ATR–Chk1 pathway. *Biochemical and Biophysical Research Communications,* 318**,** 297--302.
YAN, T., DESAI, A. B., JACOBBERGER, J. W., SRAMKOSKI, R. M., LOH, T. & KINSELLA, T. J. 2004. CHK1 and CHK2 are differentially involved in mismatch repair–mediated 6-thioguanine-induced cell cycle checkpoint responses. *Molecular Cancer Therapeutics,* 3**,** 1147--1157.
YAN, Z., DELANNOY, M., LING, C., DAEE, D., OSMAN, F., MUNIANDY, P. A., SHEN, X., OOSTRA, A. B., DU, H., STELTENPOOL, J., LIN, T., SCHUSTER, B., DÉCAILLET, C., STASIAK, A., STASIAK, A. Z., STONE, S., HOATLIN, M. E., SCHINDLER, D., WOODCOCK, C. L., JOENJE, H., SEN, R., WINTER, J. P. D., LI, L., SEIDMAN, M. M., WHITBY, M. C., MYUNG, K., CONSTANTINOU, A. & WANG, W. 2010. A Histone-Fold Complex and FANCM Form a Conserved DNA-Remodeling Complex to Maintain Genome Stability. *Molecular Cell,* 37**,** 865--878.
YANG, H., JEFFREY, P. D., MILLER, J., KINNUCAN, E., SUN, Y., THOMÄ, N. H., ZHENG, N., CHEN, P.-L., LEE, W.-H. & PAVLETICH, N. P. 2002a. BRCA2 Function in DNA Binding and Recombination from a BRCA2-DSS1-ssDNA Structure. *Science,* 297**,** 1837--1848.
YANG, J., ZAITLEN, N. A., GODDARD, M. E., VISSCHER, P. M. & PRICE, A. L. 2014. Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics,* 46**,** 100--106.
YANG, Z.-G., LIU, Y., MAO, L. Y., ZHANG, J.-T. & ZOU, Y. 2002b. Dimerization of Human XPA and Formation of XPA2−RPA Protein Complex †. *Biochemistry,* 41**,** 13012--13020.
YAZINSKI, S. A., COMAILLS, V., BUISSON, R., GENOIS, M.-M., NGUYEN, H. D., HO, C. K., KWAN, T. T., MORRIS, R., LAUFFER, S., NUSSENZWEIG, A., RAMASWAMY, S., BENES, C. H., HABER, D. A., MAHESWARAN, S., BIRRER, M. J. & ZOU, L. 2017. ATR inhibition disrupts rewired homologous recombination and fork protection pathways in PARP inhibitor-resistant BRCA-deficient cancer cells. *Genes \& Development,* 31**,** 318--332.
YOKOI, M., MASUTANI, C., MAEKAWA, T., SUGASAWA, K., OHKUMA, Y. & HANAOKA, F. 2000. The Xeroderma Pigmentosum Group C Protein Complex XPC-HR23B Plays an Important Role in the Recruitment of Transcription Factor IIH to Damaged DNA*. *Journal of Biological Chemistry,* 275**,** 9870--9875.
YONEKURA, S.-I., NAKAMURA, N., YONEI, S. & ZHANG-AKIYAMA, Q.-M. 2009. Generation, Biological Consequences and Repair Mechanisms of Cytosine Deamination in DNA. *Journal of Radiation Research,* 50**,** 19--26.
YOSHIDA, R. 2021. Hereditary breast and ovarian cancer (HBOC): review of its molecular characteristics, screening, treatment, and prognosis. *Breast Cancer,* 28**,** 1167--1180.
YOUDS, J. L., BARBER, L. J. & BOULTON, S. J. 2009. C. elegans: A model of Fanconi anemia and ICL repair. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 668**,** 103--116.

YU, Y., WANG, W., DING, Q., YE, R., CHEN, D., MERKLE, D., SCHRIEMER, D., MEEK, K. & LEES-MILLER, S. P. 2003. DNA-PK phosphorylation sites in XRCC4 are not required for survival after radiation or for V(D)J recombination. *DNA Repair,* 2**,** 1239--1252.

YURCHENKO, V., XUE, Z. & SADOFSKY, M. J. 2006. SUMO Modification of Human XRCC4 Regulates Its Localization and Function in DNA Double-Strand Break Repair. *Molecular and Cellular Biology,* 26**,** 1786--1794.

YURGELUN, M. B., ALLEN, B., KALDATE, R. R., BOWLES, K. R., JUDKINS, T., KAUSHIK, P., ROA, B. B., WENSTRUP, R. J., HARTMAN, A.-R. & SYNGAL, S. 2015. Identification of a Variety of Mutations in Cancer Predisposition Genes in Patients With Suspected Lynch Syndrome. *Gastroenterology,* 149**,** 604--613.e20.

YURGELUN, M. B., KULKE, M. H., FUCHS, C. S., ALLEN, B. A., UNO, H., HORNICK, J. L., UKAEGBU, C. I., BRAIS, L. K., MCNAMARA, P. G., MAYER, R. J., SCHRAG, D., MEYERHARDT, J. A., NG, K., KIDD, J., SINGH, N., HARTMAN, A.-R., WENSTRUP, R. J. & SYNGAL, S. 2017. Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *Journal of Clinical Oncology,* 35**,** JCO.2016.71.001.

ZATREANU, D., ROBINSON, H. M. R., ALKHATIB, O., BOURSIER, M., FINCH, H., GEO, L., GRANDE, D., GRINKEVICH, V., HEALD, R. A., LANGDON, S., MAJITHIYA, J., MCWHIRTER, C., MARTIN, N. M. B., MOORE, S., NEVES, J., RAJENDRA, E., RANZANI, M., SCHAEDLER, T., STOCKLEY, M., WIGGINS, K., BROUGH, R., SRIDHAR, S., GULATI, A., SHAO, N., BADDER, L. M., NOVO, D., KNIGHT, E. G., MARLOW, R., HAIDER, S., CALLEN, E., HEWITT, G., SCHIMMEL, J., PREVO, R., ALLI, C., FERDINAND, A., BELL, C., BLENCOWE, P., BOT, C., CALDER, M., CHARLES, M., CURRY, J., EKWURU, T., EWINGS, K., KRAJEWSKI, W., MACDONALD, E., MCCARRON, H., PANG, L., PEDDER, C., RIGOREAU, L., SWARBRICK, M., WHEATLEY, E., WILLIS, S., WONG, A. C., NUSSENZWEIG, A., TIJSTERMAN, M., TUTT, A., BOULTON, S. J., HIGGINS, G. S., PETTITT, S. J., SMITH, G. C. M. & LORD, C. J. 2021. Polθ inhibitors elicit BRCA-gene synthetic lethality and target PARP inhibitor resistance. *Nature Communications,* 12**,** 3636.

ZEMAN, M. K. & CIMPRICH, K. A. 2014. Causes and consequences of replication stress. *Nature Cell Biology,* 16**,** 2--9.

ZHANG, H., CHEN, Z., YE, Y., YE, Z., CAO, D., XIONG, Y., SRIVASTAVA, M., FENG, X., TANG, M., WANG, C., TAINER, J. A. & CHEN, J. 2019a. SLX4IP acts with SLX4 and XPF–ERCC1 to promote interstrand crosslink repair. *Nucleic Acids Research,* 47**,** 10181--10201.

ZHANG, J. & WALTER, J. C. 2014. Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA Repair,* 19**,** 135--142.

ZHANG, M., HU, C., MOSES, N., HAAKENSON, J., XIANG, S., QUAN, D., FANG, B., YANG, Z., BAI, W., BEPLER, G., LI, G.-M. & ZHANG, X. M. 2019b. HDAC6 regulates DNA damage response via deacetylating MLH1. *Journal of Biological Chemistry,* 294**,** 5813--5826.

ZHANG, Q., KARNAK, D., TAN, M., LAWRENCE, T. S., MORGAN, M. A. & SUN, Y. 2016. FBXW7 Facilitates Nonhomologous End-Joining via K63-Linked Polyubiquitylation of XRCC4. *Molecular Cell,* 61**,** 419--433.

ZHANG, X.-Y., LANGENICK, J., TRAYNOR, D., BABU, M. M., KAY, R. R. & PATEL, K. J. 2009. Xpf and Not the Fanconi Anaemia Proteins or Rev3 Accounts for the Extreme Resistance to Cisplatin in Dictyostelium discoideum. *PLoS Genetics,* 5**,** e1000645.

ZHAO, B., ROTHENBERG, E., RAMSDEN, D. A. & LIEBER, M. R. 2020. The molecular basis and disease relevance of non-homologous DNA end joining. *Nature Reviews Molecular Cell Biology***,** 1--17.

ZHAO, H. & PIWNICA-WORMS, H. 2001. ATR-Mediated Checkpoint Pathways Regulate Phosphorylation and Activation of Human Chk1. *Molecular and Cellular Biology,* 21**,** 4129--4139.

ZHOU, H., PALMA, S. D., PREISINGER, C., PENG, M., POLAT, A. N., HECK, A. J. R. & MOHAMMED, S. 2013. Toward a Comprehensive Characterization of a Human Cancer Cell Phosphoproteome. *Journal of Proteome Research,* 12**,** 260--271.

ZHOU, J., GELOT, C., PANTELIDOU, C., LI, A., YÜCEL, H., DAVIS, R. E., FARKKILA, A., KOCHUPURAKKAL, B., SYED, A., SHAPIRO, G. I., TAINER, J. A., BLAGG, B. S. J., CECCALDI, R. & D'ANDREA, A. D. 2021. A first-in-class Polymerase Theta Inhibitor selectively targets Homologous-Recombination-Deficient Tumors. *Nature cancer,* 2**,** 598--610.

ZHOU, L.-P., LUAN, H., DONG, X.-H., JIN, G.-J., MA, D.-L. & SHANG, H. 2012a. Association of Functional Polymorphisms of the XRCC4 Gene with the Risk of Breast Cancer: A Meta-analysis. *Asian Pacific Journal of Cancer Prevention,* 13**,** 3431--3436.

ZHOU, W., OTTO, E. A., CLUCKEY, A., AIRIK, R., HURD, T. W., CHAKI, M., DIAZ, K., LACH, F. P., BENNETT, G. R., GEE, H. Y., GHOSH, A. K., NATARAJAN, S., THONGTHIP, S., VETURI, U., ALLEN, S. J., JANSSEN, S., RAMASWAMI, G., DIXON, J., BURKHALTER, F., SPOENDLIN, M., MOCH, H., MIHATSCH, M. J., VERINE, J., READE, R., SOLIMAN, H., GODIN, M., KISS, D., MONGA, G., MAZZUCCO, G., AMANN, K., ARTUNC, F., NEWLAND, R. C., WIECH, T., ZSCHIEDRICH, S., HUBER, T. B., FRIEDL, A., SLAATS, G. G., JOLES, J. A., GOLDSCHMEDING, R., WASHBURN, J., GILES, R. H., LEVY, S., SMOGORZEWSKA, A. & HILDEBRANDT, F. 2012b. FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. *Nature Genetics,* 44**,** 910--915.

ZHU, Z., CHUNG, W.-H., SHIM, E. Y., LEE, S. E. & IRA, G. 2008. Sgs1 Helicase and Two Nucleases Dna2 and Exo1 Resect DNA Double-Strand Break Ends. *Cell,* 134**,** 981--994.

ZIANI, S., NAGY, Z., ALEKSEEV, S., SOUTOGLOU, E., EGLY, J.-M. & COIN, F. 2014. Sequential and ordered assembly of a large DNA repair complex on undamaged chromatin. *J Cell Biol,* 206**,** 589--598.

ZIMMERMANN, M., LOTTERSBERGER, F., BUONOMO, S. B., SFEIR, A. & LANGE, T. D. 2013. 53BP1 Regulates DSB Repair Using Rif1 to Control 5' End Resection. *Science,* 339**,** 700--704.

ZIMMERMANN, M., MURINA, O., REIJNS, M. A. M., AGATHANGGELOU, A., CHALLIS, R., TARNAUSKAITĖ, Ž., MUIR, M., FLUTEAU, A., AREGGER, M., MCEWAN, A., YUAN, W., CLARKE, M., LAMBROS, M. B., PANEESHA, S., MOSS, P., CHANDRASHEKHAR, M., ANGERS, S., MOFFAT, J., BRUNTON, V. G., HART, T., BONO, J. D., STANKOVIC, T., JACKSON, A. P. & DUROCHER, D. 2018. CRISPR screens identify genomic ribonucleotides as a source of PARP-trapping lesions. *Nature,* 559**,** 285--289.

ZOU, L. & ELLEDGE, S. J. 2003. Sensing DNA Damage Through ATRIP Recognition of RPA-ssDNA Complexes. *Science,* 300**,** 1542--1548.

ZOU, X., KOH, G. C. C., NANDA, A. S., DEGASPERI, A., URGO, K., ROUMELIOTIS, T. I., AGU, C. A., BADJA, C., MOMEN, S., YOUNG, J., AMARANTE, T. D., SIDE, L., BRICE, G., PEREZ-ALONSO, V., RUEDA, D., GOMEZ, C., BUSHELL, W., HARRIS, R., CHOUDHARY, J. S., CONSORTIUM, G. E. R., AMBROSE, J. C., ARUMUGAM, P., BAPLE, E. L., BLEDA, M., BOARDMAN-PRETTY, F., BOISSIERE, J. M., BOUSTRED, C. R., BRITTAIN, H., CAULFIELD, M. J., CHAN, G. C., CRAIG, C. E. H., DAUGHERTY, L. C., BURCA, A. D., DEVEREAU, A., ELGAR, G., FOULGER, R. E., FOWLER, T., FURIÓ-TARÍ, P., GIESS, A., HACKETT, J. M., HALAI, D., HAMBLIN, A., HENDERSON, S., HOLMAN, J. E., HUBBARD, T. J. P., IBÁÑEZ, K., JACKSON, R., JONES, L. J., KASPERAVICIUTE, D., KAYIKCI, M., KOUSATHANAS, A., LAHNSTEIN, L., LAWSON, K., LEIGH, S. E. A., LEONG, I. U. S., LOPEZ, J. F., MALEADY-CROWE, F., MASON, J., MCDONAGH, E. M., MOUTSIANAS, L., MUELLER, M., MURUGAESU, N., NEED, A. C., O'DONOVAN, P., ODHAMS, C. A., ORIOLI, A., PATCH, C., PEREIRA, M. B., PEREZ-GIL, D., POLYCHRONOPOULOS, D., PULLINGER, J., RAHIM, T., RENDON, A., RIESGO-FERREIRO, P., ROGERS, T., RYTEN, M., SAVAGE, K., SAWANT, K., SCOTT, R. H., SIDDIQ, A., SIEGHART, A., SMEDLEY, D., SMITH, K. R., SMITH, S. C., SOSINSKY, A., SPOONER, W., STEVENS, H. E., STUCKEY, A., SULTANA, R., TANGUY, M. L., THOMAS, E. R. A., THOMPSON, S. R., TREGIDGO, C., TUCCI, A., WALSH, E., WATTERS, S. A., WELLAND, M. J., WILLIAMS, E., WITKOWSKA, K., WOOD, S. M., et al. 2021. A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nature Cancer,* 2**,** 643--657.

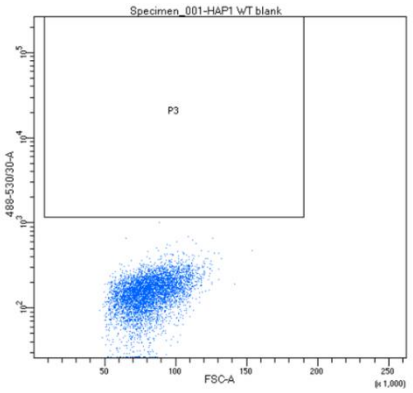
10 Appendices

10.1 FACS plots from vertebrate XPA orthologue UV survival assay

HAP1 WT Blank

Tube: HAP1 WT blank

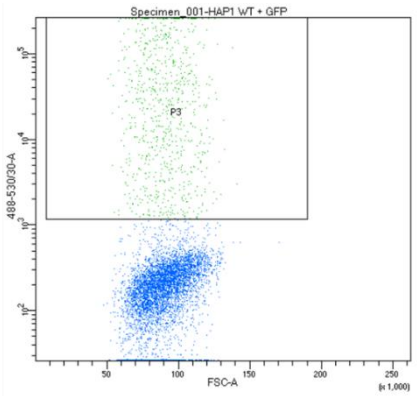
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
P1	9,142	91.4	91.4
P2	5,411	59.2	54.1
P3	0	0.0	0.0



HAP1 WT + GFP

Tube: HAP1 WT + GFP

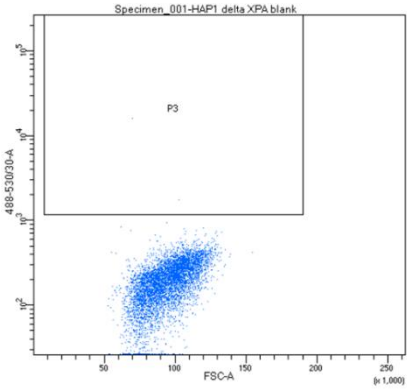
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
P1	8,503	85.0	85.0
P2	6,659	78.3	66.6
P3	809	12.1	8.1



HAP1 Δ*XPA* Blank

Tube: HAP1 delta XPA blank

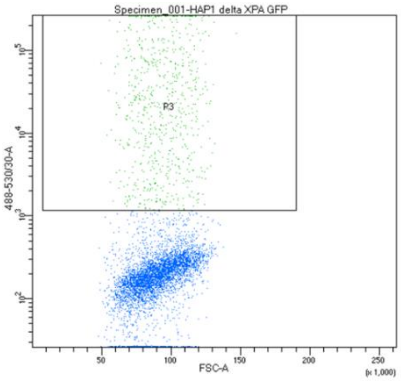
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
P1	8,585	85.9	85.9
P2	6,126	71.4	61.3
P3	2	0.0	0.0



HAP1 Δ*XPA* + GFP

Tube: HAP1 delta XPA GFP

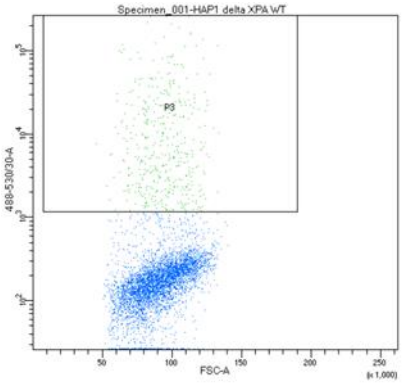
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
P1	8,516	85.2	85.2
P2	5,524	64.9	55.2
P3	703	12.7	7.0



HAP1 Δ*XPA* + GFP-human XPA

Tube: HAP1 delta XPA WT

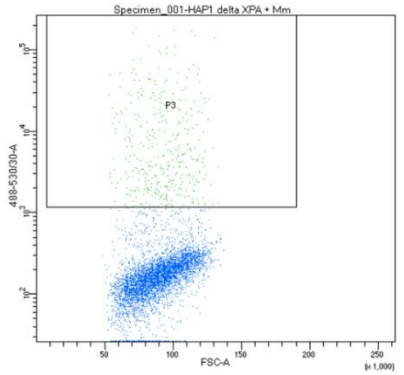
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
P1	8,337	83.4	83.4
P2	5,154	61.8	51.5
P3	446	8.7	4.5



HAP1 Δ*XPA* + GFP-mouse XPA

Tube: HAP1 delta XPA + Mm

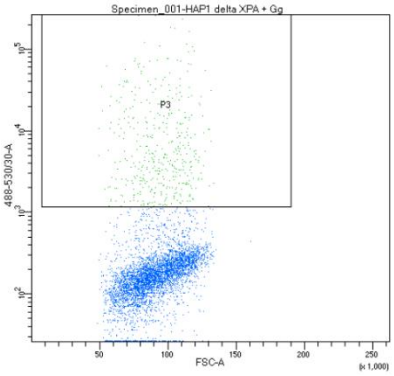
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
P1	8,568	85.7	85.7
P2	5,236	61.1	52.4
P3	392	7.5	3.9



HAP1 Δ*XPA* + GFP-chicken XPA

Tube: HAP1 delta XPA + Gg

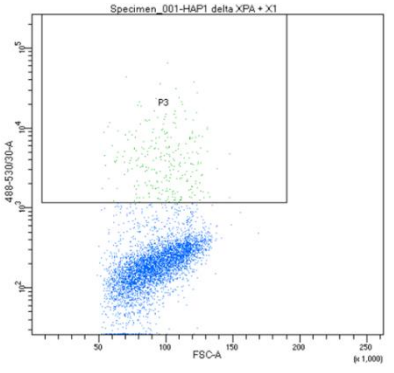
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
P1	8,285	82.8	82.8
P2	5,064	61.1	50.6
P3	427	8.4	4.3



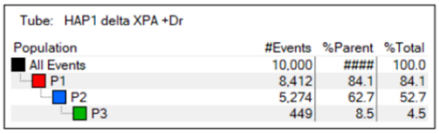
HAP1 Δ*XPA* + GFP-frog XPA

Tube: HAP1 delta XPA + X1

Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
P1	8,283	82.8	82.8
P2	5,088	61.4	50.9
P3	274	5.4	2.7



HAP1 Δ*XPA* + GFP-fish XPA



Tube: HAP1 delta XPA +Dr			
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
P1	8,412	84.1	84.1
P2	5,274	62.7	52.7
P3	449	8.5	4.5

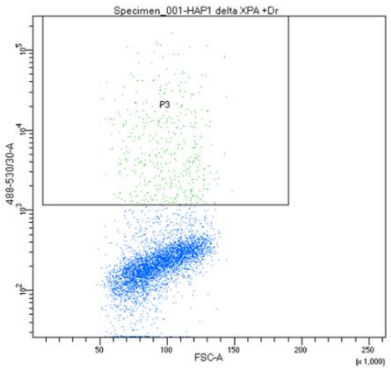


Figure 10-1: Representative FACS plots from HAP1 WT and HAP1 Δ*XPA* cells either untransfected or transfected with GFP and GFP-tagged vertebrate XPA orthologues

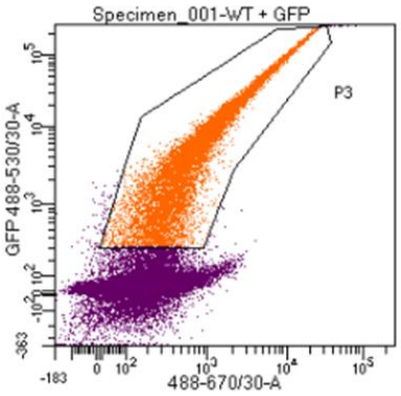
FACS plots obtained from HAP1 WT and HAP1 Δ*XPA* cells either untransfected or transfected with GFP and GFP-tagged vertebrate XPA orthologues as well as untransfected blank controls. % GFP + cells indicated by P3 gate in table and on FACS plot.

10.2 FACS plots from invertebrate XPA orthologue UV survival assay

HAP1 WT + GFP

Tube: WT + GFP

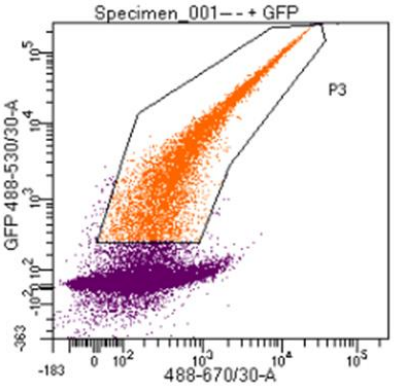
Population	#Events	%Parent	%Total
All Events	114,170	####	100.0
P1	95,726	83.8	83.8
P2	95,121	99.4	83.3
P3	10,083	10.6	8.8



HAP1 Δ*XPA* + GFP

Tube: - - + GFP

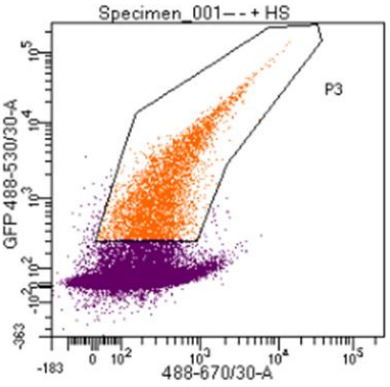
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	81,210	81.2	81.2
P2	80,288	98.9	80.3
P3	6,535	8.1	6.5



HAP1 Δ*XPA* + GFP-human XPA

Tube: - - + HS

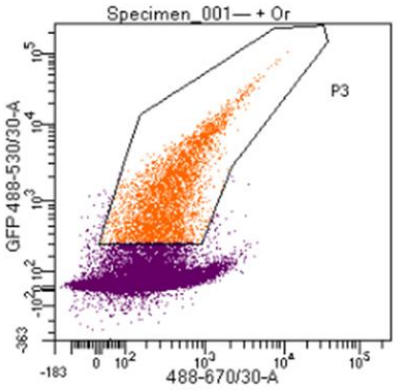
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	84,012	84.0	84.0
P2	83,115	98.9	83.1
P3	4,404	5.3	4.4



HAP1 Δ*XPA* + GFP-fish XPA

Tube: -- + Or

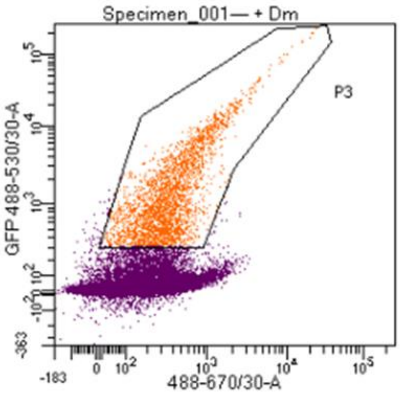
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	81,685	81.7	81.7
P2	80,543	98.6	80.5
P3	4,069	5.1	4.1



HAP1 Δ*XPA* + GFP-fly XPA

Tube: -- + Dm

Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	83,564	83.6	83.6
P2	82,600	98.8	82.6
P3	2,883	3.5	2.9



HAP1 Δ*XPA* + GFP-yeast XPA

Tube: -- + Sp

Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	85,431	85.4	85.4
P2	84,742	99.2	84.7
P3	1,782	2.1	1.8

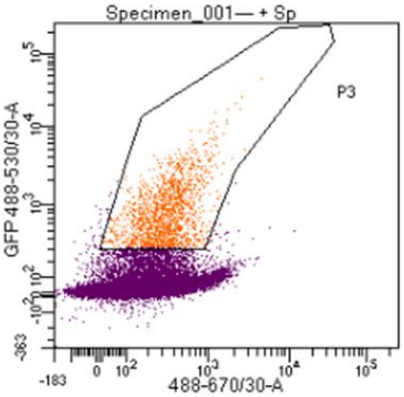


Figure 10-2: Representative FACS plots from HAP1 WT and HAP1 Δ*XPA* cells transfected with GFP and GFP-tagged vertebrate and invertebrate XPA orthologues

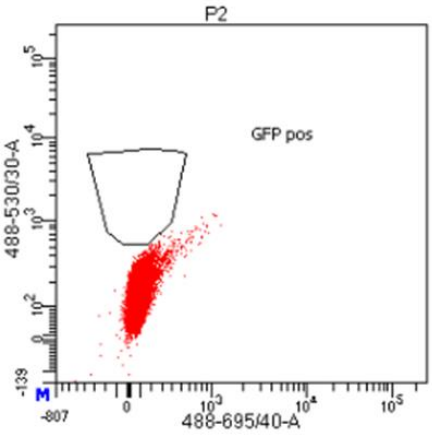
FACS plots obtained from HAP1 WT and HAP1 Δ*XPA* cells transfected with GFP and GFP-tagged vertebrate and invertebrate XPA orthologues. % GFP + cells indicated by P3 gate in table and on FACS plot.

10.3 FACS plots from XPA variants UV survival assay

HAP1 WT Blank

Tube: blank

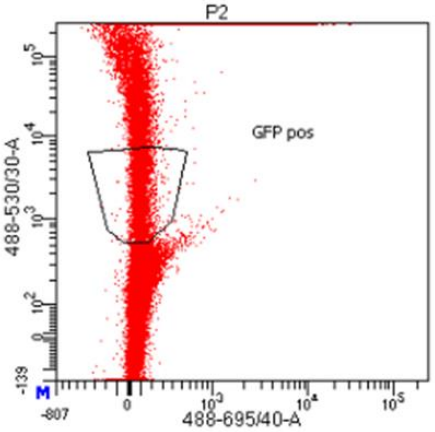
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	80,442	80.4	80.4
P2	80,382	99.9	80.4
GFP pos	0	0.0	0.0



HAP1 WT + GFP

Tube: WT+GFP

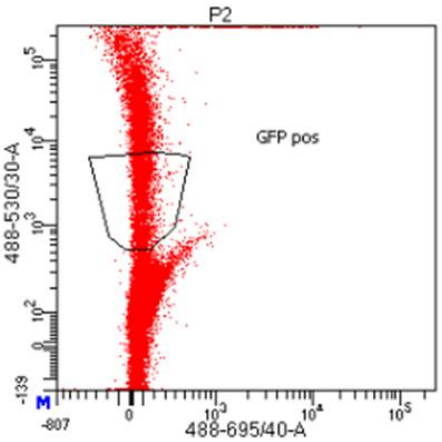
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	80,555	80.6	80.6
P2	80,496	99.9	80.5
GFP pos	3,950	4.9	4.0



HAP1 Δ*XPA* + GFP

Tube: D+GFP

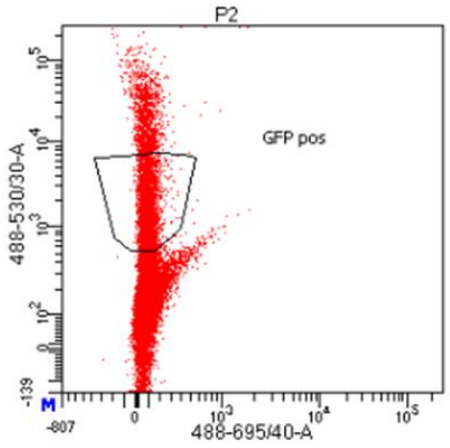
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	78,809	78.8	78.8
P2	78,790	100.0	78.8
GFP pos	3,073	3.9	3.1



HAP1 Δ*XPA* + GFP-human XPA

Tube: WT+HS

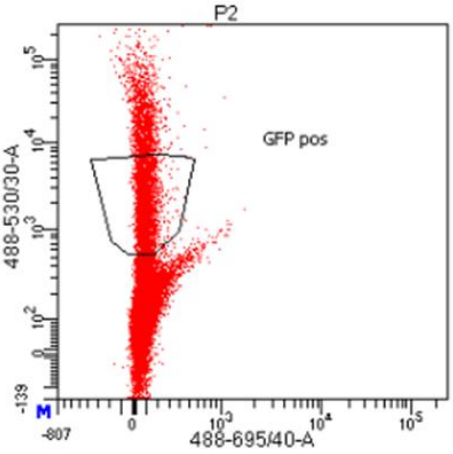
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	77,367	77.4	77.4
P2	77,339	100.0	77.3
GFP pos	3,470	4.5	3.5



HAP1 Δ*XPA* + GFP-XPA I68T

Tube: D+I68T

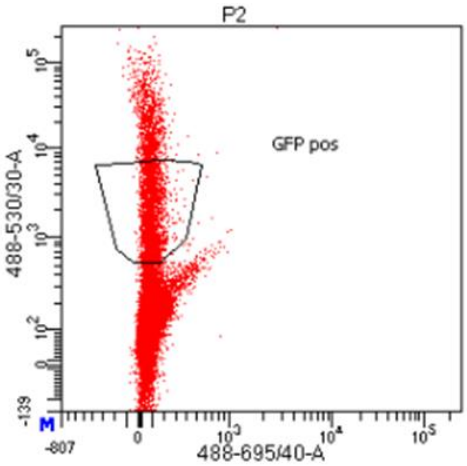
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	78,191	78.2	78.2
P2	78,165	100.0	78.2
GFP pos	3,472	4.4	3.5



HAP1 Δ*XPA* + GFP-XPA K141R

Tube: D+K141R

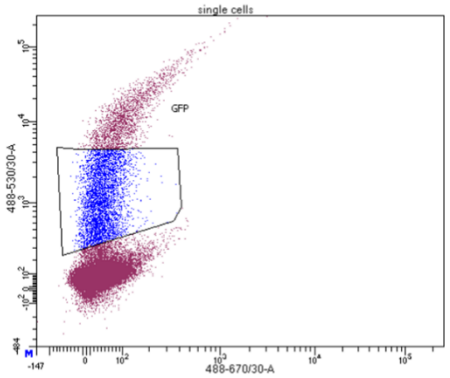
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	71,334	71.3	71.3
P2	71,286	99.9	71.3
GFP pos	2,857	4.0	2.9



HAP1 Δ*XPA* + GFP-XPA V234L

Tube: 1_D+V234L

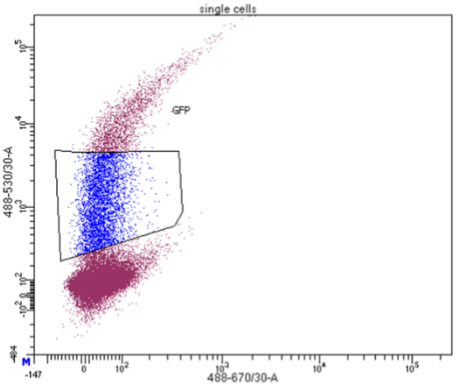
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	81,792	81.8	81.8
single cells	80,079	97.9	80.1
GFP	4,072	5.1	4.1



HAP1 Δ*XPA* + GFP-XPA L191V

Tube: 1_D+L191V

Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	81,048	81.0	81.0
single cells	79,593	98.2	79.6
GFP	4,061	5.1	4.1



HAP1 Δ*XPA* + GFP-XPA '3A'

Tube: 1_D+3A

Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	81,355	81.4	81.4
single cells	79,241	97.4	79.2
GFP	3,008	3.8	3.0

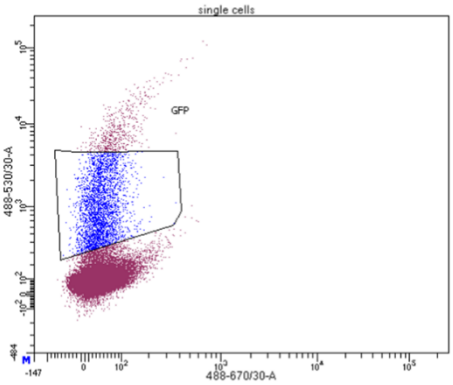


Figure 10-3: Representative FACS plots from HAP1 WT and HAP1 Δ*XPA* cells either untransfected or transfected with GFP and GFP-tagged XPA variant constructs.

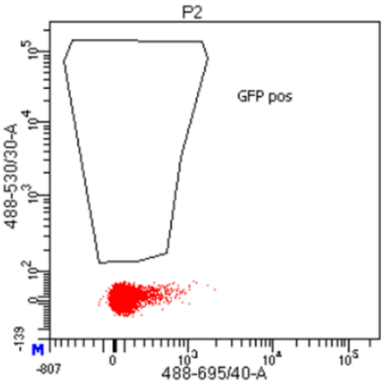
FACS plots obtained from HAP1 WT and HAP1 Δ*XPA* cells either untransfected or transfected with GFP and GFP-tagged XPA variant constructs. % GFP + cells indicated by P3 gate in table and on FACS plot

10.4 FACS plots from XRCC4 vertebrate orthologue IR survival assay

Blank

Tube: WT blank

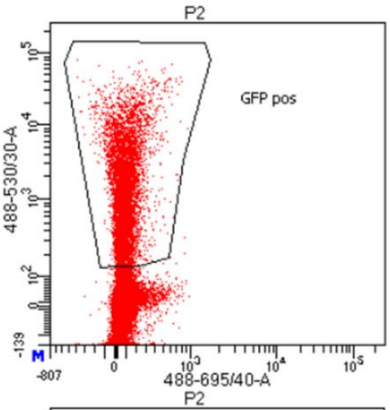
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	80,164	80.2	80.2
P2	72,209	90.1	72.2
GFP pos	0	0.0	0.0



HAP1 WT + GFP

Tube: WT + GFP

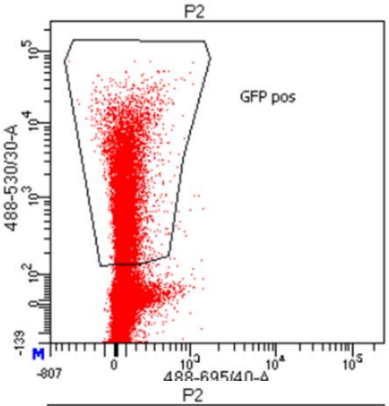
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	74,827	74.8	74.8
P2	69,273	92.6	69.3
GFP pos	7,738	11.2	7.7



HAP1 Δ*XRCC4* + GFP

Tube: delta GFP

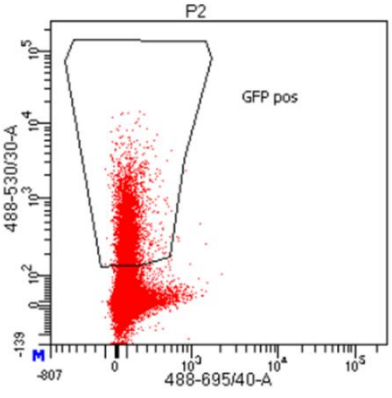
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	71,737	71.7	71.7
P2	68,150	95.0	68.2
GFP pos	10,694	15.7	10.7



HAP1 Δ*XRCC4* + GFP-human XRCC4

Tube: delta HS

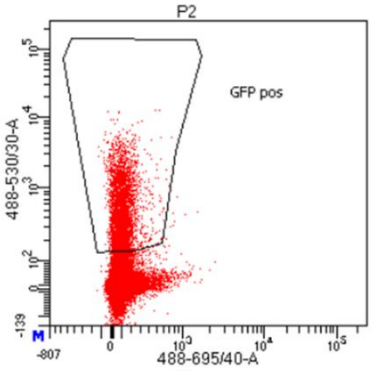
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	72,842	72.8	72.8
P2	70,618	96.9	70.6
GFP pos	3,892	5.5	3.9



HAP1 ΔXRCC4 + GFP-mouse XRCC4

Tube: delta MM

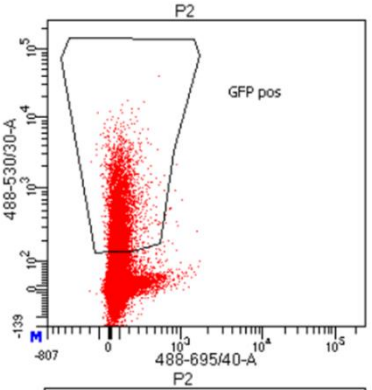
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	72,806	72.8	72.8
P2	70,332	96.6	70.3
GFP pos	4,747	6.7	4.7



HAP1 ΔXRCC4 + GFP-chicken XRCC4

Tube: delta Gg

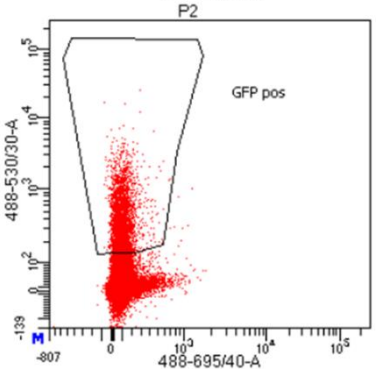
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	71,894	71.9	71.9
P2	69,107	96.1	69.1
GFP pos	5,115	7.4	5.1



HAP1 ΔXRCC4 + GFP-frog XRCC4

Tube: delta X1

Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	77,345	77.3	77.3
P2	74,447	96.3	74.4
GFP pos	3,597	4.8	3.6



HAP1 ΔXRCC4 + GFP-fish XRCC4

Tube: delta Dr

Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	79,502	79.5	79.5
P2	76,093	95.7	76.1
GFP pos	6,829	9.0	6.8

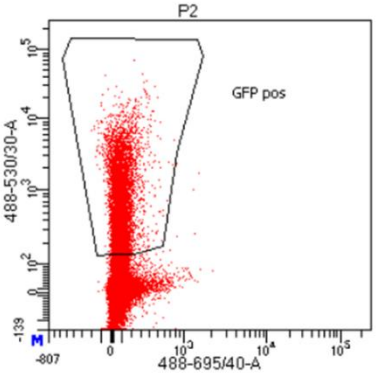


Figure 10-4: Representative FACS plots from HAP1 WT and HAP1 Δ*XRCC4* cells either untransfected or transfected with GFP and GFP-tagged vertebrate XRCC4 orthologues

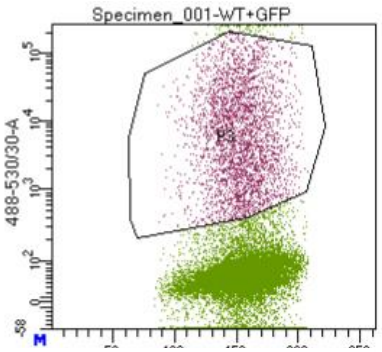
FACS plots obtained from HAP1 WT and HAP1 Δ*XRCC4* cells either untransfected or transfected with GFP and GFP-tagged vertebrate XRCC4 orthologues as well as untransfected blank controls. % GFP + cells indicated by P2 gate in table and on FACS plot.

10.5 FACS plots from XRCC4 6A variant IR survival assay

HAP1 WT + GFP

Tube: WT+GFP

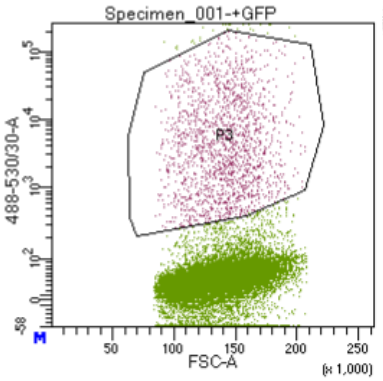
Population	#Events	%Parent	%Total
All Events	50,000	####	100.0
P1	26,838	53.7	53.7
P2	20,322	75.7	40.6
P3	3,427	16.9	6.9



HAP1 Δ*XRCC4* + GFP

Tube: +GFP

Population	#Events	%Parent	%Total
All Events	50,000	####	100.0
P1	29,689	59.4	59.4
P2	24,983	84.1	50.0
P3	1,744	7.0	3.5



HAP1 Δ*XRCC4* + GFP-XRCC4

Tube: +WT

Population	#Events	%Parent	%Total
All Events	50,000	####	100.0
P1	29,761	59.5	59.5
P2	25,133	84.4	50.3
P3	1,158	4.6	2.3

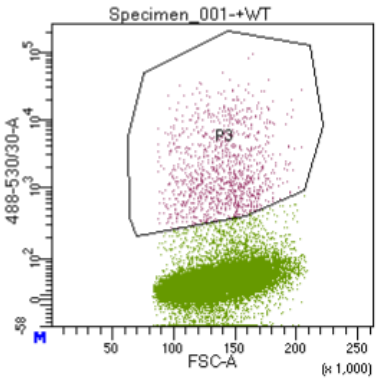
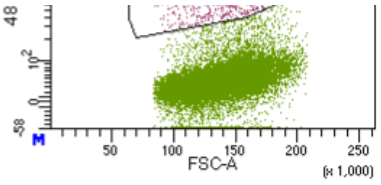


Figure 10-5: **Representative FACS plots from HAP1 WT and HAP1 Δ*XRCC4* cells transfected with GFP and GFP-tagged XRCC4 WT or variant constructs**

FACS plots obtained from HAP1 WT and HAP1 Δ*XRCC4* cells transfected with GFP and GFP-tagged WT or variant XRCC4 constructs. % GFP + cells indicated in table and in gate on FACS plot.

GFP-XRCC4 '6A'

P2	28,253	85.6	56.5
P3	1,404	5.0	2.8

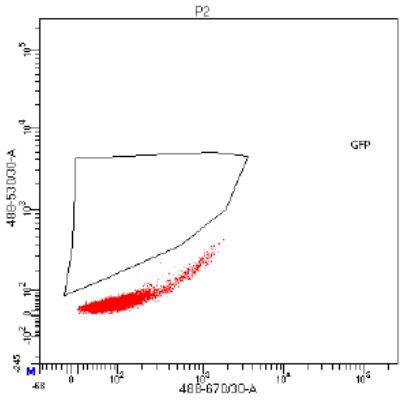


10.6 FACS plots from HAP1 ΔMLH1 MLH1 orthologue 6-TG survival assay

Blank (HAP1 WT)

Tube: Blank

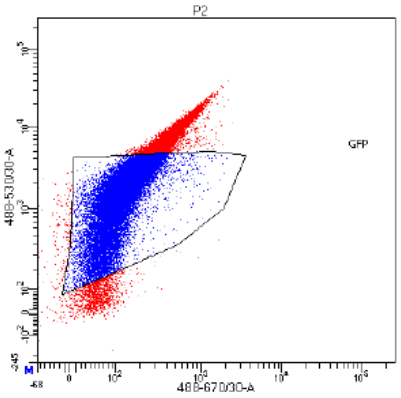
Population	#Events	%Parent	%Total
All Events	31,830	####	100.0
P1	29,501	92.7	92.7
P2	26,007	88.2	81.7
GFP	0	0.0	0.0



HAP1 WT + GFP

Tube: WT + GFP

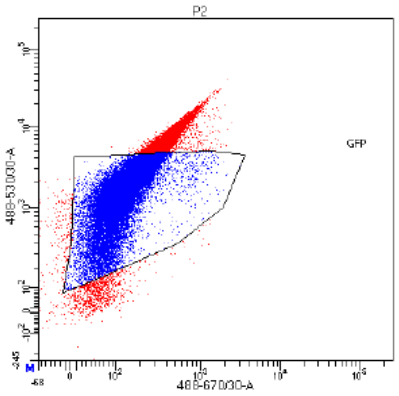
Population	#Events	%Parent	%Total
All Events	136,687	####	100.0
P1	112,759	82.5	82.5
P2	94,149	83.5	68.9
GFP	65,574	69.6	48.0



HAP1 Δ*MLH1* + GFP

Tube: D + GFP

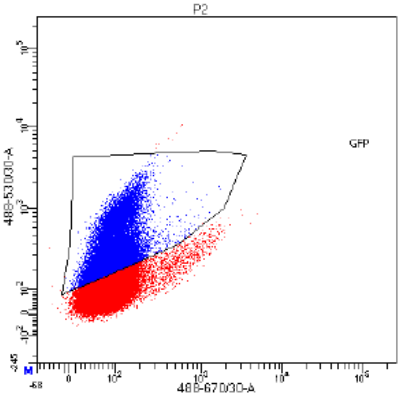
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	83,289	83.3	83.3
P2	68,538	82.3	68.5
GFP	53,142	77.5	53.1



HAP1 Δ*MLH1* + GFP-human MLH1

Tube: D + WT

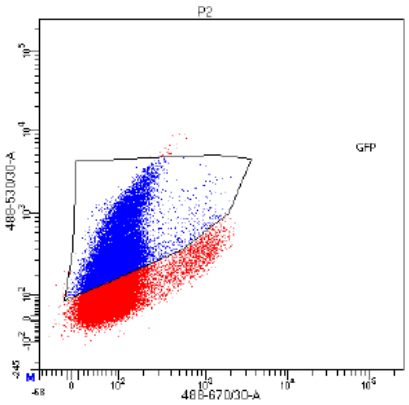
Population	#Events	%Parent	%Total
All Events	142,956	####	100.0
P1	123,946	86.7	86.7
P2	108,474	87.5	75.9
GFP	37,872	34.9	26.5



HAP1 Δ*MLH1* + GFP-mouse MLH1

Tube: D + MM

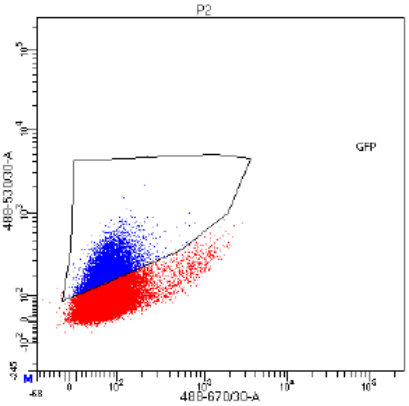
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	86,382	86.4	86.4
P2	76,365	88.4	76.4
GFP	31,274	41.0	31.3



HAP1 Δ*MLH1* + GFP-chicken MLH1

Tube: D + Gg

Population	#Events	%Parent	%Total
All Events	100,052	####	100.0
P1	87,191	87.1	87.1
P2	77,021	88.3	77.0
GFP	10,770	14.0	10.8



HAP1 Δ*MLH1* + GFP-frog MLH1

Tube: D + Xl

Population	#Events	%Parent	%Total
All Events	100,012	####	100.0
P1	87,853	87.8	87.8
P2	79,234	90.2	79.2
GFP	17,692	22.3	17.7

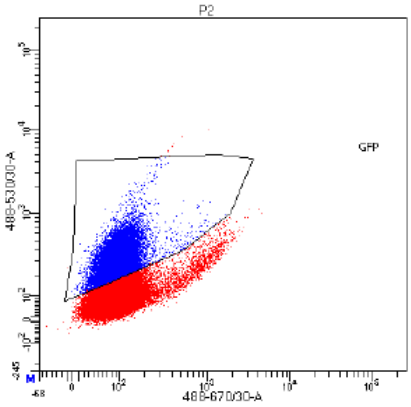


Figure 10-6: Representative FACS plots from HAP1 WT and HAP1 Δ*MLH1* cells constitutively expressing GFP or GFP-tagged MLH1 orthologues used for 6-TG sensitivity assays and PMS2 stability blots.

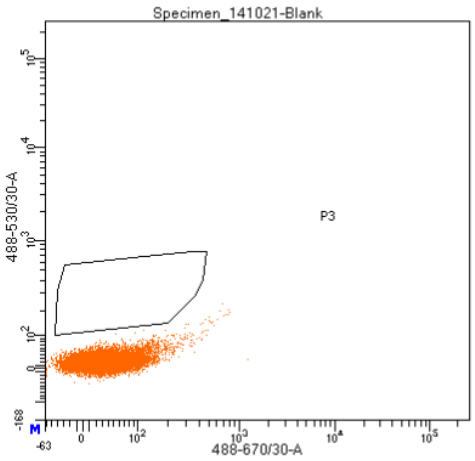
FACS plots obtained from HAP1 WT and HAP1 Δ*MLH1* cells transduced with lentivirus encoding GFP and GFP-tagged MLH1 orthologues. % GFP + cells indicated in table and shown in gate on FACS plot.

10.7 FACS plots from HCT116 MLH1 orthologue 6-TG survival assay

Blank HCT116

Tube: Blank

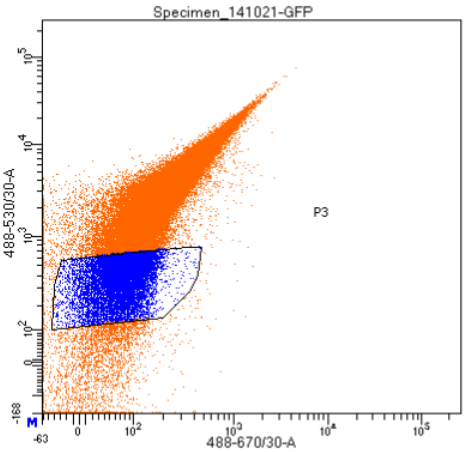
Population	#Events	%Parent	%Total
All Events	34,580	####	100.0
P1	26,138	75.6	75.6
P2	26,129	100.0	75.6
P3	0	0.0	0.0



HCT116 + GFP

Tube: GFP

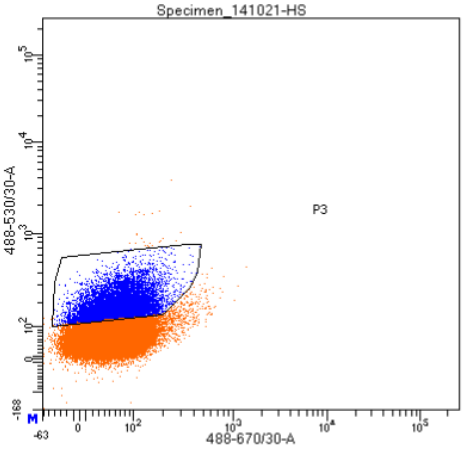
Population	#Events	%Parent	%Total
All Events	142,575	####	100.0
P1	113,821	79.8	79.8
P2	112,683	99.0	79.0
P3	13,315	11.8	9.3



HCT116 + GFP-human MLH1

Tube: HS

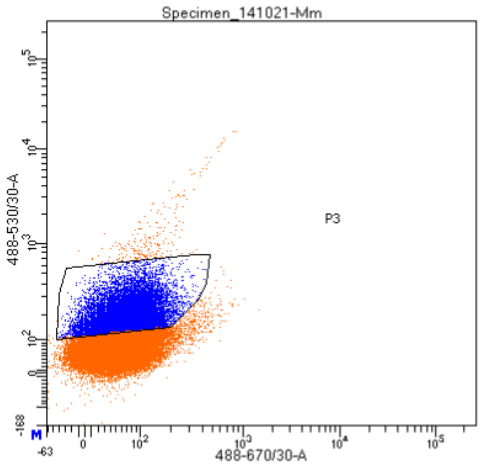
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	87,225	87.2	87.2
P2	86,500	99.2	86.5
P3	14,778	17.1	14.8



HCT116 + GFP-mouse MLH1

Tube: Mm

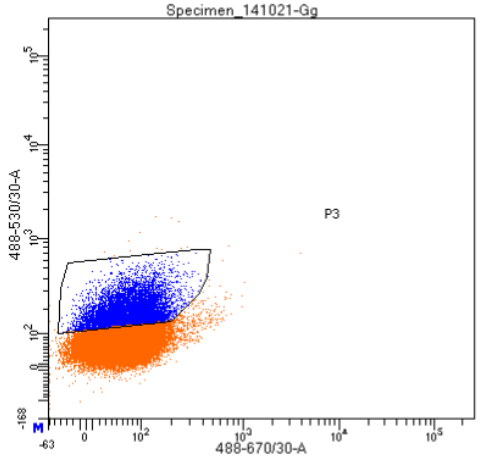
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	88,028	88.0	88.0
P2	87,441	99.3	87.4
P3	20,105	23.0	20.1



HCT116 + GFP-chicken MLH1

Tube: Gg

Population	#Events	%Parent	%Total
All Events	115,440	####	100.0
P1	81,247	70.4	70.4
P2	74,305	91.5	64.4
P3	10,977	14.8	9.5



HCT116 + GFP-frog MLH1

Tube: Xi

Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	87,226	87.2	87.2
P2	86,826	99.5	86.8
P3	2,149	2.5	2.1

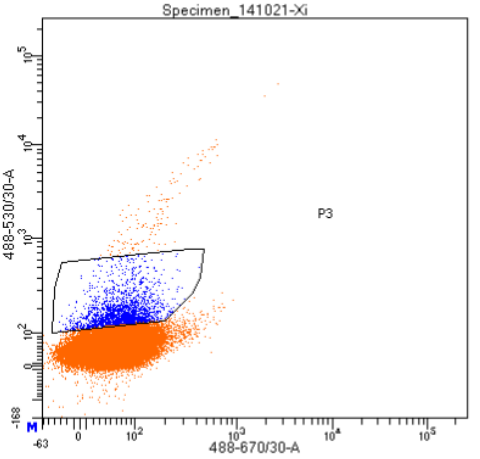


Figure 10-7: Representative FACS plots from HCT116 cells constitutively expressing GFP or GFP-tagged MLH1 orthologues used for 6-TG sensitivity assays and PMS2 stability blots.

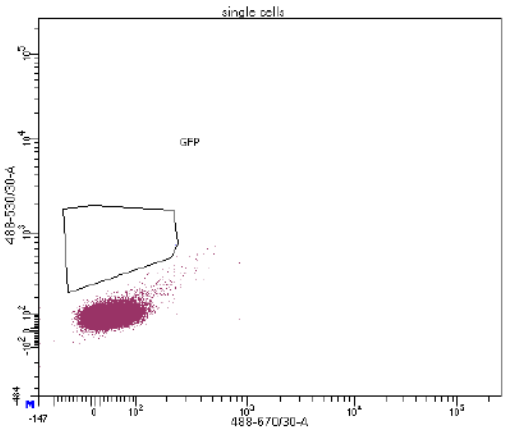
FACS plots obtained from HCT116 cells transduced with lentivirus encoding GFP and GFP-tagged MLH1 orthologues. % GFP + cells indicated in table and shown in gate on FACS plot.

10.8 FACS plots from MLH1 variant constructs 6-TG survival assay

Blank (HAP1 WT)

Tube: Blank

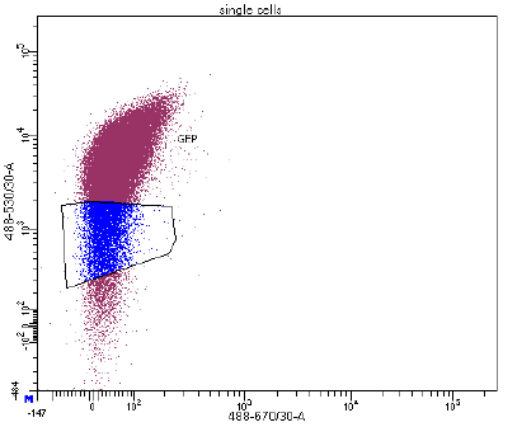
Population	#Events	%Parent	%Total
All Events	80,370	####	100.0
cells	63,973	79.6	79.6
single cells	61,760	96.5	76.8
GFP	1	0.0	0.0



HAP1 WT + GFP

Tube: WT + GFP

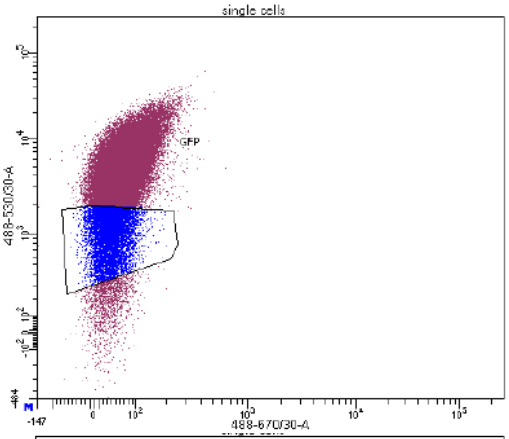
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	72,215	72.2	72.2
single cells	70,518	97.7	70.5
GFP	7,699	10.9	7.7



HAP1 Δ*MLH1* + GFP

Tube: D+GFP

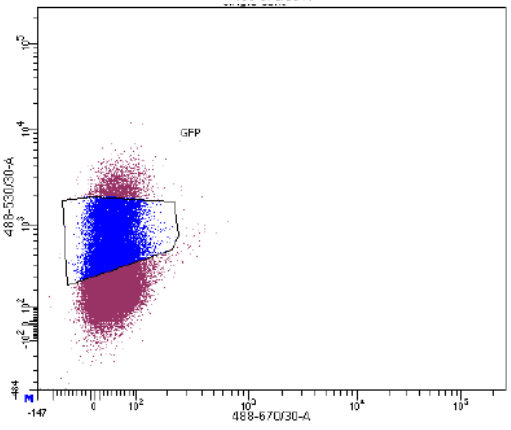
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	75,817	75.8	75.8
single cells	74,018	97.6	74.0
GFP	10,604	14.3	10.6



HAP1 Δ*MLH1* + GFP-WT MLH1

Tube: D+WT

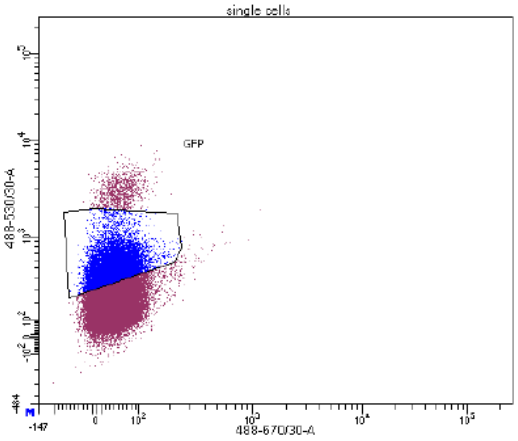
Population	#Events	%Parent	%Total
All Events	105,214	####	100.0
cells	67,705	64.3	64.3
single cells	62,822	92.8	59.7
GFP	19,722	31.4	18.7



HAP1 Δ*MLH1* + GFP-MLH1 S46N

Tube: D+S46N

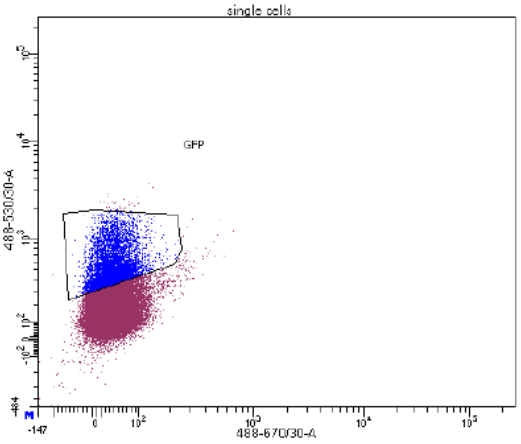
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	72,902	72.9	72.9
single cells	70,606	96.9	70.6
GFP	16,016	22.7	16.0



HAP1 Δ*MLH1* + GFP-MLH1 T82I

Tube: D+T82I

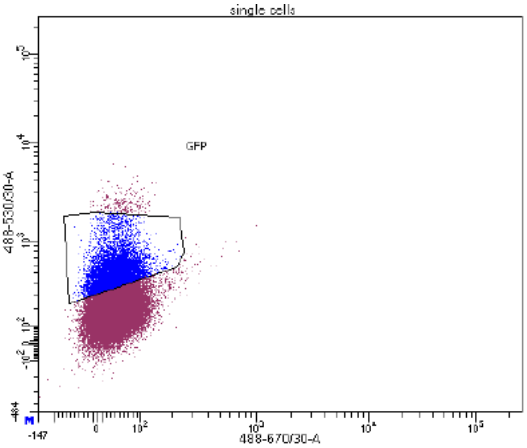
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	68,557	68.6	68.6
single cells	65,679	95.8	65.7
GFP	7,819	11.9	7.8



HAP1 Δ*MLH1* + GFP-MLH1 V185I

Tube: D+V185I

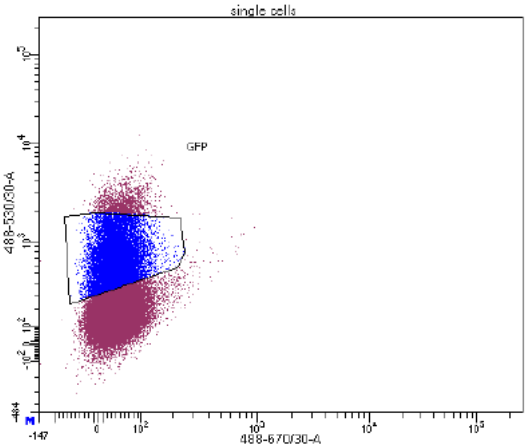
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	61,623	61.6	61.6
single cells	57,804	93.8	57.8
GFP	9,850	17.0	9.8



HAP1 Δ*MLH1* + GFP-MLH1 V185G

Tube: D+V185G

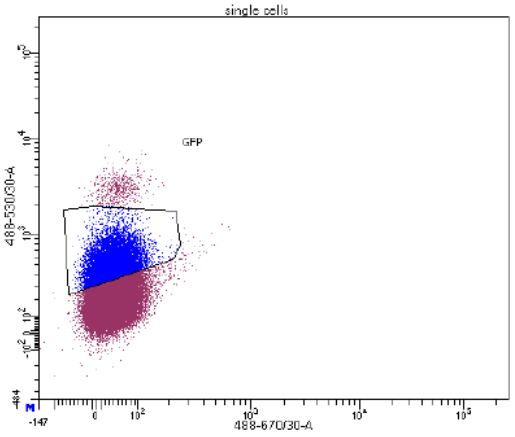
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	70,736	70.7	70.7
single cells	67,628	95.6	67.6
GFP	16,655	24.6	16.7



HAP1 Δ*MLH1* + GFP-MLH1 P208S

Tube: D + P208S

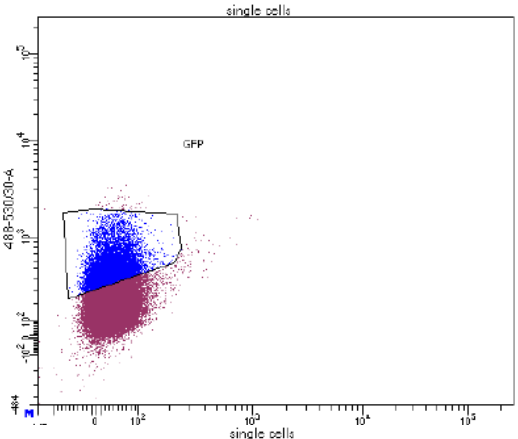
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	74,467	74.5	74.5
single cells	71,881	96.5	71.9
GFP	17,036	23.7	17.0



HAP1 Δ*MLH1* + GFP-MLH1 A619T

Tube: D + A619T

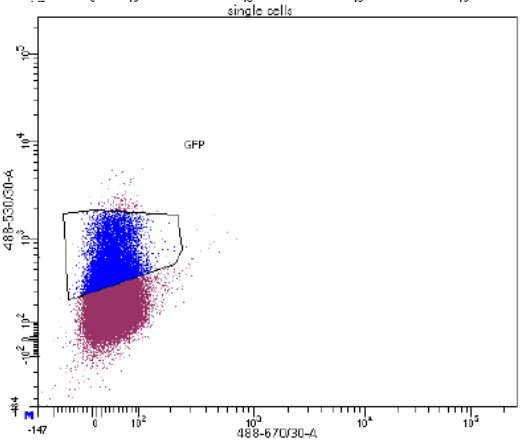
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	75,279	75.3	75.3
single cells	72,983	97.0	73.0
GFP	13,453	18.4	13.5



HAP1 Δ*MLH1* + GFP-MLH1 A681S

Tube: D + A681S

Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
cells	70,073	70.1	70.1
single cells	68,003	97.0	68.0
GFP	13,078	19.2	13.1



HAP1 Δ*MLH1* + GFP-MLH1 A681T

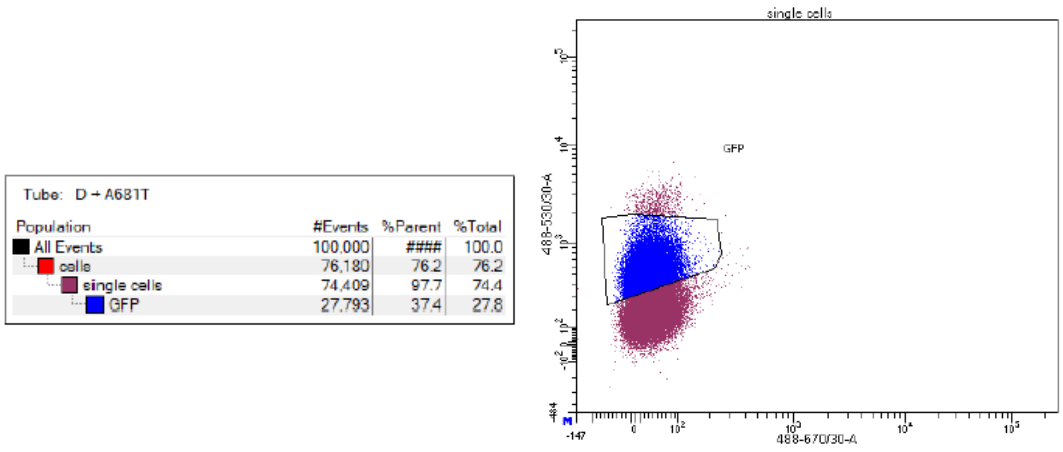


Figure 10-8: Representative FACS plots from HAP1 Δ*MLH1* cells constitutively expressing GFP or GFP-tagged MLH1 variant constructs used for 6-TG sensitivity assays and PMS2 stability/interaction blots.

FACS plots obtained from HAP1 WT or HAP1 Δ*MLH1* cells transduced with lentivirus encoding GFP GFP-tagged MLH1 variant constructs. % GFP + cells indicated in table and shown in gate on FACS plot.

10.9 Sequence alignment of vertebrate and invertebrate MLH1 orthologues highlighting A681

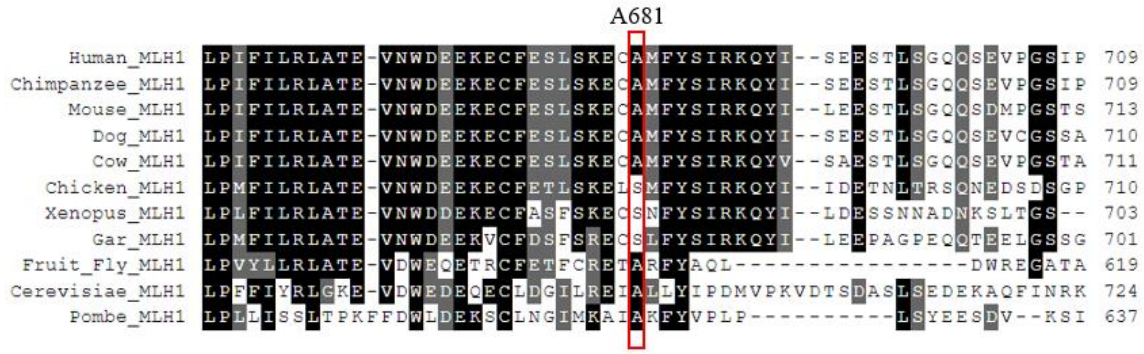


Figure 10-9: Sequence alignment of vertebrate and invertebrate MLH1 orthologues with A681 highlighted

10.10 Structural diagram of MLH1 with focus on alpha helix containing A681

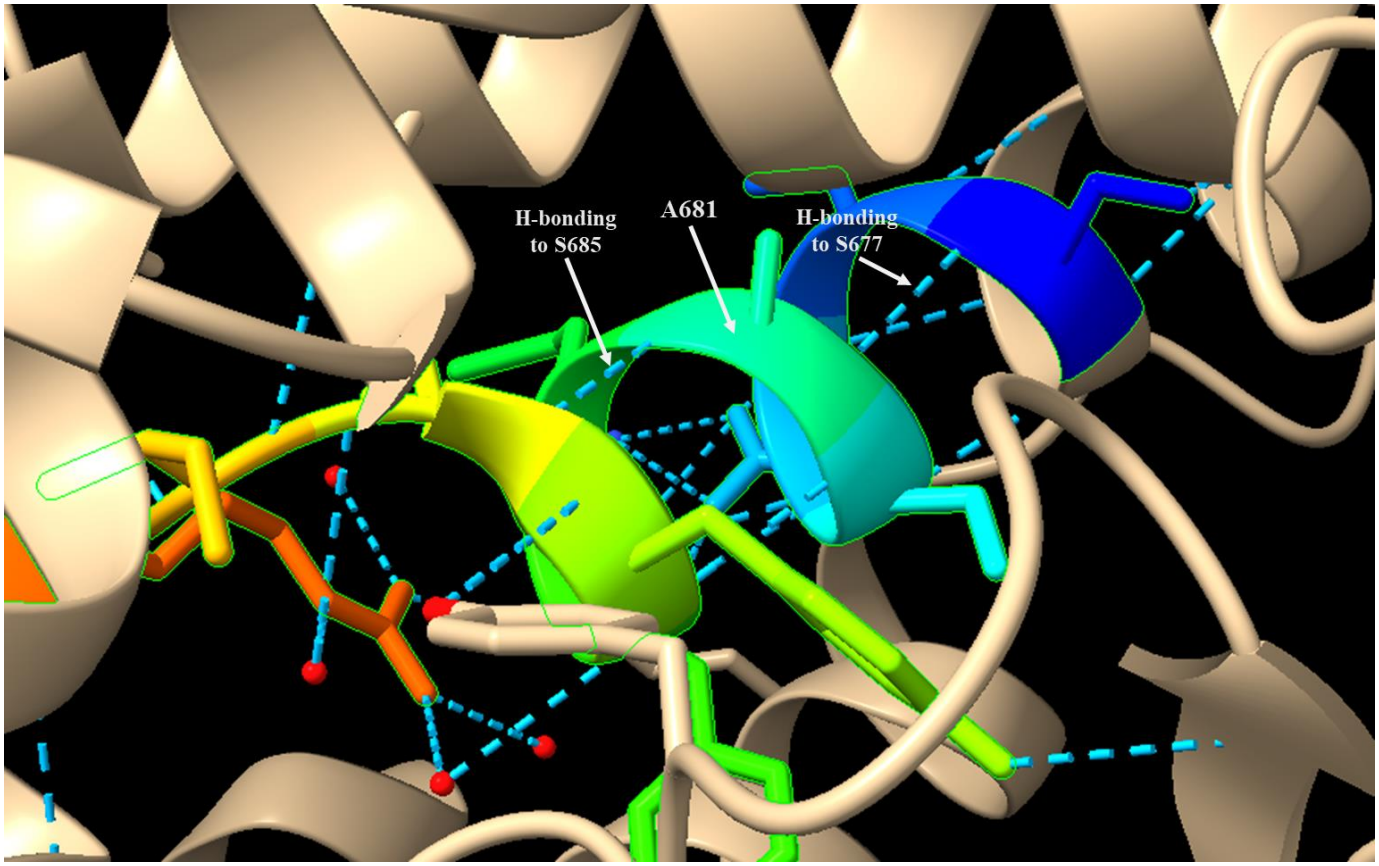
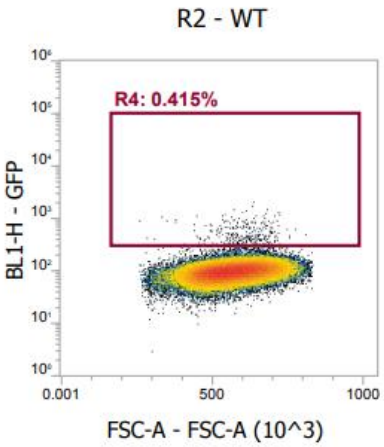


Figure 10-10: Structural prediction diagram for MLH1 with focus on alpha helix containing A681

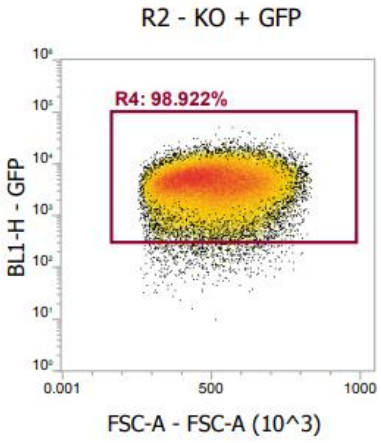
The structure of human MLH1 structure was predicted using AlphaFold (Jumper et al., 2021). A681 and the surrounding residues are highlighted. Hydrogen bonding by A681 to the surrounding residues is indicated.

10.11 FACS plots showing GFP % of stable pools of FANCL orthologue cell lines

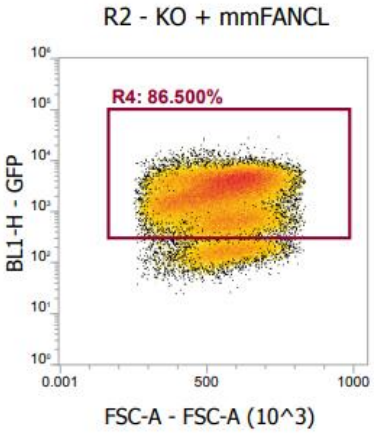
HAP1 WT (blank)



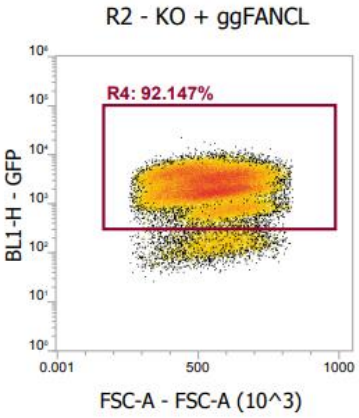
HAP1 Δ*FANCL* + GFP



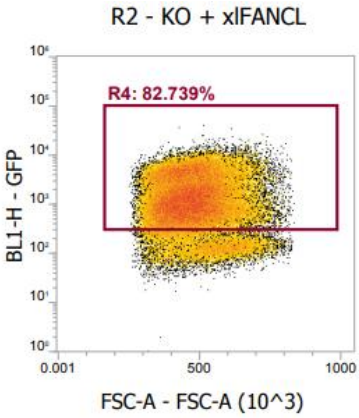
HAP1 Δ*FANCL* + mClover3-mouse FANCL



HAP1 Δ*FANCL* + mClover3-chicken FANCL



HAP1 Δ*FANCL* + mClover3-frog FANCL



HAP1 Δ*FANCL* + mClover3-fish FANCL

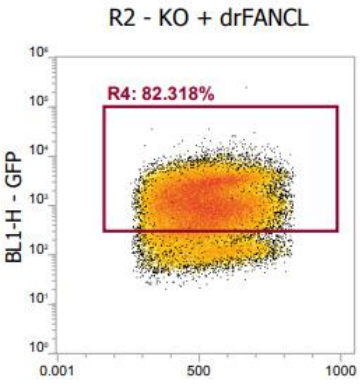


Figure 10-11: FACS plots from analysis of HAP1 Δ*FANCL* cells transduced with lentiviruses encoding either GFP or mClover3-tagged FANCL orthologues

FACS plots showing the % GFP + cells in stable pools of HAP1 Δ*FANCL* cells transduced with lentiviruses encoding either GFP or mClover3-tagged FANCL orthologues.

nature reviews cancer

https://doi.org/10.1038/s41568-022-00535-5

Review article

Check for updates

Targeting DNA damage response pathways in cancer

Florian J. Groelly[1], Matthew Fawkes[2], Rebecca A. Dagg[1], Andrew N. Blackford[2] ✉ & Madalena Tarsounas[1] ✉

Abstract

Cells have evolved a complex network of biochemical pathways, collectively known as the DNA damage response (DDR), to prevent detrimental mutations from being passed on to their progeny. The DDR coordinates DNA repair with cell-cycle checkpoint activation and other global cellular responses. Genes encoding DDR factors are frequently mutated in cancer, causing genomic instability, an intrinsic feature of many tumours that underlies their ability to grow, metastasize and respond to treatments that inflict DNA damage (such as radiotherapy). One instance where we have greater insight into how genetic DDR abrogation impacts on therapy responses is in tumours with mutated *BRCA1* or *BRCA2*. Due to compromised homologous recombination DNA repair, these tumours rely on alternative repair mechanisms and are susceptible to chemical inhibitors of poly(ADP-ribose) polymerase (PARP), which specifically kill homologous recombination-deficient cancer cells, and have become a paradigm for targeted cancer therapy. It is now clear that many other synthetic-lethal relationships exist between DDR genes. Crucially, some of these interactions could be exploited in the clinic to target tumours that become resistant to PARP inhibition. In this Review, we discuss state-of-the-art strategies for DDR inactivation using small-molecule inhibitors and highlight those compounds currently being evaluated in the clinic.

Sections

Introduction

The DDR

PARP inhibitors in targeted cancer therapy

Targeting DDR kinases in the clinic

Emerging therapies and new targets

Conclusion and perspectives

[1]Genome Stability and Tumourigenesis Group, Oxford Institute for Radiation Oncology, Department of Oncology, University of Oxford, Oxford, UK. [2]Department of Oncology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK. ✉e-mail: andrew.blackford@imm.ox.ac.uk; madalena.tarsounas@oncology.ox.ac.uk

Introduction

DNA is a relatively stable organic molecule, but our genomes are nonetheless subjected to constant assault from various sources of endogenous and exogenous damage. Cells have therefore evolved a complex system of biochemical pathways to deal with this threat, collectively termed the 'DNA damage response' (DDR)[1].

The physiological importance of the DDR in humans is highlighted by the fact that mutations in genes encoding enzymes involved in the DDR (DDR genes) are found in cancer cells and in the germ lines of individuals with a hereditary predisposition to cancer[2]. Indeed, genome instability caused by deficiencies in the DDR is a hallmark of cancer, because a greater mutational burden increases the chances of oncogene activation and loss of tumour suppressor genes, leading to tumorigenesis[3]. A more genetically diverse population of cancer cells within a tumour also increases the chances that resistant clones will arise after radiotherapy or chemotherapy, thus facilitating cancer recurrence[4].

However, the inherent genome instability of rapidly proliferating tumours also provides us with therapeutic opportunities to target DDR pathways and thus specifically kill cancer cells via additional replication stress, exogenous DNA damage and/or DDR inhibition. This probably explains in large part the early therapeutic success of radiotherapy and various DNA-damaging chemotherapeutics, such as alkylating agents and topoisomerase inhibitors[1]. However, such agents tend to be indiscriminate and damage healthy tissues, leading to substantial side effects.

Because specific genetic vulnerabilities have been identified in certain cancers, we are currently in a unique position to exploit the concept of synthetic lethality in the clinic, whereby loss of one cellular pathway results in high reliance on another pathway, which is not essential under normal settings (Tables 1, 2). This is perhaps best exemplified by the use of poly(ADP-ribose) polymerase (PARP) inhibitors to specifically target tumours deficient in the homologous recombination (HR) DNA repair factor BRCA1 or BRCA2 (BRCA1/2)[5,6] or tumours with otherwise compromised HR (Boxes 1, 2). It has subsequently become apparent that some tumours with wild type *BRCA1*/*BRCA2* genes can also be hypersensitive to PARP inhibition through mechanisms independent of HR inactivation[7]. In addition, targeting the DDR in cancer is not confined to PARP inhibition (Table 2), as other potential DDR targets have been identified in recent years; small-molecule inhibitors of several targets have been developed, and some are being tested in clinical trials (Table 1).

In this Review, we describe the major DNA damage checkpoint and repair pathways and discuss the strategies for DDR inactivation using small-molecule inhibitors that are currently being evaluated in the clinic. Furthermore, we highlight the mechanisms of DDR inhibitor drug resistance and how these might be overcome through understanding and targeting the molecular pathways underlying resistant tumours.

The DDR

DNA damage activates a signalling cascade mediated by checkpoint kinases

The DDR starts with lesion recognition and engagement of DNA repair pathways[8]. Depending on the type and complexity of genotoxic stress, cellular signalling cascades may be launched that alter the surrounding chromatin, activate cell-cycle checkpoints and modify gene expression via changes to transcription or translation. If lesions are not rapidly repaired, persistent DDR signalling can also alter cell fate by promoting differentiation, senescence or programmed cell death. The DDR thereby maintains genome stability to maximize the chances of the hereditary material being passed on to the next generation intact and unchanged.

ATM and ATR are the kinases responsible for orchestrating cellular responses to DNA double-strand breaks (DSBs) and replication stress (Fig. 1), which include DNA repair, checkpoint activation, apoptosis, senescence, and alterations in chromatin structure, transcription and pre-mRNA splicing[8]. To achieve this, they phosphorylate hundreds of substrates in response to DNA damage (see, for example, ref.[9]). The downstream cell-cycle checkpoint kinases CHK1 and CHK2 are major substrates for ATR and ATM, respectively, and are responsible for downregulating the activity of cyclin-dependent kinases (CDKs) to halt cell-cycle progression in response to genotoxic stress. This is achieved in G1 phase cells by ATM and CHK2 through phosphorylation and stabilization of p53, which results in transcription of *CDKN1A*, which encodes the CDK inhibitor p21 (ref.[10]) (Fig. 1). ATM is recruited at DSBs and promotes histone H2AX phosphorylation, which in turn recruits mediator of DNA damage checkpoint protein 1 (MDC1). In interphase cells, MDC1 recruitment and phosphorylation by ATM catalyse further phosphorylation and ubiquitylation events that result in recruitment of the DNA damage mediator proteins 53BP1 and BRCA1; however, neither 53BP1 nor BRCA1 are recruited during mitosis[11]. Instead, casein kinase 2 (CK2)-dependent MDC1 phosphorylation mediates the interaction of MDC1 and DNA topoisomerase 2-binding protein 1 (TOPBP1)[12], which is recruited to DSBs together with cellular inhibitor of PP2A (CIP2A)[13] (Fig. 1). Together, TOPBP1 and CIP2A form filamentous structures that can bridge two MDC1 foci, thus tethering both DSB ends together[12,13]. This ensures correct chromosome segregation until DSBs are repaired in the following G1 phase.

In S and G2 phases of the cell cycle, ATR and its downstream target CHK1 are primarily responsible for the checkpoint activation via phosphorylation and inactivation of the CDC25 family of phosphatases[14–16]. These remove the inhibitory phosphorylation from CDK1 and CDK2, deposited by the kinase WEE1 (refs.[17,18]) and the WEE1-like kinase PKMYT1 (refs.[19,20]) (Fig. 1). Moreover, ATR controls the S–G2 checkpoint that prevents entry into mitosis before replication is completed[21,22]. In contrast to ATM, which is activated by binding of the MRE11–RAD50–NBS1 (MRN) complex to DSBs[23,24], ATR is recruited by ATR-interacting protein (ATRIP), which binds to replication protein A (RPA)-coated single-stranded DNA (ssDNA) generated at stalled replication forks[25]. ATR is then activated by the replication stress response protein ETAA1 (refs.[26–28]) or TOPBP1 (ref.[29]), which bind to RPA at ssDNA or to the RAD9–RAD1–HUS1 (9-1-1) complex at ssDNA–double-stranded DNA junctions, respectively[8]. Therefore, ATR is activated in S phase in response to a broad range of genotoxic lesions, including stalled replication forks, leading or lagging strand ssDNA gaps or resected DSBs[21,22].

Multiple DNA repair pathways deal with different DNA lesions

The most common types of DNA damage are lesions where only one strand of the DNA double helix is affected, either by single-strand breaks (SSBs) in the phosphate backbone or by chemical modification of DNA bases[30]. Most such lesions are recognized and repaired by the base excision repair or nucleotide excision repair pathways or can be bypassed entirely during replication via translesion synthesis. During DNA replication, the wrong nucleotide can be incorporated, resulting in two mismatched bases; these are dealt with by the mismatch repair (MMR) pathway[31]. Alternatively, ribonucleotides can be incorporated into the genome instead of deoxyribonucleotides; these are excised

Table 1 | Selected clinical trials for DDR inhibitors

Target	Agent[a]	Combination	Phase	Cancer types	Biomarkers	Clinical trial ID[b] or reference
ATM	AZD0156	Monotherapy, olaparib, irinotecan or FOLFIRI	I	ASTs	–	NCT02588105
	AZD1390	Radiotherapy	I	Grade IV glioma	–	NCT05182905
			I	Brain	–	NCT03423628
			I	Soft tissue sarcoma	–	NCT05116254
			I	Lung	–	NCT04550104
	M4076	–	I	ASTs	–	NCT04882917
ATM and DNA-PKcs	XRD-0394	Radiotherapy (palliative)	I	ASTs, MSTs, RSTs	–	NCT05002140
ATR	ART0380	Monotherapy, gemcitabine or irinotecan	I/II	ASTs, MSTs	ATM deficiency	NCT04657068
	ATRN-119	–	I/II	ASTs	DDR gene mutations	NCT04905914
	BAY1895344	–	I	ASTs, lymphomas	–	NCT03188965 (ref.[196])
		Niraparib	I	ASTs, ovarian	–	NCT04267939
		Pembrolizumab	I	Solid tumours	DDR gene mutations	NCT04095273
	Berzosertib (VX-970, M6620, VE-822)	Veliparib or cisplatin	I	Solid tumours	–	NCT02723864
		Various chemotherapies	I	ASTs	–	NCT02157792
						Yap et al.[196]
		Gemcitabine±berzosertib	II	Ovarian	–	NCT02595892 (ref.[197])
		Topotecan	II	Lung	–	NCT02487095 (ref.[134])
		Avelumab	I/II	Solid tumours	DDR gene mutations	NCT04266912
	Ceralasertib (AZD6738)	Olaparib	II	Gynaecological	–	NCT04065269
		Durvalumab	I	Head and neck, lung	–	NCT02264678
	IMP9064	–	I	ASTs	–	NCT05269316
	M4344	Monotherapy or carboplatin	I	ASTs	*ARID1A*, *ATRX*, *DAXX* or *ATM* mutations	NCT02278250
	RP-3500	Monotherapy or talazoparib+gemcitabine	I/II	ASTs	–	NCT04497116
		Olaparib	I/II	CLL	*TP53*, *ATM*, *SF3B1*, *XPO1* or *POT1* mutations	NCT05405309
		Olaparib or niraparib	I/II	ASTs	–	NCT04972110
		RP-6306	I	ASTs	–	NCT04855656
CHK1	Prexasertib (LY2606368)	Irinotecan	I/II	DSRCT, rhabdomyosarcoma	–	NCT04095221
		–	II	Lung	–	NCT02735980
		–	II	Solid tumours	*MYC* or *CCNE1* amplification, *RB* loss or *FBXW7*, *BRCA1*/*BRCA2*, *PALB2*, *RAD51C*, *RAD51D*, *ATR*, *ATM*, *CHK2* or Fanconi anaemia gene mutations	NCT02873975
		–	II	Ovarian	*BRCA1*/*BRCA2* mutations	NCT03414047
		–	II	Breast, ovarian, mCRPC	*BRCA1*/*BRCA2* mutations	NCT02203513
	SRA737	Gemcitabine±cisplatin	I/II	ASTs	Predicted sensitivity to CHK1 inhibition[c]	NCT02797977
		–	I/II	ASTs, NHL	Predicted sensitivity to CHK1 inhibition[c]	NCT02797964
	LY2880070	Gemcitabine	I/II	ASTs, MSTs	–	NCT02632448
			II	Ewing sarcoma	–	NCT05275426

Table 1 (continued) | Selected clinical trials for DDR inhibitors

Target	Agent[a]	Combination	Phase	Cancer types	Biomarkers	Clinical trial ID[b] or reference
DNA-PKcs	AZD7648	Radiotherapy	I	Soft tissue sarcoma	–	NCT05116254
		Monotherapy or PLD	I	ASTs	–	NCT03907969
	M3814	–	I	ASTs, CLL	–	NCT02316197
		Lutetium Lu 177 dotatate	I	Neuroendocrine	–	NCT04750954
		Radiotherapy	I/II	Pancreatic	–	NCT04172532
		Radiotherapy + avelumab	I/II	ASTs, MSTs	–	NCT04068194
		Radiotherapy + capecitabine	I/II	Rectal	–	NCT03770689
		Radiotherapy + temozolomide	I	Glioblastoma	–	NCT04555577
		Radium-223 dichloride ± avelumab	I/II	Prostate	–	NCT04071236
DNA-PKcs and mTOR	CC-115	Enzalutamide	I	Prostate	–	NCT02833883
		–	I	ASTs	–	NCT01353625
	Samotolisib[d] (LY3023414)	Prexasertib	I	ASTs, MSTs	*PIK3CA* mutations	NCT02124148 (ref.[198])
		–	II	ASTs, MSTs, NHL	*TSC1*, *TSC2*, *PIK3CA* or *MTOR* mutations	NCT03155620
		–	II	AST, NHL, CNS tumours	*TSC1*, *TSC2*, *PI3K* or *MTOR* mutations	NCT03213678
		–	II	Metastatic breast cancers	–	NCT04032080
		–	II	Endometrial	–	NCT02549989
PKMYT1	RP-6306	FOLFIRI	I	ASTs	–	NCT05147350
		Gemcitabine	I	–	–	NCT05147272
		RP-3500	I	–	–	NCT04855656
WEE1	Adavosertib (AZD1775)	Gemcitabine, paclitaxel, carboplatin or PLD	II	Ovarian	-	NCT02272790 (ref.[199])
		Radiotherapy + gemcitabine	I	Pancreatic	–	NCT02037230 (ref.[200])
		–	II	ASTs, RSTs, lymphomas, plasma cell myeloma	*BRCA1*/*BRCA2* mutations	NCT04439227
		–	II	ASTs, MSTs	*SETD2* mutations	NCT03284385
	IMP7068	–	I	ASTs	–	NCT04768868

AST, advanced solid tumour; CLL, chronic lymphocytic leukaemia; DDR, DNA damage response; DNA-PKcs, DNA-dependent protein kinase catalytic subunit; DSRCT, desmoplastic small round cell tumour; FOLFIRI, folinic acid, fluorouracil and irinotecan; mCRPC, metastatic castration-resistant prostate cancer; MST, metastatic solid tumour; RST, refractory solid tumour; NHL, non-Hodgkin lymphoma; PLD, pegylated liposomal doxorubicin. [a]Other names for agents are given in parentheses where applicable. [b]Clinical trials are accessible at https://clinicaltrials.gov/. [c]Predicted sensitivity to CHK1 inhibition includes DDR gene mutations, loss of tumour suppressor genes involved in G1 cell cycle progression (*RB* and *TP53*), gain of functions or amplification of oncogenic drivers (for example, *CCNE1*), amplification of *CHK1* or *ATR*, and human papilloma virus positivity. [d]Samotolisib also inhibits PI3K, whereas CC-115 does not.

by ribonuclease H2 (RNase H2)[32]. Occasionally, the two strands of the double helix can become crosslinked; depending on their chemical nature, such DNA interstrand crosslinks can be repaired by the Fanconi anaemia pathway, the DNA glycosylase NEIL3 or other incompletely characterized interstrand crosslink repair pathways[33]. Proteins can also become covalently trapped on DNA, requiring DNA–protein crosslink repair mediated by specialized proteases such as SPRTN[34].

SSBs, or single-strand nicks, are primarily recognized by the enzyme PARP1 or PARP2, which catalyse the formation of poly(ADP-ribose) (PAR) chains on themselves and adjacent target proteins[30]. PARP1 and its poly(ADP-ribosyl)ation (PARylation) activity at SSBs recruit the scaffold protein XRCC1, which brings DNA ligase 3 (LIG3) and accessory repair factors to re-ligate the break. PARylation is a highly dynamic and often transient process, because PAR chains can be degraded rapidly by poly(ADP-ribose) glycohydrolase (PARG). Rather than simply being a negative regulator of SSB repair, PARG activity restores PARPs and PARylated proteins to their de-(ADP-ribosylated) state to promote subsequent rounds of SSB repair. Thus, PARG inhibition, similarly to PARP inhibition, reduces the rate of SSB repair[35].

While SSBs are the most common DNA lesions, they are relatively easily repaired. By contrast, DNA DSBs, which pose a higher threat to genome integrity, are far more difficult to repair[8]. There are two major DSB repair pathways in human cells (Fig. 2). The first is the non-homologous end joining (NHEJ) pathway, which repairs the vast majority of two-ended DSBs. By contrast, if DNA replication forks collapse during S phase to form one-ended DSBs, NHEJ is toxic because it can

Table 2 | Selected clinical trials for PARP inhibitors

Agent[a]	Combination	Phase	Cancer type	Biomarkers	Clinical trial ID[b]
Monotherapy					
Fuzuloparib	–	III	Ovarian	–	NCT03863860
Olaparib	–	II	Breast	–	NCT04191135
		III	Lung	–	NCT04624204
Pamiparib (BGB-290)	–	III	Ovarian	–	NCT03519230
Rucaparib	–	III	Ovarian	Wild-type *BRCA1*/*BRCA2*	NCT04227522
Monotherapy in *BRCA1*/*BRCA2*-mutated cancers					
Fuzuloparib	–	III	Pancreatic	*BRCA1*/*BRCA2* mutation or *PALB2* mutation	NCT04300114
IMP4297	–	II	Ovarian	*BRCA1*/*BRCA2* mutation	NCT04089189
		III	Ovarian	*BRCA1*/*BRCA2* mutation	NCT04169997
Rucaparib	–	III	Ovarian	*BRCA1*/*BRCA2* mutation	NCT02855944 (ref.[201])
In prostate cancer					
AZD5305	Enzalutamide, abiraterone or darolutamide	I/II	Metastatic prostate cancer	–	NCT05367440
Fuzuloparib	Abiraterone and prednisone	II	mCRPC	–	NCT04691804
Olaparib	Abiraterone	III	Prostate	–	NCT03732820 NCT05171816
Rucaparib	Enzalutamide	III	Prostate	–	NCT04455750
Talazoparib	Enzalutamide	III	Prostate	–	NCT03395197
In DDR-deficient prostate cancer					
Fuzuloparib	Monotherapy or apatinib	II	mCRPC	HR gene mutations	NCT04869488
IMP4297	–	II	mCRPC	HR gene mutations	NCT04822961
Niraparib	Abiraterone and prednisone	III	Prostate	HR gene mutations	NCT04497844 NCT03748641
Rucaparib	–	III	Prostate	*BRCA1*/*BRCA2* mutation or *ATM* mutation	NCT02975934
Talazoparib	Enzalutamide	III	Prostate	DDR gene mutations	NCT04821622
In combination with chemotherapy					
AZD5305	Various chemotherapies	I/II	Solid tumours	–	NCT04644068
Olaparib	Paclitaxel	III	Gastric	–	NCT01924533 (ref.[202])
Veliparib	Carboplatin and paclitaxel	III	Lung	–	NCT02106546 (ref.[203])
			Lung	–	NCT02264990 (ref.[204])
			Breast	–	NCT02032277 (ref.[205])
			Breast	*BRCA1*/*BRCA2* mutation	NCT02163694 (ref.[206])
			Ovarian	HR deficiency or *BRCA1*/*BRCA2* mutation	NCT02470585 (ref.[207])
Fuzuloparib	Temozolomide	II	Glioblastoma	–	NCT04552977
Veliparib	Temozolomide	II/III	Glioblastoma	*MGMT* promoter hypermethylation	NCT02152982
In combination with immunotherapy					
Niraparib	Bevacizumab±dostarlimab	III	Ovarian	–	NCT03602859
	Pembrolizumab	III	Lung	–	NCT04475939 (refs.[203,208])
Olaparib	Pembrolizumab	III	Lung	–	NCT03976362
					NCT03976323
					NCT04380636
Pamiparib (BGB-290)	Tislelizumab	I	Advanced solid tumours	HR deficiency or *BRCA1*/*BRCA2* mutation	NCT02660034 (ref.[209])
Pamiparib (BGB-290)	Tislelizumab or temozolomide	III	Advanced solid tumours	–	NCT04164199
Rucaparib	Nivolumab	III	Ovarian	–	NCT03522246

DDR, DNA damage response; HR, homologous recombination; mCRPC, metastatic castration-resistant prostate cancer; PARP, poly(ADP-ribose) polymerase. [a]Other names for agents are given in parentheses where applicable. [b]Clinical trials are accessible at https://clinicaltrials.gov/.

Box 1

Current clinical applications of PARP inhibitors

Since the discovery of the synthetic-lethal interaction between poly(ADP-ribose) polymerase (PARP) inhibition and BRCA1/2 deficiencies[5,6], multiple studies have demonstrated the clinical benefits of PARP inhibitors, in particular for patients with *BRCA1*/*BRCA2*-mutated tumours[4]. Consequently, six different PARP inhibitors have been approved for clinical use, including against specific subsets of *BRCA1*/*BRCA2*-mutated cancers: olaparib, rucaparib, niraparib, talazoparib, fuzuloparib and pamiparib. For example, the US Food and Drug Administration (FDA) has approved olaparib and talazoparib to treat advanced or metastatic HER2-negative breast cancers in patients carrying deleterious germ line *BRCA1*/*BRCA2* mutations[211,212]. In March 2022, olaparib was approved by the FDA for adjuvant treatment of patients with inherited *BRCA1*/*BRCA2* mutations and HER2-negative high-risk early breast cancer[213]. PARP inhibitors are also used in patients with germ line or somatic *BRCA1*/*BRCA2*-mutated ovarian cancer, as maintenance treatment (olaparib)[214] or treatment after chemotherapy (olaparib and rucaparib)[201]. Moreover, olaparib is used for maintenance treatment of patients with germ line *BRCA1*/*BRCA2*-mutated metastatic pancreatic cancers[215]. Rucaparib is used as second-line treatment for patients with germ line or somatic *BRCA1*/*BRCA2*-mutated metastatic castration-resistant prostate cancer[216].

The clinical applications of PARP inhibitors are not restricted to *BRCA1*/*BRCA2*-mutated tumours. Olaparib and niraparib are available to target advanced ovarian cancers associated with homologous recombination deficiency defined not only by *BRCA1*/*BRCA2* mutations but also by genomic instability evaluated with an approved companion diagnostic test[217,218] (Box 2). Moreover, olaparib has been approved for treatment of patients with metastatic castration-resistant prostate cancer whose tumours have progressed following anti-androgen therapy and who harbour germ line or somatic mutations in homologous recombination genes, including *BRCA1*, *BRCA2*, *PALB2*, *RAD51C*, *RAD51D* and *ATM*[219]. It is noteworthy that PARP inhibitors have also been approved for maintenance therapy in patients with recurrent or advanced ovarian cancers (olaparib, rucaparib and niraparib) and who were not selected on the basis of known *BRCA1*/*BRCA2* mutations or homologous recombination deficiencies[220–223].

These inhibitors have also been approved by the European Medicines Agency (EMA) to treat some of the tumour types mentioned above. Moreover, the China National Medical Products Administration (NMPA) has approved the use of PARP inhibitors to treat germ line *BRCA1*/*BRCA2*-mutated recurrent ovarian cancers (fuzuloparib)[224] or germ line *BRCA1*/*BRCA2*-mutated recurrent advanced ovarian cancers (pamiparib)[225].

However, the emergence of resistance to PARP inhibitors limits their clinical efficacy (see Box 2). Because PARP trapping mediates the antitumour activity of PARP inhibitors and acquired resistance[93,189], modulation of their trapping ability may affect their clinical efficacy. Therefore, novel PARP inhibitors with increased PARP trapping capacity, such as the recently developed veliparib derivative[226], are promising clinical candidates.

generate chromosomal rearrangements by re-ligating DNA ends on different chromosomes. Thus, NHEJ is actively suppressed at replication forks by components of the second major DSB repair pathway, HR. HR is also favoured in specific contexts, such as interstrand crosslink repair by the Fanconi anaemia pathway[36] or repair of programmed DSBs induced during meiosis[37], when NHEJ would be similarly disruptive.

NHEJ is often referred to as error-prone (for example, in the context of gene editing via CRISPR–Cas9[38]) because it does not use a homologous template for repair. In reality, NHEJ is remarkably efficient and accurate most of the time[8], which explains why cells have evolved to use it preferentially. The pathway is initiated when Ku, a basket-shaped heterodimeric protein complex consisting of Ku70 and Ku80 subunits[39], recognizes and binds broken DNA ends (Fig. 2). Ku primarily acts as a recruitment platform for downstream NHEJ components, while simultaneously protecting DNA ends from unwinding and degradation by cellular helicases and nucleases, respectively. Ku-bound DNA ends rapidly recruit the DNA-dependent protein kinase (DNA-PK) catalytic subunit (DNA-PKcs) to form the DNA-PK holoenzyme, which is essential for NHEJ in part by promoting DNA-end tethering and thus facilitating DNA ligation. Interestingly, the best-known substrate for DNA-PK kinase activity in NHEJ is DNA-PKcs itself. This autophosphorylation is nonetheless essential for the repair process because it promotes DNA-end synapsis[40] and DNA-PK dissociation from DNA ends[41,42]. DNA-end ligation is subsequently performed by LIG4 and its stable binding partner XRCC4 (refs.[43,44]), a step which also requires either XLF or PAXX, two partially redundant NHEJ core factors[45–49]. NHEJ accessory factors, including specialized nucleases (for example, Artemis) and polymerases (for example, DNA polymerase-μ and DNA polymerase-λ), also promote accurate DSB repair in specific circumstances but are not essential for all NHEJ repair[50].

In contrast to NHEJ, the HR pathway uses a homologous DNA molecule (usually the sister chromatid) as a template for repair. HR is initiated when nucleases digest the double-stranded DNA ends at DSB sites to produce ssDNA overhangs (Fig. 2), a process termed 'DNA-end resection'[51]. This results in removal of Ku from DNA ends and production of tracts of ssDNA, which become rapidly coated by RPA. DNA-end resection is tightly controlled during the cell cycle, such that it occurs only during S and G2, due to CDK-dependent phosphorylation of CtIP (also known as RBBP8), a factor known to stimulate end resection[52,53]. This also prevents diploid cells from using the homologous chromosome rather than the sister chromatid as a template for repair (which could otherwise lead to loss of heterozygosity). Exceptions may occur at repetitive regions of the genome, such as ribosomal or centromeric DNA, where resection and HR can occur even in G1 phase cells[54,55].

Two distinct but possibly redundant pathways promote end resection after initial ssDNA overhang production by CtIP and the MRN complex. The first pathway depends on the nuclease EXO1, while the second requires the nuclease DNA2 and the helicase BLM[56–58]. After resection, RPA is replaced on ssDNA with the recombinase RAD51, a step that requires BRCA1 and BRCA2 (ref.[36]). This is because BRCA1,

via its interaction with PALB2, promotes BRCA2 recruitment to DSB sites, where the latter loads RAD51 directly onto the ssDNA ends. RAD51 forms nucleoprotein filaments on ssDNA, which promotes strand invasion and displacement loop (D-loop) formation. This DNA structure allows DNA synthesis across the break site using the sister chromatid as the template so as to prevent alteration of genetic information. Finally, the invading strand can be displaced and HR completed via synthesis-dependent strand annealing. Alternatively, if second-end capture occurs, a double Holliday junction is formed, which is either dissolved by the Bloom syndrome complex (containing BLM and the topoisomerase TOP3A) or resolved by SLX4–MUS81 or GEN1, structure-specific nucleases[59]. The extent of DNA-end resection determines the use of alternative DNA repair pathways (Fig. 2), most notably single-strand annealing and alternative end joining (also known as DNA polymerase-θ (POLQ)-mediated end joining or microhomology-mediated end joining), which are reliant on RAD52 and POLQ, respectively. Single-strand annealing and alternative end joining act mainly during S phase and mitosis[60], respectively, as backup repair pathways when HR is inactivated (discussed later).

Box 2

Biomarkers and functional assays for patient selection

Patient stratification has a key role in the development and clinical application of targeted antitumour therapies, and novel biomarkers are being used to determine whether a patient will benefit from a given treatment. The presence of germ line or somatic mutations is a major criterion for patient selection, as illustrated by the use of poly(ADP-ribose) polymerase (PARP) inhibitors to target *BRCA1*-, *BRCA2*- or *ATM*-mutated tumours (Box 1, Tables 1, 2). Other companion diagnostic tests that measure complex 'genomic scars' are also being used to select eligible patients. For instance, one companion diagnostic test measures loss of heterozygosity, telomeric allelic imbalance and large-scale state transitions and calculates a genomic instability score[227].

As an alternative to genetic tests, functional assays measuring the formation of RAD51 foci, as a surrogate for homologous recombination activity, have been shown to predict PARP inhibitor response ex vivo. Indeed, in cells derived from patients with epithelial ovarian cancers that were challenged with the PARP inhibitor rucaparib ex vivo, the formation of RAD51 foci correlated with cell survival in congenic assays[228]. Recent studies demonstrated the possibility to perform formation of RAD51 foci assays in formalin-fixed, paraffin-embedded tumour samples, thereby removing the need to isolate live cancer cells and treat them ex vivo[229,230]. Functional homologous recombination assays in formalin-fixed, paraffin-embedded tumour samples from patients with germ line *BRCA1*/*BRCA2* mutations and treated with PARP inhibitors predicted clinical efficacy and drug resistance[230]. Altogether, the formation of RAD51 foci is a promising tool for the selection of patients who may benefit from PARP inhibitors.

Genetic and functional assays are likely to become routine for selection of patients for DNA damage response inhibitor therapy in the clinic. However, functional assays are available only for PARP inhibitors, and future efforts should focus on the development of functional assays predicting the efficacy of other DNA damage response-targeting therapies.

Chromatin modifications at DNA damage sites control DNA repair pathway choice

While cell cycle-dependent phosphorylation events (for example, CtIP phosphorylation by CDKs) enable initiation of DNA resection and HR during S and G2 phases of the cell cycle, NHEJ remains the predominant DSB repair pathway throughout interphase[61,62]. This is due, in part, to DSB repair regulation at the chromatin level by BARD1, the stable binding partner of BRCA1, and 53BP1, which compete to promote and antagonize end resection, respectively. BARD1 and 53BP1 act as readers of histone H2A Lys15 (H2AK15) ubiquitylation, a DNA damage-induced histone modification deposited by RNF168 (Fig. 1), and of histone H4 Lys20 (H4K20) methylation status. Histone H4 Lys20 dimethylation is a histone modification deposited on mature chromatin and is thus absent from the nascent daughter DNA strand during S phase[63,64]. By directly binding to nucleosomes containing H2AK15ub[65] and H4K20me2 (ref.[66]), 53BP1 recruits downstream effector proteins, including RIF1 and the shieldin complex, to counteract DNA-end resection by modulating 3D chromatin structure, protecting DSB ends and/or recruiting DNA polymerase-α to fill in ssDNA gaps generated during resection[67–70]. Thus, by inhibiting the production of ssDNA, 53BP1 and its binding partners ensure that RAD51 filaments of sufficient length cannot be assembled, thereby suppressing HR.

53BP1-mediated NHEJ is highly toxic during S phase, when collapsed replication forks lead to the production of one-ended DSBs[71]. When two such ends from different chromosomes are ligated, chromosomal translocations arise. Thus, DSBs that occur during replication must be repaired by HR rather than NHEJ. This is achieved by the BRCA1–BARD1 complex, which, like 53BP1, recognizes the RNF168-mediated histone H2A Lys15 ubiquitylation mark (Fig. 1), but binds unmethylated histone H4 Lys20 rather than dimethylated histone H4 Lys20 (refs.[72–75]). Importantly, nascent chromatin in the vicinity of a replication fork is enriched in unmethylated histone H4 Lys20, in contrast to mature chromatin, which becomes progressively methylated on H4K20 after replication. In this way, cells ensure that 53BP1 promotes NHEJ throughout interphase at DSBs that occur in mature chromatin, whereas BRCA1 is recruited to chromatin proximal to DSBs that occur during DNA replication and are repaired by HR. However, it is important to note that 53BP1 is still recruited to more distal chromatin sites at DSBs even in S phase. This contributes to maintaining HR accuracy by preventing extensive long-range resection that could otherwise activate single-strand annealing (Fig. 2), a RAD52-mediated repair process that produces large genomic deletions[76]. If DNA ends are not extensively resected, then alternative end joining (Fig. 2) can be used to repair DSBs[77]. However, the alternative end joining and single-strand annealing pathways are both highly mutagenic and are used as backups only when the more accurate repair pathways (NHEJ or HR) are compromised. This model is supported by recent evidence showing that POLQ-dependent DSB repair is restricted until the onset of mitosis, before cell division occurs[60]. This last-resort mutagenic DSB repair is preferable to unrepaired DSBs being carried into mitosis, where chromosome segregation could lead to micronuclei and catastrophic events such as breakage–fusion–bridge cycles and chromothripsis, both sources of chromosomal instability in cancer[11].

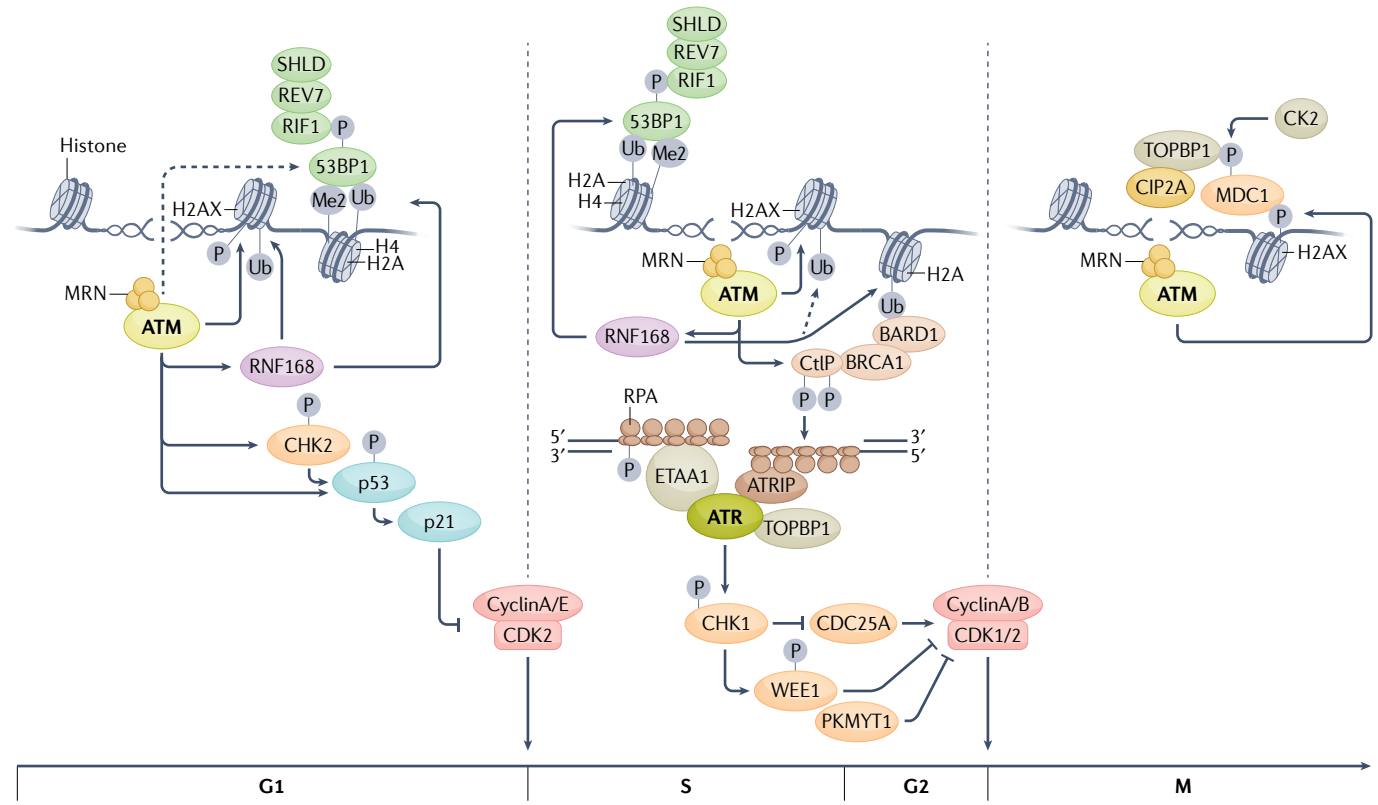


Fig. 1 | Cell cycle-dependent DDR activation at DSBs. The MRE11–RAD50–NBS1 (MRN) complex senses DNA double-strand breaks (DSBs) and recruits and activates ATM at break sites to orchestrate DNA damage response (DDR) signalling. Left: During G1, ATM promotes activation of RNF168, which ubiquitylates histone H2A. This modification, together with the histone H4 Lys20 dimethylation mark leads to the recruitment of 53BP1 and the ATM-mediated phosphorylation of 53BP1, which promotes its interaction with RIF1, as well as REV7 and other members of the shieldin complex (SHLD). ATM kinase activity also controls p53 stability, which triggers G1 arrest in response to DNA damage. Centre: During S phase, the BRCA1–BARD1 complex is recruited to DSBs and counteracts 53BP1 via recognition of unmethylated H4 Lys20 and H2A Lys15 ubiquitylation to promote DNA-end resection and RAD51 loading onto resected DNA ends. Replication protein A (RPA)-coated single-stranded DNA also recruits ATR, via interaction with ATR-interacting protein (ATRIP). ETAA1 binding to RPA-coated single-stranded DNA or topoisomerase 2-binding protein 1 (TOPBP1) binding to the RAD9–RAD1–HUS1 (9-1-1) complex (not shown) at double-stranded DNA–single-stranded DNA junctions activates ATR, which phosphorylates CHK1 to promote CDC25A degradation, thus preventing activation of cyclin-dependent kinase 1 (CDK1) and CDK2. The kinases WEE1 (a phosphorylation target of CHK1) and PKMYT1 mediate inhibitory phosphorylation of CDK1 and CDK2, which prevents cells progressing into mitosis. Right: In mitosis, ATM-dependent H2AX phosphorylation recruits mediator of DNA damage checkpoint protein 1 (MDC1), which binds TOPBP1 and cellular inhibitor of PP2A (CIP2A) via its casein kinase 2 (CK2)-mediated phosphorylation. TOPBP1 and CIP2A form filamentous structures which tether both DSB ends together (not shown). Me2, dimethylation; P, phosphorylation; Ub, ubiquitylation.

PARP inhibitors in targeted cancer therapy

Germ line heterozygous mutations in *BRCA1* or *BRCA2* predispose affected individuals to multiple types of cancer, including breast, ovarian, prostate and pancreatic cancer[78,79]. As discussed earlier herein, BRCA1 and BRCA2 promote accurate DNA repair by the HR pathway, which is essential for preserving genome integrity following exposure to DNA-damaging treatments such as ionizing radiation or DNA interstrand crosslinking agents. In the absence of exogenous challenges, BRCA1 and BRCA2 are crucial to maintaining genome stability during DNA replication[36]. Although not required for replication per se, BRCA1 and BRCA2 facilitate it by preventing nascent DNA degradation by nucleases and promoting HR repair of broken replication forks[80–83]. When BRCA1 or BRCA2 is abrogated, DNA lesions caused by failed replication accumulate. Their repair by error-prone backup pathways generates genomic alterations (insertions, deletions and chromosomal rearrangements[84]) that pose a threat to genome integrity and, in the long term, promote tumorigenesis[85,86].

As discussed earlier herein, PARP plays a key role in SSB repair. It was therefore originally postulated that PARP inhibition specifically kills BRCA1/2-deficient cells due to accumulation of SSBs, which are converted into DSBs when they are encountered by replication forks. Replication-associated DSBs are one-ended and must therefore be repaired by HR, as NHEJ can produce toxic genomic rearrangements. The exquisite hypersensitivity of BRCA1/2-deficient and, more generally, HR-deficient cells to PARP inhibition[5,6] was thought to be due to suppression of PARP-dependent SSB repair, leading to accumulation of DNA lesions (SSBs and DSBs) during replication. However, this model does not completely explain the observation that the degree to which BRCA1/2-deficient cells are sensitive to PARP inhibitors correlates with the extent to which they 'trap' PARP on DNA[87]. Indeed, loss or inhibition

of factors such as ALC1 (also known as CHD1L), RNase H2 or the ATPase p97 (also known as VCP), which increase PARP trapping, are synthetic lethal with BRCA1/2 loss and increase sensitivity to PARP inhibitors[88–93]. Furthermore, recent studies have demonstrated that PARP inhibition leads to accumulation of ssDNA gaps behind the replication fork, which originate from defects in processing Okazaki fragments[94–96]. Progression through mitosis and the subsequent S phase are required for conversion of these gaps into DSBs that require BRCA1/2 for repair[97,98].

Further work is required to establish the precise mechanism of PARP inhibitor-induced lethality in BRCA1/2-deficient cells. However,

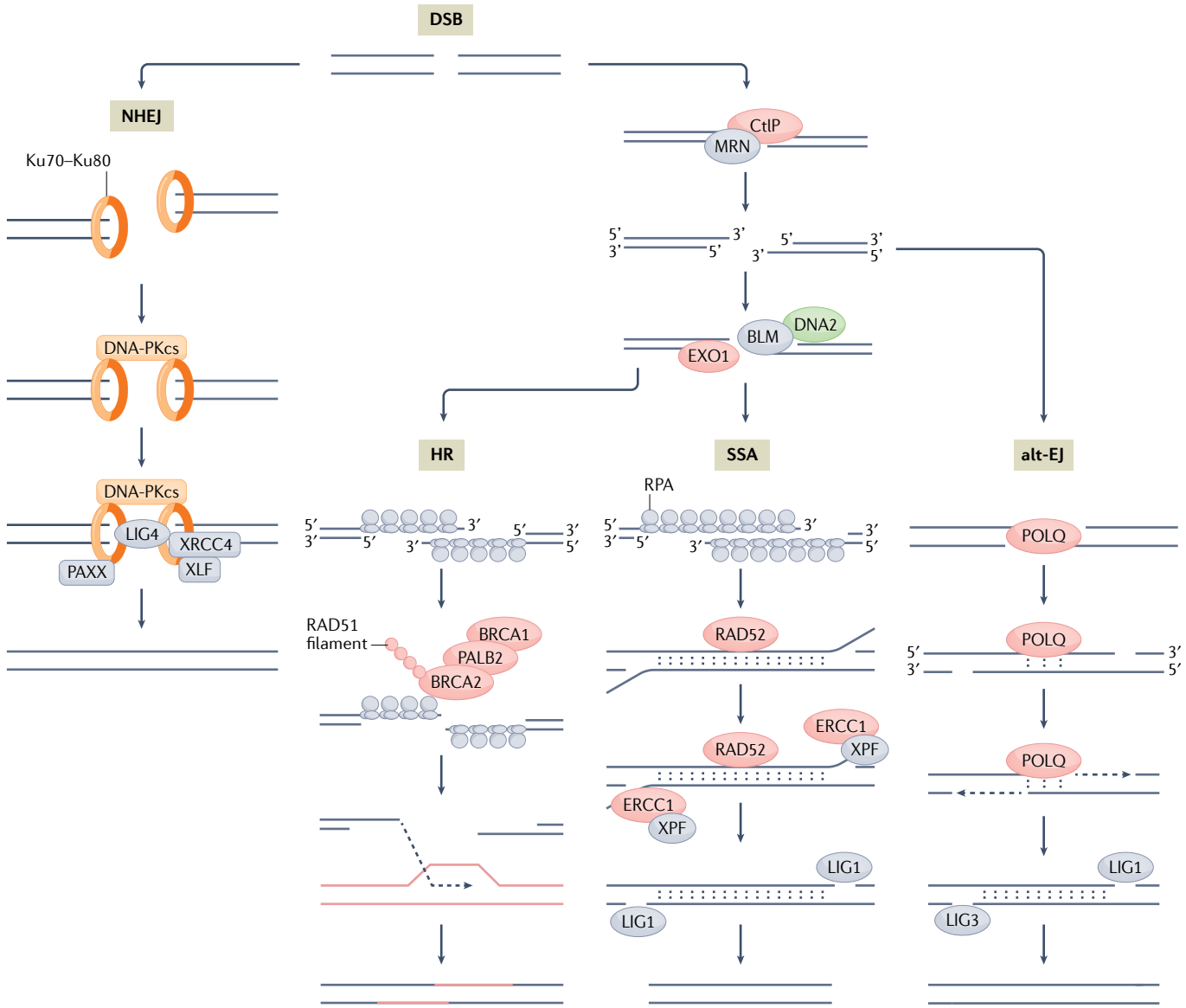


Fig. 2 | Major and backup pathways of DSB repair. DNA double-strand break (DSB) repair pathway choice depends on the cell-cycle stage and the extent of DNA-end resection. Most DSBs are repaired by non-homologous end joining (NHEJ), initiated by binding of Ku70–Ku80 heterodimer to DNA ends. The subsequent recruitment and autophosphorylation of DNA-dependent protein kinase catalytic subunit (DNA-PKcs) brings the DNA ends together and allows their ligation by XRCC4–DNA ligase 4 (LIG4), together with either XLF or PAXX. During S phase, MRE11 as part of the MRE11–RAD50–NBS1 (MRN) complex and CtIP initiate resection to promote homology-directed repair. Long-range resection requires BLM–DNA2 helicase–nuclease or EXO1 nuclease activities and leads to the formation of replication protein A (RPA)-coated single-stranded DNA overhangs. Homologous recombination (HR) requires the concerted action of BRCA1, PALB2 and BRCA2 to load RAD51 onto single-stranded DNA, displacing RPA. RAD51 nucleoprotein filaments promote invasion into the sister chromatid (pink), which serves as a template for DNA synthesis. Alternatively, extensive resection creates a substrate for single-strand annealing (SSA), where RAD52 promotes annealing of homologous sequences on each single-stranded DNA end. The subsequent processing of 3′ single-stranded flaps by ERCC1 and XPF allows LIG1-mediated DNA ligation. RAD52 is also involved in break-induced replication (not shown), which is distinct from SSA and may contribute to the genomic alterations in cells defective for HR[210]. By contrast, short-range resection is sufficient to activate alternative end joining (alt-EJ), where DNA polymerase-θ (POLQ) promotes the annealing of short homologous sequences, followed by DNA synthesis and LIG1- or LIG3-dependent re-ligation of DNA ends.

the discovery of the synthetic-lethal relationship between PARP inhibition and BRCA1/2 deficiency has established a paradigm for selective targeting of DDR factors in tumours with HR defects (Box 1). This provides an exciting field for future cancer research and clinical applications, despite the acquired drug resistance (Box 3), which still poses a major drawback to the use of PARP inhibitors in the clinic. Indeed, as we discuss next, novel synthetic-lethal interactions between DDR factors and pathways have been identified, some of which are already being tested in early clinical trials.

Targeting DDR kinases in the clinic

ATM inhibitors

ATM is the apical DDR kinase orchestrating DSB repair, and therefore multiple compounds have been developed for its selective inhibition (Table 1). Given the key role of ATM in DSB signalling and repair, ATM inhibition coupled with radiotherapy is an obvious and attractive therapeutic combination for tumour eradication, and is currently being explored in several clinical trials. Early on, it was revealed that specific ATM inhibitors can be used to increase the anticancer activity of DNA-damaging agents such as topoisomerase inhibitors[99] or PARP inhibitors[100]. In addition, ATM deficiency sensitizes cancer cells to topoisomerase 1 inhibitors[71] or PARP inhibitors[71,101]. This is of clinical relevance since *ATM* is frequently mutated or inactivated in sporadic cancers, including lung, breast, brain[102] or pancreatic[103] cancers. Mechanistically, PARP and topoisomerase 1 inhibitors lead to one-ended DSBs which require HR for their accurate repair. Loss of ATM signalling delays end resection and channels the repair of one-ended DSBs to NHEJ, which results in toxic chromosome fusions. Conversely, abrogation of XRCC4, LIG4 or XLF, which are critical for NHEJ, promotes resistance to PARP or topoisomerase 1 inhibitors in ATM-deficient cells, because one-ended DSBs can be repaired by HR[71]. However, ATM inhibition reverses PARP inhibitor resistance in BRCA1–53BP1-deficient or BRCA1–REV7-deficient mouse cells[104,105], an effect mediated by reduced end resection. Thus, the combination of ATM inhibitors with PARP or topoisomerase 1 inhibitors may be a promising therapeutic strategy, which is currently being investigated in the clinic (Table 1).

In addition, previous studies have indicated that ATM inhibition is synthetic lethal with loss of genes in the Fanconi anaemia pathway[106,107], an observation recently validated in multiple CRISPR–Cas9 screens[108,109]. Thus, tumours that are defective in one or more of the more than 20 Fanconi anaemia genes (for example, ~5% of ovarian cancers[110]) may be attractive candidates for therapy using ATM inhibitors.

Despite several ongoing clinical trials (Table 1), the clinical efficacy of ATM inhibitors is yet to be demonstrated. It is also noteworthy that germ line heterozygous *ATM* mutation predisposes to haematological malignancies and other cancer types[111,112]. Therefore, careful consideration should be given to ensure that therapeutic benefits of ATM inhibition outweigh the risks for patients with cancer, as prolonged exposure to ATM inhibitors might lead to de novo tumour formation in multiple tissues.

DNA-PKcs inhibitors

DNA-PKcs, the catalytic subunit of the holoenzyme DNA-PK, which also includes the proteins Ku70 and Ku80, is critical for DSB repair via NHEJ throughout the cell cycle. In line with this, cells lacking DNA-PKcs are sensitive to DSB-inducing agents[113–115], and treatment with DNA-PKcs inhibitors recapitulates this effect[116]. Similarly to ATM, DNA-PKcs inhibition in combination with radiotherapy is an attractive anticancer strategy[116,117]. Importantly, because ATM and DNA-PKcs display only partial redundancy in substrate phosphorylation in response to DSBs, their concomitant loss is synthetic lethal[118,119]. Therefore, DNA-PKcs inhibitors, especially those that are currently in clinical trials (Table 1), might also be effective against ATM-deficient tumours[120]. This concept would presumably not apply to inhibitors such as XRD-0394, a dual inhibitor

Box 3

Clinically relevant mechanisms of PARP inhibitor resistance

The major drawback to the use of poly(ADP-ribose) polymerase (PARP) inhibitors in the clinic is the emergence of resistant disease. Recent in vitro studies have identified several gene mutations that could underlie PARP inhibitor resistance. However, only a subset of these mutations was found in patients. Reactivation of the *BRCA1* gene or the *BRCA2* gene through reversion mutations that restore the open reading frames is one of the most common causes of PARP inhibitor resistance in the clinic[231]. Often, this is a partial open reading frame restoration, sufficient to sustain some of the key BRCA1/2 activities (for example, RAD51 loading at double-strand breaks). A frequent mechanism of BRCA1 inactivation in sporadic triple-negative breast tumours is *BRCA1* promoter hypermethylation, a predictor of PARP inhibitor sensitivity[232]. *BRCA1* silencing is reversed by demethylation, which enables residual transcription, sufficient to trigger resistance[229]. Resistant tumours that re-express *BRCA1* while retaining *BRCA1* promoter hypermethylation have been reported[233]. One such case was caused by a chromosomal rearrangement which placed the *BRCA1* gene under the control of a heterologous active promoter.

Another mechanism of PARP inhibitor resistance specific to *BRCA1*-mutated tumours is restoration of homologous recombination repair via abrogation of 53BP1 (ref.[234]) or of its interacting partners RIF1 (refs.[235–238]) and the shieldin complex (that is, REV7, SHLD1, SHLD2 and SHLD3)[239–242]. Abrogation of any of these factors in BRCA1-deficient cells reactivates end resection and enables loading of RAD51 at double-strand break sites to levels sufficient for partial homologous recombination restoration. This mechanism is clinically relevant because mutations in or downregulation of the genes encoding 53BP1 pathway components have been found in PARP inhibitor-resistant triple-negative human breast cancers[230,239].

Mechanisms of PARP inhibitor resistance independent of homologous recombination reactivation have also been reported in patients with cancer. For example, loss of expression of poly(ADP-ribose) glycohydrolase (PARG), the enzyme responsible for the degradation of poly(ADP-ribose) chains, stabilizes poly(ADP-ribosyl)ation and promotes PARP inhibitor resistance in BRCA1/BRCA2-deficient cells and tumours[243]. PARG suppression, as detected by immunohistochemistry, may be relevant to a wide spectrum of tumours. Similarly, loss of the helicase SLFN11, a factor with elusive roles in DNA replication and repair, promotes resistance to PARP inhibitors in BRCA1/2-deficient cells[143]. An *SLFN11* missense mutation was reported in a patient with PARP inhibitor-resistant *BRCA1*-deleted metastatic breast cancer[230].

Glossary

cGAS–STING axis
A pathway that detects cytosolic DNA and activates innate immune responses.

Chromothripsis
The process of chromosome shattering followed by incorrect repair.

Displacement loop
(D-loop). A DNA structure formed when a double helix is separated by invasion of a complementary single-stranded DNA end during homologous recombination.

Double Holliday junction
A four-way joint DNA molecule and homologous recombination intermediate that can form after second-end capture and ligation.

Micronuclei
Cytosolic structures containing a chromosome or chromosome fragment that are not incorporated in the nucleus during mitosis.

One-ended DSBs
Double-stranded DNA ends formed when a replication fork collapses, where there is no second DNA end available for ligation.

Second-end capture
The annealing step that pairs the second resected single-stranded DNA end from one side of a double-strand break with the joint molecule formed by invasion of a template DNA by the first resected DNA end during homologous recombination.

Shieldin complex
A protein complex consisting of SHLD1–SHLD2–SHLD3, which acts together with 53BP1–RIF1–REV7 to limit end resection at DNA double-strand breaks.

Two-ended DSBs
DNA double-strand breaks (DSBs) where both ends are available for ligation.

of both ATM and DNA-PKcs that has recently entered phase I clinical trials but for which no data have yet been made publicly available.

ATR inhibitors

The kinase ATR acts during S phase to ensure timely and accurate DNA replication by regulating origin firing and fork progression. Because cancer cells frequently accumulate stalled replication forks, they become reliant on ATR signalling to sustain DNA replication. Cells with defective control of the G1–S transition due to p53 deficiency or oncogene activation are also dependent on ATR signalling[121–123]. Thus, ATR is an attractive target for cancer therapy, with multiple specific and potent ATR inhibitors having been developed in recent years[123–125] (Table 1). ATR inhibitors increase replication fork stalling and promote chromosomal breakage, leading to cytotoxicity. Cancer cells with replication stress caused by overexpression of oncogenes such as the oncogene encoding cyclin E1 (ref.[123]) are particularly susceptible to ATR inhibitors. Similar results were observed in xenograft mouse models of *ALK*- and *MYCN*-amplified neuroblastoma, where ATR inhibition led to inhibition of tumour growth[126]. Cancer cells harbouring *ATM* mutations that interfere with DNA repair[127,128] are also susceptible to ATR inhibitors[129].

ATR inhibitors sensitize tumours to ionizing radiation, and their combination is effective against patient-derived xenograft (PDX) models of triple-negative breast cancer[130]. Interestingly, recent studies have demonstrated that ATR inhibitors potentiate radiation-induced interferon signalling[131,132], suggesting potential therapeutic combinations of ATR inhibitors with immunotherapy. This concept was validated in mouse models for castration-resistant prostate cancer, where ATR inhibitors combined with immunotherapy targeting the immune checkpoint molecule PDL1 showed synergistic antitumour activity[133]. Combinations of ATR inhibitors and immune checkpoint blockade are currently being tested in clinical trials (Table 1); for example, a combination of the ATR inhibitor VE-822 and the anti-PDL1 immunotherapy avelumab is being used to treat solid tumours with mutations in DDR genes.

Several studies have revealed synergies between ATR inhibitors and established anticancer drugs, which interfere with DNA replication or inflict DNA damage (for example, cisplatin, gemcitabine and temozolomide)[124,128,134–137]. Importantly, ATR inhibitors have also been reported to act synergistically with PARP inhibitors against tumours harbouring *BRCA1*/*BRCA2* mutations[138], Mechanistically, ATR inhibitors exacerbate some of the pathologies induced by PARP inhibitors in BRCA2-deficient cells, including accelerated mitotic entry, accumulation of chromatin bridges and broken and/or lagging chromosomes, leading to increased PARP inhibitor cytotoxicity[139]. The synergy between these two drugs has recently been extended to wild type *BRCA1*/*BRCA2* tumours[140,141].

One of the most promising clinical applications of ATR inhibitors is their potential to treat PARP inhibitor-resistant tumours. BRCA1-deficient cancer cells can overcome PARP inhibitor toxicity via BRCA1-independent loading of RAD51 at DSBs (Box 3). In vitro, ATR inhibitors block this BRCA1-independent function, leading to resensitization of tumour cells to PARP inhibition[142]. Furthermore, ATR inhibitors reversed PARP inhibitor resistance caused by inactivation of the helicase SLFN11 in both HR-proficient and HR-deficient cancer cells[143].

CHK1 inhibitors

As a downstream target of ATR, the kinase CHK1 is similarly activated by defects in DNA replication. CHK1 inhibitors selectively kill cancer cells with high levels of replication stress[140,144,145]. CHK1 inhibitors were shown to potentiate the cytotoxicity of DNA-damaging agents such as gemcitabine, cisplatin and camptothecin in p53-deficient cancer cells[146] and of irinotecan in p53-deficient mouse models of triple-negative breast cancer[147].

Similarly to ATR inhibitors, CHK1 inhibitors exacerbate the DNA damage caused by PARP inhibition[148]. Furthermore, treatment with the CHK1 inhibitor prexasertib activates type I interferon signalling, and its combination with anti-PDL1 immunotherapy results in synergistic antitumour responses in mouse models of small cell lung cancer[149]. Although some CHK1 inhibitors showed substantial toxicity in initial clinical trials, several novel candidate compounds with better safety profiles are currently being tested in the clinic (Table 1).

WEE1 and PKMYT1 inhibitors

The tyrosine kinase WEE1 is responsible for the inhibitory phosphorylation of CDK1, which prevents mitotic entry[17,18]. Therefore, treatment with WEE1 inhibitors results in aberrant entry into mitosis, which underlies the cytotoxicity of these compounds. CDK1 activation in S phase also leads to enhanced cleavage of stalled replication forks by the SLX4–MUS81 complex and toxic accumulation of DNA damage[150].

Importantly, inhibition of WEE1 is synthetic lethal with loss of histone H3 Lys36 trimethylation (H3K36me3), which is frequently found in cancers lacking the histone methyltransferase SETD2 or overexpressing the histone lysine demethylase KDM4A[151]. Given that *SETD2* is frequently mutated in cancer, the results obtained in xenograft models and cell lines[151] provided a rationale for the use of anti-H3K36me3 antibodies as biomarkers for the identification of *SETD2*-mutated tumours that are susceptible to WEE1 inhibitors, which is currently being tested clinically[152].

Recently, the WEE1-like kinase PKMYT1 was identified in a CRISPR–Cas9 screen for synthetic-lethal interactions with cyclin E overexpression[153]. Treatment with the PKMYT1 inhibitor RP-6306 triggers unscheduled CDK1 activation in cancer cells overexpressing cyclin E, and this inhibitor exhibits potent in vivo antitumour activity in xenograft or PDX models, alone or in combination with gemcitabine[153]. This inhibitor is currently being tested in phase I clinical trials in patients with solid tumours (Table 1).

Emerging therapies and new targets

In addition to PARP1 and the kinases discussed earlier herein, which have been validated as bona fide targets for cancer therapy (Tables 1, 2), several additional DDR factors have been identified as potential targets for specific tumours. These, together with inhibitory compounds currently in clinical development, are discussed in the following subsections.

POLQ inhibitors target HR-deficient tumours

POLQ has a central role in the alternative end joining pathway, which repairs resected DNA ends with internal microhomologies[154] (Fig. 2). POLQ contains a helicase-like ATPase domain required to displace RPA from resected ssDNA ends and to unravel homologous sequences[155]. Upon their alignment and annealing, the polymerase activity of POLQ fills in the gaps, which are subsequently sealed by LIG1 or LIG3 (refs.[156–158]). Reflecting its key role in DNA repair, POLQ inhibition has a radiosensitizing effect in a panel of different tumour cell lines in vitro[159].

The discovery of *POLQ* synthetic-lethal interactions with DNA repair genes, including *BRCA1*, *BRCA2* and *ATM*[84,160,161], has underscored POLQ as a promising target for elimination of tumours with compromised HR repair. Concerted efforts of several groups worldwide towards identification of POLQ inhibitors have led to the identification of two compounds, which showed efficacy in preclinical studies and for which clinical trials are currently being designed. The antibiotic novobiocin was identified as a specific and potent POLQ inhibitor in a small-molecule library screen[162]. Binding of novobiocin to the ATPase domain of POLQ triggers elimination of BRCA1/2-deficient cells and potentiates the toxic effects of PARP inhibitors. Importantly, novobiocin inhibits growth of *BRCA1*-deleted tumours in genetically engineered mouse models of triple-negative breast cancer, as well as in PDXs with acquired PARP inhibitor resistance through loss of 53BP1, indicating the unique clinical potential of this drug. Moreover, high *POLQ* mRNA levels could potentially be used as a predictive biomarker of novobiocin sensitivity in PARP inhibitor-resistant cancers.

A second POLQ inhibitor, ART558, shows nanomolar POLQ affinity in vitro and specifically inhibits growth of BRCA2-deficient cells, thus recapitulating the effect of *POLQ* genetic ablation[163]. Similarly to novobiocin, ART558 is toxic to *BRCA1*$^{-/-}$ cells, and organoids derived from a BRCA1-mutant breast cancer that have acquired resistance to PARP inhibitors. Although in vivo data for this inhibitor are not available, an additional POLQ inhibitor with increased bioavailability and lower clearance in animal models developed by the same company, ART812, suppresses the growth of PARP inhibitor-resistant *BRCA1*$^{-/-}$ tumours in rats.

USP1 inhibitors target BRCA1/2-deficient cancers

The deubiquitylating enzyme USP1 is a negative regulator of translesion synthesis via reversal of proliferating cell nuclear antigen (PCNA) ubiquitylation[164–167]. PCNA ubiquitylation is required for postreplicative DNA repair via translesion synthesis[168,169]. The persistence of monoubiquitylated PCNA upon USP1 inactivation destabilizes replication forks, which become reliant on BRCA1. Consequently, USP1 inactivation is synthetic lethal with BRCA1 loss in cancer cells. The recently developed USP1 inhibitor KSQ-4279 specifically targets BRCA1/2-deficient cancer cells. Importantly, its combination with PARP inhibitors led to a prolonged antitumour response in PDX models of triple-negative breast cancer, regardless of BRCA1/2 status[170]. KSQ-4279 is currently being tested in a phase I clinical trial in patients with advanced solid tumours[171].

RAD51 inhibitors

RAD51 is the central enzyme of the HR pathway of DSB repair, and as such is essential for cell survival. RAD51 assembles nucleoprotein filaments on ssDNA formed at resected DSBs, which catalyse strand invasion into a homologous DNA template to initiate HR repair. BRCA2 is the loader of RAD51 at DSB sites and, as expected, RAD51 abrogation is synthetic lethal with PARP inhibition[172–174]. Several compounds that modulate RAD51 activity and prevent RAD51 filament formation and/or inhibit HR repair in vitro and in vivo have been identified (reviewed in[175]). The RAD51 inhibitor CYT-0851 exhibits selective cytotoxic activity against activation-induced cytidine deaminase (AID)-expressing chronic lymphocytic leukaemia mouse models and human cancer cells[176,177] and is currently in a phase I/II clinical trial, alone or in combination with chemotherapies, for treatment of B cell haematological malignancies and advanced solid tumours[178].

Targeting the helicase WRN in cancers with microsatellite instability

Microsatellite instability (MSI) is a common hallmark of cancers with defective MMR (for example, *MLH1*, *MSH2*, *MSH3* or *MSH6* gene mutations) and is frequent in stomach (~28%), uterine (~22%) and colorectal (~15%) cancers[179–181]. The helicase WRN is a synthetic-lethal target in cells with MSI, particularly in hereditary non-polyposis colon cancer[182,183]. However, inactivation of MMR genes alone was not sufficient to explain the observed synthetic-lethal interaction. Instead, long-term MMR deficiency leads to the accumulation of large TA dinucleotide repeats[184], which form non-B-DNA secondary structures. In the absence of WRN, the SLX4–MUS81 endonuclease complex cleaves these structures, resulting in chromosomal fragmentation and apoptosis[184]. Therefore, inhibition of WRN helicase activity might be a potent and selective therapy for MSI-positive cancers.

Conclusion and perspectives

The DDR network is highly coordinated to ensure chromosome integrity during cellular processes involving DNA metabolic reactions (replication and repair). Given that these activities are essential for cell survival and proliferation, it became clear early on that targeting the DDR could be a feasible strategy for suppressing growth of cancer cells[1,185]. The obvious challenge to this approach is to target the DDR specifically in the tumour, leaving the normal tissues unaffected. Substantial progress has been made towards tackling this challenge with the identification of specific genetic alterations that promote tumour growth, while making the tumour vulnerable to DDR-targeting treatments with selective toxicity in this specific genetic background. The well-characterized paradigm of the synthetic-lethal interaction between *BRCA1*/*BRCA2* mutations in tumours and PARP inhibitors illustrates this concept. Applying the same principle, multiple studies have identified synthetic-lethal interactions between other DDR genes mutated in cancer, thereby supplying potential druggable targets

against tumours carrying those mutations. However, the approach has proven cumbersome. This is because three unique factors underlie the remarkable hypersensitivity of *BRCA1*/*BRCA2*-mutated cells and tumours to PARP inhibitors: (1) PARP inhibitors prevent processing of unligated Okazaki fragments, thereby generating SSBs[94]; (2) SSBs are the most common DNA lesions in cancers[186] and require PARP1 for repair; (3) unrepaired SSBs in BRCA1/2-mutated tumours treated with PARP inhibitors cause further DNA damage, which in turn triggers potent innate immune responses[187,188]. It is therefore possible that other synthetic-lethal interactions may not be able to replicate the success of PARP inhibitors against HR-deficient tumours.

Moreover, most tools used for the identification of novel druggable targets (for example, CRISPR–Cas9 screens) monitor synthetic-lethal interactions on the genetic level. Although this approach can lead to the identification of novel druggable targets (for example, RNase H in BRCA1-deficient cancer cells[88]), the deletion of a target protein may not have as strong anticancer effects as its chemical inhibition. For example, the cytotoxicity of PARP1 inhibitors with regard to HR-deficient cancers is thought to be mediated by PARP1 trapping rather than by loss of PARP activity[87,189]. In other cases, loss of an enzyme's catalytic activity is worse than not having the protein at all. For example, *Atm*-knockout mice are viable although tumour-prone, whereas mutation of the ATM catalytic site is embryonic lethal[190,191]. The same is true for DNA-PKcs kinase-dead mice, which are also embryonic lethal, although mice lacking DNA-PKcs are viable[192].

Furthermore, inactivation of DDR genes elicits genomic instability, leading to alterations in multiple other genes, making it difficult to ascertain the genetic origin of the tumour. The emergence of precision medicine emphasized the benefits of matching a specific genetic alteration found in patients with cancer with a targeted therapy for this alteration. The aim is to maximize therapeutic impact, while diminishing adverse effects (that is, toxicity and resistance). However, the advances made in precision medicine in recent years have also unravelled the complexity of the task. For example, although technologies that enable sequencing of patient samples are readily available, deconvoluting and interpreting the genomics and transcriptomics data remain major obstacles[193].

Another consequence of the genomic instability in DDR-mutated tumours is the rewiring of their genetic make-up to escape drug toxicity and restore tumour growth despite the targeted treatment. This phenomenon is known as acquired resistance to therapy, and is still an insurmountable obstacle in the clinic. The design of drug combinations, with optimized timing and dosing regimens, that reduce toxicity and limit or suppress resistance could tackle these problems[194]. In this respect, the use of DDR inhibitors together with immune checkpoint blockade is a promising therapeutic strategy. The DNA lesions inflicted by DDR inhibitors lead to the accumulation of cytosolic DNA fragments (for example, in the form of micronuclei), which activate the cGAS–STING axis and trigger antitumour immune responses. The interplay between the DDR and cGAS–STING pathways and clinical applications have been discussed in other recent reviews[187,195].

Multiple novel compounds targeting DDR pathways have been recently developed, as discussed in this Review, and those showing promising results in the preclinical setting are being evaluated in patients. However, it is important to recognize that such empirical approaches may not work in complex diseases such as cancer. It is imperative therefore to concentrate research efforts towards the discovery of novel targets for drug development, as well as towards the optimization of existing drugs through superior formulations or identification of potent drug combinations that can reduce side effects and maximize clinical efficacy.

Published online: 5 December 2022

References

1. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071–1078 (2009).
2. Taylor, A. M. R. et al. Chromosome instability syndromes. *Nat. Rev. Dis. Prim.* **5**, 64 (2019).
3. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
4. Lord, C. J. & Ashworth, A. PARP inhibitors: synthetic lethality in the clinic. *Science* **355**, 1152–1158 (2017).
5. Bryant, H. E. et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
6. Farmer, H. et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
   Along with Bryant et al. (2005), this article shows that PARP inhibition targets BRCA1/2-deficient cells.
7. Fugger, K., Hewitt, G., West, S. C. & Boulton, S. J. Tackling PARP inhibitor resistance. *Trends Cancer* **7**, 1102–1118 (2021).
8. 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).
9. Matsuoka, S. et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160–1166 (2007).
10. Kastan, M. B. & Bartek, J. Cell-cycle checkpoints and cancer. *Nature* **432**, 316–323 (2004).
11. Blackford, A. N. & Stucki, M. How cells respond to DNA breaks in mitosis. *Trends Biochem. Sci.* **45**, 321–331 (2020).
12. Leimbacher, P. A. et al. MDC1 interacts with TOPBP1 to maintain chromosomal stability during mitosis. *Mol. Cell* **74**, 571–583 e578 (2019).
13. De Marco Zompit, M. et al. The CIP2A-TOPBP1 complex safeguards chromosomal stability during mitosis. *Nat. Commun.* **13**, 4143 (2022).
14. Mailand, N. et al. Rapid destruction of human Cdc25A in response to DNA damage. *Science* **288**, 1425–1429 (2000).
15. Peng, C. Y. et al. Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216. *Science* **277**, 1501–1505 (1997).
16. Sanchez, Y. et al. Conservation of the Chk1 checkpoint pathway in mammals: linkage of DNA damage to Cdk regulation through Cdc25. *Science* **277**, 1497–1501 (1997).
17. Heald, R., McLoughlin, M. & McKeon, F. Human wee1 maintains mitotic timing by protecting the nucleus from cytoplasmically activated Cdc2 kinase. *Cell* **74**, 463–474 (1993).
18. Gu, Y., Rosenblatt, J. & Morgan, D. O. Cell cycle regulation of CDK2 activity by phosphorylation of Thr160 and Tyr15. *EMBO J.* **11**, 3995–4005 (1992).
19. Booher, R. N., Holman, P. S. & Fattaey, A. Human Myt1 is a cell cycle-regulated kinase that inhibits Cdc2 but not Cdk2 activity. *J. Biol. Chem.* **272**, 22300–22306 (1997).
20. Liu, F., Stanton, J. J., Wu, Z. & Piwnica-Worms, H. The human Myt1 kinase preferentially phosphorylates Cdc2 on threonine 14 and localizes to the endoplasmic reticulum and Golgi complex. *Mol. Cell Biol.* **17**, 571–583 (1997).
21. Simoneau, A. & Zou, L. An extending ATR-CHK1 circuitry: the replication stress response and beyond. *Curr. Opin. Genet. Dev.* **71**, 92–98 (2021).
22. Lecona, E. & Fernandez-Capetillo, O. Targeting ATR in cancer. *Nat. Rev. Cancer* **18**, 586–595 (2018).
23. Dupre, A., Boyer-Chatenet, L. & Gautier, J. Two-step activation of ATM by DNA and the Mre11-Rad50-Nbs1 complex. *Nat. Struct. Mol. Biol.* **13**, 451–457 (2006).
24. Lee, J. H. & Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science* **308**, 551–554 (2005).
25. Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542–1548 (2003).
26. Bass, T. E. et al. ETAA1 acts at stalled replication forks to maintain genome integrity. *Nat. Cell Biol.* **18**, 1185–1195 (2016).
27. 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).
28. Haahr, P. et al. Activation of the ATR kinase by the RPA-binding protein ETAA1. *Nat. Cell Biol.* **18**, 1196–1207 (2016).
29. Kumagai, A., Lee, J., Yoo, H. Y. & Dunphy, W. G. TopBP1 activates the ATR-ATRIP complex. *Cell* **124**, 943–955 (2006).
30. Caldecott, K. W. DNA single-strand break repair. *Exp. Cell Res.* **329**, 2–8 (2014).
31. Jiricny, J. Postreplicative mismatch repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012633 (2013).
32. Klein, H. L. Genome instabilities arising from ribonucleotides in DNA. *DNA Repair* **56**, 26–32 (2017).
33. Semlow, D. R. & Walter, J. C. Mechanisms of vertebrate DNA interstrand cross-link repair. *Annu. Rev. Biochem.* **90**, 107–135 (2021).
34. Ruggiano, A. & Ramadan, K. DNA-protein crosslink proteases in genome stability. *Commun. Biol.* **4**, 11 (2021).
35. Davidovic, L., Vodenicharov, M., Affar, E. B. & Poirier, G. G. Importance of poly(ADP-ribose) glycohydrolase in the control of poly(ADP-ribose) metabolism. *Exp. Cell Res.* **268**, 7–13 (2001).

36. Tarsounas, M. & Sung, P. The antitumorigenic roles of BRCA1-BARD1 in DNA repair and replication. *Nat. Rev. Mol. Cell Biol.* **21**, 284–299 (2020).
37. Spies, M. & Fishel, R. Mismatch repair during homologous and homologous recombination. *Cold Spring Harb. Perspect. Biol.* **7**, a022657 (2015).
38. Nambiar, T. S., Baudrier, L., Billon, P. & Ciccia, A. CRISPR-based genome editing through the lens of DNA repair. *Mol. Cell* **82**, 348–388 (2022).
39. Walker, J. R., Corpina, R. A. & Goldberg, J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**, 607–614 (2001).
40. Graham, T. G. W., Walter, J. C. & Loparo, J. J. Two-stage synapsis of DNA ends during non-homologous end joining. *Mol. Cell* **61**, 850–858 (2016).
41. Dobbs, T. A., Tainer, J. A. & Lees-Miller, S. P. A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA Repair* **9**, 1307–1314 (2010).
42. Hammel, M. et al. Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. *J. Biol. Chem.* **285**, 1414–1423 (2010).
43. Critchlow, S. E., Bowater, R. P. & Jackson, S. P. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr. Biol.* **7**, 588–598 (1997).
44. Grawunder, U. et al. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* **388**, 492–495 (1997).
45. Ahnesorg, P., Smith, P. & Jackson, S. P. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* **124**, 301–313 (2006).
46. Balmus, G. et al. Synthetic lethality between PAXX and XLF in mammalian development. *Genes Dev.* **30**, 2152–2157 (2016).
47. Lescale, C. et al. Specific roles of XRCC4 paralogs PAXX and XLF during V(D)J recombination. *Cell Rep.* **16**, 2967–2979 (2016).
48. Buck, D. et al. Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* **124**, 287–299 (2006).
49. Ochi, T. et al. DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science* **347**, 185–188 (2015).
50. Zhao, B., Rothenberg, E., Ramsden, D. A. & Lieber, M. R. The molecular basis and disease relevance of non-homologous DNA end joining. *Nat. Rev. Mol. Cell Biol.* **21**, 765–781 (2020).
51. Cejka, P. & Symington, L. S. DNA end resection: mechanism and control. *Annu. Rev. Genet.* **55**, 285–307 (2021).
52. Huertas, P. & Jackson, S. P. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J. Biol. Chem.* **284**, 9558–9565 (2009).
53. Sartori, A. A. et al. Human CtIP promotes DNA end resection. *Nature* **450**, 509–514 (2007).
54. 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).
55. Yilmaz, D. et al. Activation of homologous recombination in G1 preserves centromeric integrity. *Nature* **600**, 748–753 (2021).
56. 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).
57. 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).
58. Mimitou, E. P. & Symington, L. S. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* **455**, 770–774 (2008).
59. West, S. C. et al. Resolution of recombination intermediates: mechanisms and regulation. *Cold Spring Harb. Symp. Quant. Biol.* **80**, 103–109 (2015).
60. Llorens-Agost, M. et al. POLθ-mediated end joining is restricted by RAD52 and BRCA2 until the onset of mitosis. *Nat. Cell Biol.* **23**, 1095–1104 (2021).
61. Beucher, A. et al. ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *EMBO J.* **28**, 3413–3427 (2009).
62. Karanam, K., Kafri, R., Loewer, A. & Lahav, G. Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. *Mol. Cell* **47**, 320–329 (2012).
63. Pellegrino, S., Michelena, J., Teloni, F., Imhof, R. & Altmeyer, M. Replication-coupled dilution of H4K20me2 guides 53BP1 to pre-replicative chromatin. *Cell Rep.* **19**, 1819–1831 (2017).
64. Simonetta, M. et al. H4K20me2 distinguishes pre-replicative from post-replicative chromatin to appropriately direct DNA repair pathway choice by 53BP1-RIF1-MAD2L2. *Cell Cycle* **17**, 124–136 (2018).
65. Fradet-Turcotte, A. et al. 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature* **499**, 50–54 (2013).
66. Botuyan, M. V. et al. Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).
67. Feng, S. et al. RIF1-ASF1-mediated high-order chromatin structure safeguards genome integrity. *Nat. Commun.* **13**, 957 (2022).
68. Mirman, Z., Sasi, N. K., King, A., Chapman, J. R. & de Lange, T. 53BP1-shieldin-dependent DSB processing in BRCA1-deficient cells requires CST-Polα-primase fill-in synthesis. *Nat. Cell Biol.* **24**, 51–61 (2022).
69. Ochs, F. et al. Stabilization of chromatin topology safeguards genome integrity. *Nature* **574**, 571–574 (2019).
70. Paiano, J. et al. Role of 53BP1 in end protection and DNA synthesis at DNA breaks. *Genes Dev.* **35**, 1356–1367 (2021).
71. Balmus, G. et al. ATM orchestrates the DNA-damage response to counter toxic non-homologous end-joining at broken replication forks. *Nat. Commun.* **10**, 87 (2019).
72. Becker, J. R. et al. BARD1 reads H2A lysine 15 ubiquitination to direct homologous recombination. *Nature* **596**, 433–437 (2021).
73. Hu, Q. et al. Mechanisms of BRCA1-BARD1 nucleosome recognition and ubiquitylation. *Nature* **596**, 438–443 (2021).
74. Krais, J. J. et al. RNF168-mediated localization of BARD1 recruits the BRCA1-PALB2 complex to DNA damage. *Nat. Commun.* **12**, 5016 (2021).
75. Nakamura, K. et al. H4K20me0 recognition by BRCA1-BARD1 directs homologous recombination to sister chromatids. *Nat. Cell Biol.* **21**, 311–318 (2019).
This study shows that BARD1 interacts with H4K20me0, which is found only on nascent chromatin. Therefore, HR is restricted to S phase, where all the H4K20 is unmethylated.
76. Ochs, F. et al. 53BP1 fosters fidelity of homology-directed DNA repair. *Nat. Struct. Mol. Biol.* **23**, 714–721 (2016).
77. Ramsden, D. A., Carvajal-Garcia, J. & Gupta, G. P. Mechanism, cellular functions and cancer roles of polymerase-theta-mediated DNA end joining. *Nat. Rev. Mol. Cell Biol.* **23**, 125–140 (2021).
78. Thompson, D. & Easton, D. F., Breast Cancer Linkage Consortium. Cancer incidence in BRCA1 mutation carriers. *J. Natl Cancer Inst.* **94**, 1358–1365 (2002).
79. Breast Cancer Linkage Consortium. Cancer risks in BRCA2 mutation carriers. *J. Natl Cancer Inst.* **91**, 1310–1316 (1999).
80. 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).
81. Schlacher, K. et al. Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145**, 529–542 (2011).
82. Quinet, A., Lemacon, D. & Vindigni, A. Replication fork reversal: players and guardians. *Mol. Cell* **68**, 830–833 (2017).
83. Roy, R., Chun, J. & Powell, S. N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat. Rev. Cancer* **12**, 68–78 (2012).
84. Ceccaldi, R. et al. Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair. *Nature* **518**, 258–262 (2015).
85. Zhang, Y. & Jasin, M. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nat. Struct. Mol. Biol.* **18**, 80–84 (2011).
86. Yousefzadeh, M. J. et al. Mechanism of suppression of chromosomal instability by DNA polymerase POLQ. *PLoS Genet.* **10**, e1004654 (2014).
87. Murai, J. et al. Trapping of PARP1 and PARP2 by clinical PARP inhibitors. *Cancer Res.* **72**, 5588–5599 (2012).
This study proposes that the cytotoxicity of PARP inhibitors depends on their trapping activity rather than their catalytic inhibitory activity.
88. Zimmermann, M. et al. CRISPR screens identify genomic ribonucleotides as a source of PARP-trapping lesions. *Nature* **559**, 285–289 (2018).
89. Blessing, C. et al. The oncogenic helicase ALC1 regulates PARP inhibitor potency by trapping PARP2 at DNA breaks. *Mol. Cell* **80**, 862–875 e866 (2020).
90. Hewitt, G. et al. Defective ALC1 nucleosome remodeling confers PARPi sensitization and synthetic lethality with HRD. *Mol. Cell* **81**, 767–783 e711 (2021).
91. Juhasz, S. et al. The chromatin remodeler ALC1 underlies resistance to PARP inhibitor treatment. *Sci. Adv.* https://doi.org/10.1126/sciadv.abb8626 (2020).
92. Verma, P. et al. ALC1 links chromatin accessibility to PARP inhibitor response in homologous recombination-deficient cells. *Nat. Cell Biol.* **23**, 160–171 (2021).
93. Krastev, D. B. et al. The ubiquitin-dependent ATPase p97 removes cytotoxic trapped PARP1 from chromatin. *Nat. Cell Biol.* **24**, 62–73 (2022).
94. Vaitsiankova, A. et al. PARP inhibition impedes the maturation of nascent DNA strands during DNA replication. *Nat. Struct. Mol. Biol.* **29**, 329–338 (2022).
95. Hanzlikova, H. et al. The importance of poly(ADP-ribose) polymerase as a sensor of unligated Okazaki fragments during DNA replication. *Mol. Cell* **71**, e313 (2018).
96. Cong, K. et al. Replication gaps are a key determinant of PARP inhibitor synthetic lethality with BRCA deficiency. *Mol. Cell* **81**, 3128–3144 e3127 (2021).
This study proposes that the accumulation of ssDNA gaps underlies the cytotoxic activity of PARP inhibitors.
97. Schoonen, P. M. et al. Progression through mitosis promotes PARP inhibitor-induced cytotoxicity in homologous recombination-deficient cancer cells. *Nat. Commun.* **8**, 15981 (2017).
98. Simoneau, A., Xiong, R. & Zou, L. The trans cell cycle effects of PARP inhibitors underlie their selectivity toward BRCA1/2-deficient cells. *Genes. Dev.* **35**, 1271–1289 (2021).
99. Hickson, I. et al. Identification and characterization of a novel and specific inhibitor of the ataxia-telangiectasia mutated kinase ATM. *Cancer Res.* **64**, 9152–9159 (2004).
This study describes the use of a selective ATM inhibitor to potentiate the anticancer activity of DNA-damaging agents.
100. Bryant, H. E. & Helleday, T. Inhibition of poly (ADP-ribose) polymerase activates ATM which is required for subsequent homologous recombination repair. *Nucleic Acids Res.* **34**, 1685–1691 (2006).
101. Weston, V. J. et al. The PARP inhibitor olaparib induces significant killing of ATM-deficient lymphoid tumor cells in vitro and in vivo. *Blood* **116**, 4578–4587 (2010).
102. Cremona, C. A. & Behrens, A. ATM signalling and cancer. *Oncogene* **33**, 3351–3360 (2014).
103. Bailey, P. et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).
104. Bunting, S. F. et al. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* **141**, 243–254 (2010).
This study reveals the competition between 53BP1 and BRCA1 in DSB repair pathway choice.

105. Xu, G. et al. REV7 counteracts DNA double-strand break resection and affects PARP inhibition. *Nature* **521**, 541–544 (2015).
106. Blackford, A. N. et al. The DNA translocase activity of FANCM protects stalled replication forks. *Hum. Mol. Genet.* **21**, 2005–2016 (2012).
107. Kennedy, R. D. et al. Fanconi anemia pathway-deficient tumor cells are hypersensitive to inhibition of ataxia telangiectasia mutated. *J. Clin. Invest.* **117**, 1440–1449 (2007).
108. Cai, M. Y. et al. Cooperation of the ATM and Fanconi anemia/BRCA pathways in double-strand break end resection. *Cell Rep.* **30**, 2402–2415 e2405 (2020).
109. Wang, C. et al. Genetic vulnerabilities upon inhibition of DNA damage response. *Nucleic Acids Res.* **49**, 8214–8231 (2021).
110. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
111. Reiman, A. et al. Lymphoid tumours and breast cancer in ataxia telangiectasia; substantial protective effect of residual ATM kinase activity against childhood tumours. *Br. J. Cancer* **105**, 586–591 (2011).
112. Choi, M., Kipps, T. & Kurzrock, R. ATM mutations in cancer: therapeutic implications. *Mol. Cancer Ther.* **15**, 1781–1791 (2016).
113. Blunt, T. et al. Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell* **80**, 813–823 (1995).
114. Kirchgessner, C. U. et al. DNA-dependent kinase (p350) as a candidate gene for the murine SCID defect. *Science* **267**, 1178–1183 (1995).
115. Peterson, S. R. et al. Loss of the catalytic subunit of the DNA-dependent protein kinase in DNA double-strand-break-repair mutant mammalian cells. *Proc. Natl Acad. Sci. USA* **92**, 3171–3174 (1995).
116. Fok, J. H. L. et al. AZD7648 is a potent and selective DNA-PK inhibitor that enhances radiation, chemotherapy and olaparib activity. *Nat. Commun.* **10**, 5065 (2019).
117. Zhao, Y. et al. Preclinical evaluation of a potent novel DNA-dependent protein kinase inhibitor NU7441. *Cancer Res.* **66**, 5354–5362 (2006).
118. Gurley, K. E. & Kemp, C. J. Synthetic lethality between mutation in Atm and DNA-PK(cs) during murine embryogenesis. *Curr. Biol.* **11**, 191–194 (2001).
119. Callen, E. et al. Essential role for DNA-PKcs in DNA double-strand break repair and apoptosis in ATM-deficient lymphocytes. *Mol. Cell* **34**, 285–297 (2009).
120. Riabinska, A. et al. Therapeutic targeting of a robust non-oncogene addiction to PRKDC in ATM-defective tumors. *Sci. Transl. Med.* **5**, 189ra178 (2013).
121. Cliby, W. A. et al. Overexpression of a kinase-inactive ATR protein causes sensitivity to DNA-damaging agents and defects in cell cycle checkpoints. *EMBO J.* **17**, 159–169 (1998).
122. Nghiem, P., Park, P. K., Kim, Y., Vaziri, C. & Schreiber, S. L. ATR inhibition selectively sensitizes G1 checkpoint-deficient cells to lethal premature chromatin condensation. *Proc. Natl Acad. Sci. USA* **98**, 9092–9097 (2001).
123. Toledo, L. I. et al. A cell-based screen identifies ATR inhibitors with synthetic lethal properties for cancer-associated mutations. *Nat. Struct. Mol. Biol.* **18**, 721–727 (2011).
124. Charrier, J. D. et al. Discovery of potent and selective inhibitors of ataxia telangiectasia mutated and Rad3 related (ATR) protein kinase as potential anticancer agents. *J. Med. Chem.* **54**, 2320–2330 (2011).
125. Foote, K. M. et al. Discovery and characterization of AZD6738, a potent inhibitor of ataxia telangiectasia mutated and Rad3 related (ATR) kinase with application as an anticancer agent. *J. Med. Chem.* **61**, 9889–9907 (2018).
126. Szydzik, J. et al. ATR inhibition enables complete tumour regression in ALK-driven NB mouse models. *Nat. Commun.* **12**, 6813 (2021).
127. Lloyd, R. L. et al. Loss of Cyclin C or CDK8 provides ATR inhibitor resistance by suppressing transcription-associated replication stress. *Nucleic Acids Res.* **49**, 8665–8683 (2021).
128. Kwok, M. et al. ATR inhibition induces synthetic lethality and overcomes chemoresistance in TP53- or ATM-defective chronic lymphocytic leukemia cells. *Blood* **127**, 582–595 (2016).
129. Reaper, P. M. et al. Selective killing of ATM- or p53-deficient cancer cells through inhibition of ATR. *Nat. Chem. Biol.* **7**, 428–430 (2011).
This study describes the first ATR inhibitor (VE-821) and shows that it can kill ATM-deficient tumours.
130. Tu, X. et al. ATR inhibition is a promising radiosensitizing strategy for triple-negative breast cancer. *Mol. Cancer Ther.* **17**, 2462–2472 (2018).
131. Dillon, M. T. et al. ATR inhibition potentiates the radiation-induced inflammatory tumor microenvironment. *Clin. Cancer Res.* **25**, 3392–3403 (2019).
132. Feng, X. et al. ATR inhibition potentiates ionizing radiation-induced interferon response via cytosolic nucleic acid-sensing pathways. *EMBO J.* **39**, e104036 (2020).
133. Tang, Z. et al. ATR inhibition induces CDK1-SPOP signaling and enhances anti-PD-L1 cytotoxicity in prostate cancer. *Clin. Cancer Res.* **27**, 4898–4909 (2021).
134. Thomas, A. et al. Therapeutic targeting of ATR yields durable regressions in small cell lung cancers with high replication stress. *Cancer Cell* **39**, 566–579.e567 (2021).
135. El Touny, L. H. et al. ATR inhibition reverses the resistance of homologous recombination deficient $MGMT^{low}/MMR^{proficient}$ cancer cells to temozolomide. *Oncotarget* **12**, 2114–2130 (2021).
136. Wallez, Y. et al. The ATR inhibitor AZD6738 synergizes with gemcitabine in vitro and in vivo to induce pancreatic ductal adenocarcinoma regression. *Mol. Cancer Ther.* **17**, 1670–1682 (2018).
137. Peasland, A. et al. Identification and evaluation of a potent novel ATR inhibitor, NU6027, in breast and ovarian cancer cell lines. *Br. J. Cancer* **105**, 372–381 (2011).
138. Kim, H. et al. Targeting the ATR/CHK1 axis with PARP inhibition results in tumor regression in BRCA-mutant ovarian cancer models. *Clin. Cancer Res.* **23**, 3097–3108 (2017).
139. Schoonen, P. M. et al. Premature mitotic entry induced by ATR inhibition potentiates olaparib inhibition-mediated genomic instability, inflammatory signaling, and cytotoxicity in BRCA2-deficient cancer cells. *Mol. Oncol.* **13**, 2422–2440 (2019).
140. Gralewska, P. et al. PARP Inhibition increases the reliance on ATR/CHK1 checkpoint signaling leading to synthetic lethality-an alternative treatment strategy for epithelial ovarian cancer cells independent from HR effectiveness. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms21249715 (2020).
141. Nam, A. R. et al. ATR inhibition amplifies antitumor effects of olaparib in biliary tract cancer. *Cancer Lett.* **516**, 38–47 (2021).
142. Yazinski, S. A. et al. ATR inhibition disrupts rewired homologous recombination and fork protection pathways in PARP inhibitor-resistant BRCA-deficient cancer cells. *Genes Dev.* **31**, 318–332 (2017).
143. Murai, J. et al. Resistance to PARP inhibitors by SLFN11 inactivation can be overcome by ATR inhibition. *Oncotarget* **7**, 76534–76550 (2016).
144. Ferrao, P. T., Bukczynska, E. P., Johnstone, R. W. & McArthur, G. A. Efficacy of CHK inhibitors as single agents in MYC-driven lymphoma cells. *Oncogene* **31**, 1661–1672 (2012).
145. Brooks, K. et al. A potent Chk1 inhibitor is selectively cytotoxic in melanomas with high levels of replicative stress. *Oncogene* **32**, 788–796 (2013).
146. Massey, A. J. et al. mTORC1 and DNA-PKcs as novel molecular determinants of sensitivity to Chk1 inhibition. *Mol. Oncol.* **10**, 101–112 (2016).
147. Ma, C. X. et al. Targeting Chk1 in p53-deficient triple-negative breast cancer is therapeutically beneficial in human-in-mouse tumor models. *J. Clin. Invest.* **122**, 1541–1552 (2012).
148. Booth, L. et al. PARP and CHK inhibitors interact to cause DNA damage and cell death in mammary carcinoma cells. *Cancer Biol. Ther.* **14**, 458–465 (2013).
149. Sen, T. et al. Targeting DNA damage response promotes antitumor immunity through STING-mediated T-cell activation in small cell lung cancer. *Cancer Discov.* **9**, 646–661 (2019).
150. Duda, H. et al. A mechanism for controlled breakage of under-replicated chromosomes during mitosis. *Dev. Cell* **39**, 740–755 (2016).
151. Pfister, S. X. et al. Inhibiting WEE1 selectively kills histone H3K36me3-deficient cancers by dNTP starvation. *Cancer Cell* **28**, 557–568 (2015).
152. ClinicalTrials.gov https://ClinicalTrials.gov/show/NCT03284385 (2017).
153. Gallo, D. et al. CCNE1 amplification is synthetic lethal with PKMYT1 kinase inhibition. *Nature* **604**, 749–756 (2022).
This study describes the development and use of a selective PKMYT1 inhibitor to target cyclin E1-overexpressing cancer cells.
154. Roerink, S. F., van Schendel, R. & Tijsterman, M. Polymerase theta-mediated end joining of replication-associated DNA breaks in C. elegans. *Genome Res.* **24**, 954–962 (2014).
155. Mateos-Gomez, P. A. et al. The helicase domain of Poltheta counteracts RPA to promote alt-NHEJ. *Nat. Struct. Mol. Biol.* **24**, 1116–1123 (2017).
156. Lu, G. et al. Ligase I and ligase III mediate the DNA double-strand break ligation in alternative end-joining. *Proc. Natl Acad. Sci. USA* **113**, 1256–1260 (2016).
157. Kent, T., Chandramouly, G., McDevitt, S. M., Ozdemir, A. Y. & Pomerantz, R. T. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. *Nat. Struct. Mol. Biol.* **22**, 230–237 (2015).
158. Wyatt, D. W. et al. Essential roles for polymerase theta-mediated end joining in the repair of chromosome breaks. *Mol. Cell* **63**, 662–673 (2016).
159. Higgins, G. S. et al. A small interfering RNA screen of genes involved in DNA repair identifies tumor-specific radiosensitization by POLQ knockdown. *Cancer Res.* **70**, 2984–2993 (2010).
160. Mateos-Gomez, P. A. et al. Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. *Nature* **518**, 254–257 (2015).
161. Shima, N., Munroe, R. J. & Schimenti, J. C. The mouse genomic instability mutation chaos1 is an allele of Polq that exhibits genetic interaction with Atm. *Mol. Cell Biol.* **24**, 10381–10389 (2004).
162. Zhou, J. et al. A first-in-class polymerase theta inhibitor selectively targets homologous-recombination-deficient tumors. *Nat. Cancer* **2**, 598–610 (2021).
163. Zatreanu, D. et al. Poltheta inhibitors elicit BRCA-gene synthetic lethality and target PARP inhibitor resistance. *Nat. Commun.* **12**, 3636 (2021).
Along with Zhou et al. (2021), this article describes the cytotoxicity of a POLQ inhibitor with regard to cancer cells that have acquired resistance to PARP inhibitors.
164. Nijman, S. M. et al. The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. *Mol. Cell* **17**, 331–339 (2005).
165. Huang, T. T. et al. Regulation of monoubiquitinated PCNA by DUB autocleavage. *Nat. Cell Biol.* **8**, 339–347 (2006).
166. Cohn, M. A. et al. A UAF1-containing multisubunit protein complex regulates the Fanconi anemia pathway. *Mol. Cell* **28**, 786–797 (2007).
167. Lim, K. S. et al. USP1 is required for replication fork protection in BRCA1-deficient tumors. *Mol. Cell* **72**, 925–941 e924 (2018).
168. Ulrich, H. D. & Takahashi, T. Readers of PCNA modifications. *Chromosoma* **122**, 259–274 (2013).
169. Mailand, N., Gibbs-Seymour, I. & Bekker-Jensen, S. Regulation of PCNA-protein interactions for genome stability. *Nat. Rev. Mol. Cell Biol.* **14**, 269–282 (2013).

170. Shenker, S. et al. Abstract 1337: Functional genomic characterization of the USP1 inhibitor KSQ-4279 reveals a distinct mechanism of action and resistance profile relative to other DDR targeting drugs. *Cancer Res.* **81**, 1337–1337 (2021).
171. ClinicalTrials.gov https://ClinicalTrials.gov/show/NCT05240898 (2022).
172. Turner, N. C. et al. A synthetic lethal siRNA screen identifying genes mediating sensitivity to a PARP inhibitor. *EMBO J.* **27**, 1368–1377 (2008).
173. McCabe, N. et al. Deficiency in the repair of DNA damage by homologous recombination and sensitivity to poly(ADP-ribose) polymerase inhibition. *Cancer Res.* **66**, 8109–8115 (2006).
174. Neijenhuis, S., Bajrami, I., Miller, R., Lord, C. J. & Ashworth, A. Identification of miRNA modulators to PARP inhibitor response. *DNA Repair* **12**, 394–402 (2013).
175. Grundy, M. K., Buckanovich, R. J. & Bernstein, K. A. Regulation and pharmacological targeting of RAD51 in cancer. *Nar. Cancer* **2**, zcaa024 (2020).
176. Lamont, K. R. et al. Attenuating homologous recombination stimulates an AID-induced antileukemic effect. *J. Exp. Med.* **210**, 1021–1033 (2013).
177. Maclay, T. et al. CYT01B, a Novel RAD51 inhibitor, act synergistically with both targeted and chemotherapeutic anti-cancer agents. *Blood* https://doi.org/10.1182/blood-2018-99-119373 (2018).
178. ClinicalTrials.gov https://ClinicalTrials.gov/show/NCT03997968 (2019).
179. Vilar, E. & Gruber, S. B. Microsatellite instability in colorectal cancer-the stable evidence. *Nat. Rev. Clin. Oncol.* **7**, 153–162 (2010).
180. Cortes-Ciriano, I., Lee, S., Park, W. Y., Kim, T. M. & Park, P. J. A molecular portrait of microsatellite instability across multiple cancers. *Nat. Commun.* **8**, 15180 (2017).
181. Ratti, M., Lampis, A., Hahne, J. C., Passalacqua, R. & Valeri, N. Microsatellite instability in gastric cancer: molecular bases, clinical perspectives, and new treatment approaches. *Cell Mol. Life Sci.* **75**, 4151–4162 (2018).
182. Chan, E. M. et al. WRN helicase is a synthetic lethal target in microsatellite unstable cancers. *Nature* **568**, 551–556 (2019).
183. Behan, F. M. et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511–516 (2019).
184. van Wietmarschen, N. et al. Repeat expansions confer WRN dependence in microsatellite-unstable cancers. *Nature* **586**, 292–298 (2020).
This study shows that MSI rather than MMR deficiency alone renders cancer cells dependent on WRN.
185. Tutt, A. N. et al. Exploiting the DNA repair defect in BRCA mutant cells in the design of new therapeutic strategies for cancer. *Cold Spring Harb. Symp. Quant. Biol.* **70**, 139–148 (2005).
186. Lindahl, T. & Barnes, D. E. Repair of endogenous DNA damage. *Cold Spring Harb. Symp. Quant. Biol.* **65**, 127–133 (2000).
187. Reislander, T., Groelly, F. J. & Tarsounas, M. DNA damage and cancer immunotherapy: a STING in the tale. *Mol. Cell* **80**, 21–28 (2020).
188. Ding, L. et al. PARP inhibition elicits STING-dependent antitumor immunity in Brca1-deficient ovarian cancer. *Cell Rep.* **25**, 2972–2980 e2975 (2018).
189. Pettitt, S. J. et al. Genome-wide and high-density CRISPR-Cas9 screens identify point mutations in PARP1 causing PARP inhibitor resistance. *Nat. Commun.* **9**, 1849 (2018).
190. Daniel, J. A. et al. Loss of ATM kinase activity leads to embryonic lethality in mice. *J. Cell Biol.* **198**, 295–304 (2012).
191. Yamamoto, K. et al. Kinase-dead ATM protein causes genomic instability and early embryonic lethality in mice. *J. Cell Biol.* **198**, 305–313 (2012).
192. Jiang, W. et al. Differential phosphorylation of DNA-PKcs regulates the interplay between end-processing and end-ligation during nonhomologous end-joining. *Mol. Cell* **58**, 172–185 (2015).
193. Koh, G., Degasperi, A., Zou, X., Momen, S. & Nik-Zainal, S. Mutational signatures: emerging concepts, caveats and clinical applications. *Nat. Rev. Cancer* **21**, 619–637 (2021).
194. Gourley, C. et al. Moving from Poly (ADP-Ribose) polymerase inhibition to targeting DNA repair and DNA damage response in cancer therapy. *J. Clin. Oncol.* **37**, 2257–2269 (2019).
195. Chabanon, R. M. et al. Targeting the DNA damage response in immuno-oncology: developments and opportunities. *Nat. Rev. Cancer* **21**, 701–717 (2021).
This Review covers the interplay between the DDR and the innate immune response.
196. Yap, T. A. et al. First-in-human trial of the oral ataxia telangiectasia and RAD3-related (ATR) inhibitor BAY 1895344 in patients with advanced solid tumors. *Cancer Discov.* **11**, 80–91 (2021).
197. Konstantinopoulos, P. A. et al. Berzosertib plus gemcitabine versus gemcitabine alone in platinum-resistant high-grade serous ovarian cancer: a multicentre, open-label, randomised, phase 2 trial. *Lancet Oncol.* **21**, 957–968 (2020).
198. Hong, D. S. et al. Preclinical evaluation and phase Ib study of prexasertib, a CHK1 inhibitor, and samotolisib (LY3023414), a dual PI3K/mTOR inhibitor. *Clin. Cancer Res.* **27**, 1864–1874 (2021).
199. Moore, K. N. et al. Adavosertib with chemotherapy in patients with primary platinum-resistant ovarian, fallopian tube, or peritoneal cancer: an open-label, four-arm, phase II study. *Clin. Cancer Res.* **28**, 36–44 (2022).
200. Cuneo, K. C. et al. Dose escalation trial of the Wee1 inhibitor adavosertib (AZD1775) in combination with gemcitabine and radiation for patients with locally advanced pancreatic cancer. *J. Clin. Oncol.* **37**, 2643–2650 (2019).
201. Kristeleit, R. et al. Rucaparib versus standard-of-care chemotherapy in patients with relapsed ovarian cancer and a deleterious BRCA1 or BRCA2 mutation (ARIEL4): an international, open-label, randomised, phase 3 trial. *Lancet Oncol.* **23**, 465–478 (2022).
202. Bang, Y. J. et al. Olaparib in combination with paclitaxel in patients with advanced gastric cancer who have progressed following first-line therapy (GOLD): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Oncol.* **18**, 1637–1651 (2017).
203. Ramalingam, S. S. et al. Veliparib in combination with platinum-based chemotherapy for first-line treatment of advanced squamous cell lung cancer: a randomized, multicenter phase III study. *J. Clin. Oncol.* **39**, 3633–3644 (2021).
204. Govindan, R. et al. Veliparib plus carboplatin and paclitaxel versus investigator's choice of standard chemotherapy in patients with advanced non-squamous non-small cell lung cancer. *Clin. Lung Cancer* **23**, 214–225 (2022).
205. Loibl, S. et al. Addition of the PARP inhibitor veliparib plus carboplatin or carboplatin alone to standard neoadjuvant chemotherapy in triple-negative breast cancer (BrighTNess): a randomised, phase 3 trial. *Lancet Oncol.* **19**, 497–509 (2018).
206. Dieras, V. et al. Veliparib with carboplatin and paclitaxel in BRCA-mutated advanced breast cancer (BROCADE3): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **21**, 1269–1282 (2020).
207. Coleman, R. L. et al. Veliparib with first-line chemotherapy and as maintenance therapy in ovarian cancer. *N. Engl. J. Med.* **381**, 2403–2415 (2019).
208. Ramalingam, S. S. et al. JASPER: phase 2 trial of first-line niraparib plus pembrolizumab in patients with advanced non-small cell lung cancer. *Cancer* **128**, 65–74 (2022).
209. Friedlander, M. et al. Pamiparib in combination with tislelizumab in patients with advanced solid tumours: results from the dose-escalation stage of a multicentre, open-label, phase 1a/b trial. *Lancet Oncol.* **20**, 1306–1315 (2019).
210. Sotiriou, S. K. et al. Mammalian RAD52 functions in break-induced replication repair of collapsed DNA replication forks. *Mol. Cell* **64**, 1127–1134 (2016).
211. Robson, M. et al. Olaparib for metastatic breast cancer in patients with a germline BRCA mutation. *N. Engl. J. Med.* **377**, 523–533 (2017).
212. Litton, J. K. et al. Talazoparib in patients with advanced breast cancer and a germline BRCA mutation. *N. Engl. J. Med.* **379**, 753–763 (2018).
213. Tutt, A. N. J. et al. Adjuvant olaparib for patients with BRCA1- or BRCA2-mutated breast cancer. *N. Engl. J. Med.* **384**, 2394–2405 (2021).
214. Moore, K. et al. Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer. *N. Engl. J. Med.* **379**, 2495–2505 (2018).
215. Golan, T. et al. Maintenance olaparib for germline BRCA-mutated metastatic pancreatic cancer. *N. Engl. J. Med.* **381**, 317–327 (2019).
216. Abida, W. et al. Rucaparib in men with metastatic castration-resistant prostate cancer harboring a BRCA1 or BRCA2 gene alteration. *J. Clin. Oncol.* **38**, 3763–3772 (2020).
217. Moore, K. N. et al. Niraparib monotherapy for late-line treatment of ovarian cancer (QUADRA): a multicentre, open-label, single-arm, phase 2 trial. *Lancet Oncol.* **20**, 636–648 (2019).
218. Ray-Coquard, I. et al. Olaparib plus bevacizumab as first-line maintenance in ovarian cancer. *N. Engl. J. Med.* **381**, 2416–2428 (2019).
219. Hussain, M. et al. Survival with olaparib in metastatic castration-resistant prostate cancer. *N. Engl. J. Med.* **383**, 2345–2357 (2020).
220. Gonzalez-Martin, A. et al. Niraparib in patients with newly diagnosed advanced ovarian cancer. *N. Engl. J. Med.* **381**, 2391–2402 (2019).
221. Mirza, M. R. et al. Niraparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. *N. Engl. J. Med.* **375**, 2154–2164 (2016).
222. Ledermann, J. A. et al. Overall survival in patients with platinum-sensitive recurrent serous ovarian cancer receiving olaparib maintenance monotherapy: an updated analysis from a randomised, placebo-controlled, double-blind, phase 2 trial. *Lancet Oncol.* **17**, 1579–1589 (2016).
223. Ledermann, J. A. et al. Rucaparib for patients with platinum-sensitive, recurrent ovarian carcinoma (ARIEL3): post-progression outcomes and updated safety results from a randomised, placebo-controlled, phase 3 trial. *Lancet Oncol.* **21**, 710–722 (2020).
224. Lee, A. Fuzuloparib: first approval. *Drugs* **81**, 1221–1226 (2021).
225. Markham, A. Pamiparib: first approval. *Drugs* **81**, 1343–1348 (2021).
226. Zandarashvili, L. et al. Structural basis for allosteric PARP-1 retention on DNA breaks. *Science* https://doi.org/10.1126/science.aax6367 (2020).
227. Hoppe, M. M., Sundar, R., Tan, D. S. P. & Jeyasekharan, A. D. Biomarkers for homologous recombination deficiency in cancer. *J. Natl Cancer Inst.* **110**, 704–713 (2018).
228. Mukhopadhyay, A. et al. Development of a functional assay for homologous recombination status in primary cultures of epithelial ovarian tumor and correlation with sensitivity to poly(ADP-ribose) polymerase inhibitors. *Clin. Cancer Res.* **16**, 2344–2351 (2010).
229. Castroviejo-Bermejo, M. et al. A RAD51 assay feasible in routine tumor samples calls PARP inhibitor response beyond BRCA mutation. *EMBO Mol. Med.* https://doi.org/10.15252/emmm.201809172 (2018).
230. Cruz, C. et al. RAD51 foci as a functional biomarker of homologous recombination repair and PARP inhibitor resistance in germline BRCA-mutated breast cancer. *Ann. Oncol.* **29**, 1203–1210 (2018).
This study shows that a formation of RAD51 foci assay on diagnostic formalin-fixed, paraffin-embedded tumour samples can predict response to PARP inhibitors.
231. Pettitt, S. J. et al. Clinical BRCA1/2 reversion analysis identifies hotspot mutations and predicted neoantigens associated with therapy resistance. *Cancer Discov.* **10**, 1475–1488 (2020).
232. Swisher, E. M. et al. Rucaparib in relapsed, platinum-sensitive high-grade ovarian carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial. *Lancet Oncol.* **18**, 75–87 (2017).

233. Ter Brugge, P. et al. Mechanisms of therapy resistance in patient-derived xenograft models of BRCA1-deficient breast cancer. *J. Natl Cancer Inst.* https://doi.org/10.1093/jnci/djw148 (2016).
234. Jaspers, J. E. et al. Loss of 53BP1 causes PARP inhibitor resistance in Brca1-mutated mouse mammary tumors. *Cancer Discov.* **3**, 68–81 (2013).
235. Escribano-Diaz, C. et al. A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol. Cell* **49**, 872–883 (2013).
236. Zimmermann, M., Lottersberger, F., Buonomo, S. B., Sfeir, A. & de Lange, T. 53BP1 regulates DSB repair using Rif1 to control 5′ end resection. *Science* **339**, 700–704 (2013).
237. Di Virgilio, M. et al. Rif1 prevents resection of DNA breaks and promotes immunoglobulin class switching. *Science* **339**, 711–715 (2013).
238. Chapman, J. R. et al. RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. *Mol. Cell* **49**, 858–871 (2013).
239. Dev, H. et al. Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. *Nat. Cell Biol.* **20**, 954–965 (2018).
240. Gupta, R. et al. DNA repair network analysis reveals shieldin as a key regulator of NHEJ and PARP inhibitor sensitivity. *Cell* **173**, 972–988 e923 (2018).
241. Ghezraoui, H. et al. 53BP1 cooperation with the REV7-shieldin complex underpins DNA structure-specific NHEJ. *Nature* **560**, 122–127 (2018).
242. Noordermeer, S. M. et al. The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* **560**, 117–121 (2018).
243. Gogola, E. et al. Selective loss of PARG restores PARylation and counteracts PARP inhibitor-mediated synthetic lethality. *Cancer Cell* **33**, 1078–1093 (2018).

Acknowledgements
The laboratories of the authors are supported by the University of Oxford and a Cancer Research UK Programme Award (DRCPGM\100001) to M.T. and a Cancer Research UK Career Development Fellowship (C29215/A20772) to A.N.B.

Author contributions
All authors researched data for the article and contributed substantially to discussion of the content. F.J.G., A.N.B. and M.T. wrote the article and reviewed and/or edited the manuscript before submission.

Competing interests
The authors declare no competing interests

Additional information
Correspondence should be addressed to Andrew N. Blackford or Madalena Tarsounas.

Peer review information *Nature Reviews Cancer* thanks Nicola Curtin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Related links
ClinicalTrials.gov: https://clinicaltrials.gov/
European Medicines Agency: https://www.ema.europa.eu/

© Springer Nature Limited 2022