# TEX264 in DNA Replication and Genome Stability

John Fielden

St Catherine's College




Supervisor: Professor Kristijan Ramadan

A thesis submitted to the University of Oxford for the degree of Doctor of Philosophy

Trinity Term 2019

CRUK and MRC Oxford Institute for Radiation Oncology

Department of Oncology

University of Oxford

# Contents

Acknowledgments ..... iii
Declaration ..... iv
Abstract ..... v
List of figures ..... vi
List of abbreviations ..... vii
1 Introduction ..... 1
1.1 DNA replication ..... 2
1.1.1 DNA replication initiation ..... 2
1.1.2 DNA replication termination ..... 4
1.1.3 DNA replication stress ..... 6
1.2 Mechanisms of DNA repair ..... 9
1.2.1 Base excision repair (BER) ..... 9
1.2.2 Nucleotide excision repair (NER) ..... 9
1.2.3 Double strand break (DSB) repair ..... 12
1.2.4 DNA damage tolerance (DDT) ..... 15
1.2.5 Inter-strand crosslink (ICL) repair ..... 17
1.2.6 DNA-protein crosslink repair ..... 20
1.3 Aims of this study ..... 45
2 Materials and Methods ..... 46
2.1 Buffers ..... 46
2.1.1 10 x Phosphate Buffered Saline (PBS): ..... 46
2.1.2 10x Tris Buffered Saline (TBS) ..... 46
2.1.3 10x Tris/Glycine buffer ..... 46
2.1.4 5x Laemmli Buffer ..... 47
2.1.5 Luria-Bertani (LB) broth and agar ..... 47
2.1.6 Crystal violet staining solution ..... 48
2.2 Cell culture ..... 48
2.2.1 Cell line maintenance ..... 48
2.2.2 Generation of cell lines ..... 48
2.2.3 siRNA transfection ..... 50
2.2.4 Plasmid DNA transfection ..... 50
2.3 Molecular and cellular biology techniques ..... 51
2.3.1 PCR for Site-Directed Mutagenesis ..... 51
2.3.2 SDS-PAGE and immunoblotting ..... 52
2.3.3 Preparation of whole cell lysates ..... 53

2.3.4 Cellular fractionations ... 53
2.3.5 Immunoprecipitation (IP) ... 54
2.3.6 SUMO binding assay ... 55
2.3.7 Isolation of Proteins on Nascent DNA (iPOND) ... 55
2.3.8 Protein purification & *in vitro* interaction assays ... 56
2.3.9 Rapid Approach to DNA Adduct Recovery (RADAR) ... 57
2.3.10 Colony forming assay ... 57
2.3.11 *In vitro* TOP1cc repair assay ... 58
2.3.12 Immunofluorescence ... 58
2.3.13 DNA fibre assay ... 59
2.3.14 Alkaline comet assay ... 60
2.4 Immunohistochemistry and mRNA expression analysis ... 60
2.5 Mass spectrometry ... 62
2.6 Statistical analysis ... 63
2.7 Antibodies used in this study ... 63
3 Results ... 66
3.1 SPRTN promotes DPC repair and DNA replication ... 66
3.2 Identification of new TOP1cc repair proteins ... 73
3.3 Identification of TEX264 ... 79
3.4 The role of TEX264 in TOP1cc repair ... 83
3.4.1 TEX264 physically bridges p97 and TOP1 ... 83
3.4.2 TEX264 counteracts TOP1ccs ... 89
3.4.3 The GyrI-like domain of TEX264 binds TOP1 ... 96
3.4.4 TEX264 binds SUMO1 ... 98
3.4.5 TEX264 cooperates with SPRTN in TOP1cc repair ... 106
3.4.6 TEX264 acts at DNA replication forks ... 112
3.5 SPRTN as a biomarker ... 121
3.5.1 High SPRTN expression correlates with poor prognosis in metastatic colorectal cancer ... 121
4 Discussion ... 127
4.1 TEX264 at the INM ... 128
4.2 The role of SUMO ... 131
4.2.1 ACRC: a SUMO-targeted DPC repair protease? ... 132
4.3 Novel insights into SPRTN function ... 134
4.4 SPRTN as a driver of irinotecan resistance ... 138
5 Future work ... 140
References ... 142

# Acknowledgments

I have been fortunate to spend four years in the company of some of the smartest and ambitious people you could meet.

Kristijan, your passion and ambition has motivated me for four years. In your lab, I felt as if no challenge was too great. I will always be grateful for the confidence you had in me and my research and for helping to make this DPhil a success. Sam, I will miss our late-night dinners in the atrium and discussing science with you. I hope we continue to share our thoughts on the latest papers for a long time yet. Shudong, it was great having your company both inside and outside the lab. I always enjoyed discussing lab life and scientific problems with you. Annamaria, you were always the first person I looked for to share data, ideas, and general thoughts. Your sense of humour brought so much fun to the lab. Marta, thank you for helping me settle in the lab. The nights we spent at the White Horse will be some of my most enduring, albeit hazy, memories of Oxford. Bruno, I was always impressed by your ability to tackle new scientific challenges in creative ways without hesitation. It was a valuable lesson for me. Swagata, your passion for science was a vital source of inspiration for me in the early stages of my DPhil. I could listen to you talk (always with a smile on your face) about science for hours. Ignacio, your professionalism and perseverance were so important in helping the lab function. Thank you for all your valuable advice and, of course, for identifying TEX264 in your mass spec. Abhay, thank you for all your calm and rational advice through the years. Katy, thank you for the great work on TEX264.

To Judith, the cleverest and kindest person I know. Thank you for your endless love and support. You made me a better scientist and person.

To my family – Mum, Dad, and Rachel – you all encouraged me through the toughest of times and gave me confidence. Thank you for supporting and believing in me.

# Declaration

I declare that the work presented in this thesis is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree at this or any other university.

# Abstract

DNA damage drives numerous human pathologies, including cancer and ageing. However, it can also be exploited in cancer therapy, where most chemotherapeutics exert their effects by inducing DNA damage. A common type of DNA lesion induced by chemotherapeutics is DNA-protein crosslinks (DPC). DPCs are also very common endogenous DNA lesions which impede DNA replication, transcription, and mitosis. However, very little is known about the mechanisms by which they are repaired.

Our laboratory recently characterised a genetic syndrome, known as Ruijs-Aalfs syndrome (RJALS), characterised by premature ageing and early onset hepatocellular carcinoma, caused by mutations in the *SPRTN* gene. This thesis contributes to the characterisation of SPRTN as an essential DPC repair protease that promotes DNA replication progression.

We identify topoisomerase 1 cleavage complexes (TOP1cc) as SPRTN substrates. TOP1ccs are composed of the enzyme, TOP1, covalently bound to DNA. TOP1ccs are highly cytotoxic endogenous lesions which can also be induced by the frontline class of anti-cancer drugs known as camptothecins. In addition, we identify the previously uncharacterised protein, testes expressed 264 (TEX264), as being critical for TOP1cc repair. TEX264 recognises unmodified and SUMO1-modified TOP1 and promotes TOP1cc repair by recruiting p97 and SPRTN. Thus, TEX264 promotes DNA replication fork progression and preserves genome stability. Furthermore, this study identifies SPRTN as a biomarker of clinical resistance to the camptothecin derivative, irinotecan, in patients with metastatic colorectal cancer.

# List of figures

Figure 1. Basic model of DNA replication initiation, elongation, and termination. 5
Figure 2. A simplified model of ATR activation upon replication stress. 8
Figure 3. General principles of excision repair. 11
Figure 4. Pathways involved in ICL repair. 19
Figure 5. DPC repair pathways. 24
Figure 6. Schematic diagram of DNA damage tolerance by translesion synthesis (TLS). 28
Figure 7. SPRTN mutations cause early onset ageing and hepatocellular carcinoma. 32
Figure 8. Models of substrate processing and cofactor binding by p97. 35
Figure 9. Generation of ΔSPRTN HeLa cells by CRISPR-Cas9. 67
Figure 10. SPRTN counteracts DPC accumulation. 69
Figure 11. SPRTN regulates DNA replication fork progression. 72
Figure 12. p97 interacts with TOP1. 75
Figure 13. p97 counteracts TOP1ccs. 77
Figure 14. p97 ATPase activity is required for TOP1cc repair. 78
Figure 15. Identification of TEX264. 80
Figure 16. TEX264 is a vertebrate-specific Gyrase inhibitory-like protein. 82
Figure 17. TEX264 subcellular localisation and expression through the cell cycle. 83
Figure 18. TEX264 bridges p97 and TOP1. 86
Figure 19. TEX264 binds directly to the N terminus of p97 via its SHP box and bridges p97 to TOP1 *in vitro.* 88
Figure 20. TEX264 counteracts TOP1cc formation. 89
Figure 21. TEX264- and TDP1-deficient cells exhibit delayed DNA damage repair kinetics after CPT treatment. 92
Figure 22. TEX264, p97, and TDP1 act in the same TOP1cc repair pathway. 94
Figure 23. Identification of TOP1-binding residues in TEX264. 97
Figure 24. TEX264 binds SUMO1 via its SIMs. 99
Figure 25. SUMOylation promotes TOP1cc repair. 101
Figure 26. TEX264 binds SUMO1-modified TOP1. 104
Figure 27. SIM2 in TEX264 is required for TOP1cc repair. 105
Figure 28. TEX264 cooperates with SPRTN in TOP1cc repair. 108
Figure 29. p97 and TEX264 promote SPRTN-dependent TOP1cc repair *in vitro*. 110
Figure 30. TEX264 does not regulate general DPC repair. 111
Figure 31. TEX264 interacts with DNA replication factors. 112
Figure 32. Immunofluorescence confocal microscopy analysis of TEX264. 115
Figure 33. TEX264 acts at DNA replication forks. 117
Figure 34. TOP1ccs drive DNA damage in TEX264-deficient cells. 119
Figure 35. Proposed model. 120
Figure 36. SPRTN expression correlates with resistance to irinotecan in mCRC. 123
Figure 37. SPRTN protein expression correlates with resistance to irinotecan in mCRC patients. 124
Figure 38. TEX264 expression does not correlate with resistance to irinotecan in mCRC patients. 125
Figure 39. p97 expression does not correlate with resistance to irinotecan in mCRC patients. 126
Figure 40. Phylogenetic relationship of SPRT and WLM proteases. 133

# List of abbreviations

| | |
|---|---|
| 53BP1 | p53 Binding Protein 1 |
| 5-aza-dC | 5-aza-2′-deoxycytidine |
| 6-4 PP | 6-4 Photoproduct |
| 8-oxoG | 8-oxoguanine |
| A | Alanine |
| AA | Amino Acid |
| ACRC | Acidic Repeat Containing |
| ALDH | Aldehyde Dehydrogenase |
| ANOVA | Analysis Of Variance |
| AP | Apurinic/apyrimidinic |
| APC/C | Anaphase-Promoting Complex Cyclosome |
| ATM | Ataxia-Telangiectasia Mutated |
| ATP | Adenosine Triphosphate |
| ATR | Ataxia Telangiectasia and Rad3-Related Protein |
| ATRIP | ATR-Interacting Protein |
| BER | Base Excision Repair |
| BP | Base Pair |
| BRCA | Breast Cancer Susceptibility Gene |
| C | Carboxy/cytosine |
| C. elegans | Caenorhabditis elegans |
| CB-5083 | Cleave Biosciences 5083 |
| Cdc | Cell Division Cycle |
| Cdh1 | Cadherin 1 |
| CDK | Cyclin-Dependent Kinase |
| cDNA | Complementary DNA |
| CDT1 | Chromatin Licensing and DNA Replication Factor 1 |
| CHK | Checkpoint Kinase |
| CMG | CDC45, GINS, MCM |
| CPD | Cyclobutane Pyrimidine Dimers |
| CPT | Camptothecin |
| CRISPR | Clustered Regularly Interspaced Short Palindromic Repeats |
| CRL | Cullin RING-Based E3 Ligase |
| CsCl | Caesium Chloride |
| CtIP | C-Terminal Binding Protein 1 Interacting Protein |

| | |
|---|---|
| CUL | Cullin |
| CYREN | Cell Cycle Regulator of NHEJ |
| D | Aspartic Acid |
| Da | Dalton |
| DAPI | 4',6-Diamidino-2-Phenylindole |
| DDB | DNA Damage-Binding Protein |
| ddH2O | Double Distilled Water |
| Ddi1 | DNA Damage Inducible 1 |
| DDK | Dbf4-Dependent Kinase |
| DDT | DNA-Damage Tolerance |
| Dia2 | Digs Into Agar 2 |
| DMSO | Dimethyl Sulfoxide |
| DNA | Deoxyribonucleic Acid |
| DNA2 | DNA Replication Helicase/Nuclease 2 |
| DNA-PK | DNA-Dependent Protein Kinase |
| DNA-PKcs | DNA-Dependent Protein Kinase Catalytic Subunit |
| DNMT | DNA Methyltransferase |
| Doa1 | Degradation Of Alpha |
| DOX | Doxycycline |
| DPC | DNA-Protein Crosslink |
| DSB | Double Strand Break |
| dsDNA | Double Stranded DNA |
| E | Glutamic Acid |
| EDTA | Ethylene Diamine Tetraacetic Acid |
| EdU | 5-ethynyl-2'-deoxyuridine |
| ER | Endoplasmic Reticulum |
| ERCC1 | Excision Repair Cross-Complementation Group 1 (XPF) |
| ETAA1 | Ewing’s Tumour-Associated Antigen 1 |
| EXO1 | Exonuclease 1 |
| F | Phenylalanine |
| FA | Fanconi Anaemia |
| FAAP24 | Fanconi Anaemia core complex Associated Protein 24 |
| FAF1 | Fas Associated Factor 1 |
| FANC | FA Complementation Group |
| FBS | Fetal Bovine Serum |
| FDA | Food and Drug Administration |

| | |
|---|---|
| FOLFIRI | Folinic acid, Fluorouracil and Irinotecan |
| G | Glycine |
| g | Gram |
| G0/1/2 | Gap 0/1/2 |
| GCNA | Germ Cell Nuclear Antigen |
| GFP | Green Fluorescent Protein |
| GG-NER | Global Genome Nucleotide Excision Repair |
| GINS | Go, Ichi, Nii and San |
| GO | Gene Ontology |
| Gy | Gray |
| GyrI | Gyrase Inhibitory Protein |
| HEK293 | Human Embryonic Kidney 293 |
| HER2 | Human Epidermal Growth Factor Receptor 2 |
| His | Histidine |
| HR | Homologous Recombination |
| hr | Hour(s) |
| H-score | Histological score |
| I | Isoleucine |
| ICL | Inter-Strand Crosslink |
| IHC | Immunohistochemistry |
| INM | Inner Nuclear Membrane |
| iPOND | Isolation of Proteins On Nascent DNA |
| IR | Ionising Radiation |
| K | Lysine |
| kDa | Kilodalton |
| L | Leucine/Litre |
| LB | Luria-Bertani |
| LC3 | Light Chain 3 |
| LC-MS/MS | Tandem Liquid Chromatography Mass Spectrometry |
| LIR | LC3-Interacting Region |
| LRR | Leucine-Rich Repeat |
| LV | Leucovorin/Folinic Acid |
| M | Molar/Mitosis |
| MAFFT | Multiple Alignment Using Fast Fourier Transform |
| MCM | Minichromosome Maintenance |
| mCRC | Metastatic Colorectal Cancer |

| | |
|---|---|
| MDC1 | Mediator of DNA Damage Checkpoint Protein 1 |
| MEF | Mouse Embryonic Fibroblast |
| Met | Methionine |
| min | Minute(s) |
| mL | Millilitre(s) |
| mM | Millimolar |
| MMS | Methyl Methansulfonate |
| MRC | Medical Research Council |
| MRE11 | Meiotic Recombination 11 |
| MRN | MRE11-RAD50-NBS1 |
| mRNA | Messenger Ribonucleic Acid |
| MS | Mass Spectrometry |
| N | Amino |
| NBS1 | Nijmegen Breakage Syndrome Protein 1 |
| NCI | National Cancer Institute |
| NEIL3 | Nei Like DNA Glycosylase 3 |
| NER | Nucleotide Excision Repair |
| NHEJ | Non-Homologous End Joining |
| nM | Nanomole |
| nm | Nanometre(s) |
| NPC | Nuclear Pore Complex |
| NPL4 | Nuclear Protein Localisation Protein 4 |
| ns | Not Significant |
| OGG1 | 8-Oxoguanine DNA Glycosylase |
| ORC | Origin Recognition Complex |
| ORI | Origin of Replication |
| PAGE | Polyacrylamide Gel Electrophoresis |
| PALB2 | Partner and Localizer of BRCA2 |
| PARP1/2 | Poly [ADP-Ribose] Polymerase 1/2 |
| PARPi | Poly [ADP-Ribose] Polymerase Inhibitor |
| PAXX | Paralog of XRCC4 and XLF |
| PBS | Phosphate-Buffered Saline |
| PCNA | Proliferating Cell Nuclear Antigen |
| PCR | Polymerase Chain Reaction |
| Phyre | Protein Homology/AnalogY Recognition Engine |
| PIP | PCNA-Interacting Protein |

| | |
|---|---|
| Pol | Polymerase |
| Pre-RC | Pre-Replicative Complex |
| PSMD | 26S Proteasome Non-ATPase Regulatory Subunit |
| PTM | Posttranslational Modification |
| PUB | PNGase (Peptide N-Glycosidase)/Ubiquitin Associated |
| Q | Glutamine |
| Rad | Radiation Sensitive |
| RADAR | Rapid Approach to DNA Adducts Recovery |
| REV | Reversionless |
| RFC | Replication Factor C |
| RHINO | Rad9, Hus1, Rad1 Interacting Nuclear Orphan |
| RIF1 | Rap1-Interacting Factor 1 |
| RING | Really Interesting New Gene |
| RIPA | Radioimmunoprecipitation Assay |
| RJALS | Ruijs-Aalfs Syndrome |
| RNA | Ribonucleic Acid |
| RNAP | RNA polymerase |
| RNF | RING Finger Protein |
| RNS | Reactive Nitrogen Species |
| ROS | Reactive Oxygen Species |
| RPA | Replication Protein A |
| RPE-1 | Retinal Pigmented Epithelial 1 |
| RTEL1 | Regulator of Telomere Elongation helicase 1 |
| S Phase | Synthesis Phase |
| sbmC | Sensitivity B17 Microcin |
| SCAN1 | Spinocerebellar Ataxia with Axonal Neuropathy 1 |
| SCAR23 | Spinocerebellar Ataxia Autosomal Recessive 23 |
| SCF | SKP, Cullin, F-Box Containing Complex |
| SDS | Sodium Dodecyl Sulphate |
| SEM/SD | Standard Error of the Mean/Standard Deviation |
| S/Ser | Serine |
| sgRNA | Single Guide RNA |
| SHP | Src Homology 2-containing Phosphatase |
| SILAC | Stable Isotope Labelling by/with Amino acids in Cell culture |
| SIM | SUMO Interacting Motif |
| siRNA | Short Interfering RNA |

| | |
|---|---|
| Sld | Synthetic Lethality with Dpb11 |
| SLX | Synthetic Lethal of Unknown Function |
| Smt3p | Suppressor of Mif Two |
| SPRTN | SprT-like N-terminal domain |
| SSBs | Single Strand Breaks |
| ssDNA | Single Stranded DNA |
| SSH | Strep-Strep- haemagglutinin |
| STUbl | SUMO-Targeted Ubiquitin Ligase |
| SUMO | Small Ubiquitin-Like Modifier |
| SV40 | Simian Vacuolating Virus 40 |
| T-antigen | Tumour Antigen |
| TAP | Tandem Affinity Purification |
| TBS | Tris-Buffered Saline |
| TC-NER | Transcription-Coupled Nucleotide Excision Repair |
| TDP | Tyrosyl-DNA Phosphodiesterase |
| TEX264 | Testes Expressed 264 |
| TFIIH | Transcription Factor II H |
| TLS | Trans-Lesion Synthesis |
| TOP | Topoisomerase |
| TopBP1 | DNA Topoisomerase II Binding Protein 1 |
| TOP-cc | Topoisomerase Cleavage Complex |
| TRAIP | TRAF Interacting Protein |
| U2OS | U2-Osteosarcoma |
| Ub | Ubiquitin |
| UBC | Ubiquitin-Conjugating |
| UBE | Ubiquitin Conjugating Enzyme |
| UBL | Ubiquitin-like proteins |
| UBX | Ubiquitin Regulatory X |
| UBXD | UBX Domain-containing protein |
| UBXL | Ubiquitin Regulatory X Like |
| UBZ | Ubiquitin-Binding Zinc Finger |
| UFD1 | Ubiquitin Fusion Degradation Protein 1 |
| UHRF | Ubiquitin Like with PHD and Ring Finger Domains |
| UV | Ultraviolet |
| VBM | VCP-Binding Motif |
| VCP | Valosin Containing Protein |

| | |
|---|---|
| VIM | VCP Interacting Motif |
| W | Tryptophan |
| WCE | Whole Cell Extract |
| WLM | Wss1p-like Metalloproteases |
| Wss1 | Weak Suppressor of Smt3-1 |
| WT | Wild-Type |
| XLF | XRCC4-Like Factor Ubiquitin Like with PHD and Ring Finger Domains |
| XP | Xeroderma Pigmentosum Complementation Group |
| XRCC | X-Ray Repair Cross Complementing |
| Y | Tyrosine |
| YFP | Yellow Fluorescent Protein |
| ZNF | Zinc Finger |
| γH2AX | Phosphorylated histone H2AX |
| Δ | Deletion |

# 1 Introduction

During each cell cycle, genetic information must be duplicated accurately, and transmitted equally, into the dividing daughter cells. The process of duplicating genetic information is referred to as DNA replication and, in humans, requires that over 6 billion base pairs of DNA are copied within a matter of hours (Fig. 1). This challenge is made more difficult by numerous obstacles to DNA replication, such as common fragile sites, repetitive DNA sequences, and DNA secondary structures (Zeman and Cimprich, 2014). In addition to these structural impediments, accurate genome duplication is threatened by DNA lesions, induced by exogenous agents such as ultraviolet radiation (UV light) and chemotherapeutics, or by endogenous by-products of metabolism such as reactive oxygen species (ROS) and formaldehyde (Tubbs and Nussenzweig, 2017).

The failure to complete DNA replication or repair DNA lesions can lead to genetic information being lost or inaccurately copied. Ultimately, this can lead to cell death and disease. Cells have therefore evolved numerous DNA repair, DNA damage tolerance, and DNA replication stress response pathways to cope with these diverse threats. Collectively, these pathways coordinate DNA repair and replication and regulate the progression of the cell cycle to ensure genome integrity is preserved. The importance of these pathways to human health is illustrated by the numerous diseases that arise when their components are defective (Rivera et al., 2018). Moreover, while defects in these pathways can drive malignancies, they also present therapeutic vulnerabilities that can be exploited clinically (Brown et al., 2017).

## 1.1 DNA replication

DNA replication is the process by which the genome is duplicated before cell division occurs. Briefly, a replicative helicase unwinds the DNA duplex, exposing two single-stranded DNA templates. DNA polymerases then synthesise new DNA strands, complementary to each of the template strands, generating two DNA helices. This process is referred to as semi-conservative replication, because each new DNA helix is composed of one DNA strand from the parent helix and one newly synthesised strand (Meselson and Stahl, 1958).

All DNA polymerases synthesise DNA in the 5' to 3' direction. The leading strand, which is oriented in the 3' to 5' direction, can be synthesised continuously from each replication origin by polymerase epsilon (ε). The 5' to 3' lagging stranded is synthesised in discontinuous 100 – 200 base pair (bp) stretches known as Okazaki fragments by polymerase delta (δ) (Burgers, 2009; Okazaki et al., 1968).

### 1.1.1 DNA replication initiation

The initiation of DNA replication is tightly controlled to ensure that the genome is duplicated only once per cell cycle (Fig. 1). DNA replication is initiated from DNA replication origins. Replication origins (ORI) in yeast and bacteria are well-defined. However, in mammalian cells, the location and identity of ORI has been more difficult to define due in part to their larger genomes and apparent ability to initiate DNA replication from thousands of redundant ORI (Prioleau and MacAlpine, 2016). Recent evidence suggests that replication in human cells is initiated from nucleosome-depleted, poly-A/T tails between transcribed DNA regions (Tubbs et al., 2018).

The initiation of DNA replication occurs in two distinct stages. In the Gap 1 (G1) phase of the cell cycle, the minichromosome maintenance (MCM) complex, composed of MCM2-MCM7, is loaded onto DNA as a double hexamer by cell division control protein 6 (Cdc6), the origin recognition complex (ORC), and chromatin licensing and DNA replication factor 1 (Cdt1) (Remus et al., 2009). In the synthesis (S) phase of the cell cycle, origin firing is activated by the recruitment of Cdc45, the Go-Ichi-Ni-San (GINS) complex, and Pol ε. Cdc45 (along with its binding partners Sld2 and Sld3) is recruited by the Dbf4-dependent kinase (DDK, composed of Cdc7 and Dbf4) which phosphorylates MCM4 and MCM6. GINS and Pol ε are recruited by Cyclin A-cyclin-dependent kinase (CDK)-dependent phosphorylation of Sld2 and Sld3 (Gambus et al., 2006; Muramatsu et al., 2010; Zegerman and Diffley, 2007).

The CMG complex (Cdc45-MCM-GINS) initiates DNA unwinding. The resulting single-stranded (ss)DNA is bound by replication protein A (RPA) which recruits the DNA polymerase alpha (Pol α)-primase complex. Primase synthesizes 7-12 ribonucleotide-long primers on both strands of DNA, which are extended with deoxyribonucleotides by Pol α (Pellegrini, 2012). Replication factor C (RFC) recognises this RNA-DNA primer and loads the sliding clamp proliferating cell nuclear antigen (PCNA), which encircles ssDNA on both strands of the DNA duplex (Bowman et al., 2004). PCNA acts as a platform for binding by Pol ε and Pol δ, on the leading and lagging DNA strands, respectively. Pol ε and Pol δ can begin synthesising new DNA strands from the ends of the DNA primers (Fig. 1) (Burgers, 2009).

### 1.1.2 DNA replication termination

In eukaryotic cells, DNA replication forks terminate when two forks converge at a random site (Santamaria, 2000) (Fig. 1). As forks converge, the leading strand of one fork fuses with the lagging strand of the opposing fork. In yeast and bacteria, converging replication forks are thought to stall, then be dissolved through the activity of topoisomerase 2 (TOP2) (Baxter and Diffley, 2008; Fachinetti et al., 2010). However, there is mixed evidence as to whether fork stalling and TOP2 are required for replication fork termination in vertebrates (Dewar et al., 2015; Moreno et al., 2014).

A mechanism of replication fork termination conserved from yeast to vertebrates is regulated by the ATPase p97 (also known as VCP in mammals and Cdc48 in yeast) (Maric et al., 2014; Moreno et al., 2014). Late in S phase, K48-linked poly-ubiquitin chains are conjugated to MCM7. This ubiquitination is driven by the E3 ubiquitin ligases $SCF^{Dia2}$ in budding yeast and $CRL2^{Lrr1}$ in vertebrates (Dewar et al., 2017; Maric et al., 2014). p97 is recruited to terminating replication forks by its cofactors, Ufd1-Npl4, which recognise ubiquitinated MCM7 (Maric et al., 2017). p97 promotes replication fork termination by unloading MCM7 from chromatin, thereby disassembling the entire CMG helicase.

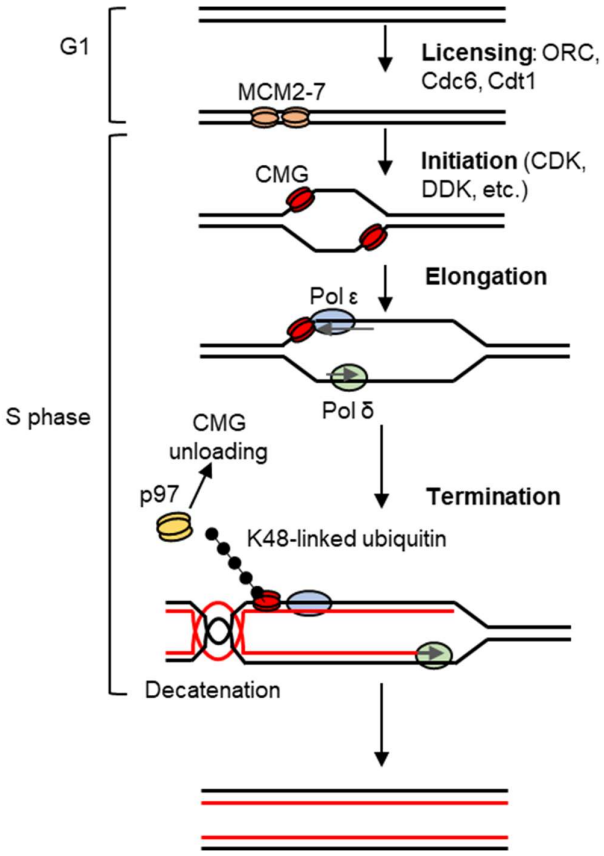



**Figure 1. Basic model of DNA replication initiation, elongation, and termination.**

Origin licensing occurs in the G1 phase of the cell cycle when the ORC complex, Cdc6, and CDT1 recruit the MCM2-7 complex to DNA origins. This complex of proteins is referred to as a pre-replicative complex (pre-RC). In S phase, a subset of pre-RCs are phosphorylated by CDK and DDK, leading to the recruitment of Cdc45 and the GINS complex and the formation of an active CMG helicase. The CMG unwinds origins to allow assembly of DNA polymerases (and PCNA). DNA polymerases (Pol) synthesise nascent DNA stands (red lines). Termination occurs when DNA replication forks converge. The MCM7 subunit is polyubiquitinated and unloaded from chromatin by the ATPase p97. DNA catenation that arises during fork convergence is resolved.

### 1.1.3 DNA replication stress

A variety of factors, including DNA lesions, nucleotide insufficiency, and difficult-to-replicate genomic regions, can hamper the progression of DNA replication forks. This can lead to fork slowing or stalling, a phenomenon referred to as DNA replication stress (Fig. 2) (Gaillard et al., 2015). The concept of DNA replication stress could also be extended to include excessive fork speed (Maya-Mendoza et al., 2018; Somyajit et al., 2017). DNA replication stress can lead to replication fork collapse, a failure to complete DNA replication, and cell death. DNA replication stress is a major driver of the genomic instability associated with human diseases such as cancer and premature aging. Cells have evolved genome surveillance mechanisms to prevent the collapse of stalled forks, initiate DNA repair, and ensure replication is completed before cell division (Zeman and Cimprich, 2014).

DNA replication stress is frequently associated with the generation of ssDNA at replication forks (Fig. 2). ssDNA can be generated when a DNA polymerase stalls (e.g. at a DNA lesion), while the CMG helicase continues to unwind the DNA duplex ahead of the polymerase (Byun et al., 2005). ssDNA can also be generated by replication fork reversal or enzymatic processing of a newly synthesised DNA strand (Thangavel et al., 2015; Zellweger et al., 2015). The mechanism by which ssDNA is generated likely depends on the type of lesion or stress that hampers fork progression. In addition to replication stress, ssDNA can be generated by the nucleolytic processing of DNA double-strand breaks (DSBs) (Jazayeri et al., 2006).

Persistent ssDNA is bound by RPA, a heterotrimeric complex composed of RPA1, RPA2, and RPA3 (Zou et al., 2006). RPA bound to ssDNA recruits the ATR (ataxia telangiectasia and Rad3-related) kinase via its interacting partner ATRIP (Cortez et al., 2001; Zou and Elledge, 2003). The kinase activity of ATR is activated by two parallel pathways in vertebrates, via either TopBP1 or ETAA1 (Bass et al., 2016; Delacroix et al., 2007; Haahr et al., 2016; Lee et al., 2016). TopBP1 is recruited to ssDNA by the 9-1-1 complex (composed of Rad9-Rad1-Hus1). The protein RHINO binds both the 9-1-1 complex and TopBP1 and is required for optimal ATR activation (Cotta-Ramusino et al., 2011). ETAA1 binds directly to RPA and stimulates ATR kinase activity, both in response to replication stress as well as during unperturbed S phase (Michelena et al., 2019; Saldivar et al., 2018). Recently, a cleavage-dependent mechanism of CHK1 activation during unperturbed S phase was proposed (Halder et al., 2019). The metalloprotease SPRTN cleaves the inhibitory N-terminus from CHK1 to release kinase active C-terminal products. This mechanism appears to be crucial for replication progression, when long stretches of ssDNA are absent. Intriguingly, ATR can also be activated by mechanical stress at the nuclear envelope, independently of RPA (Kumar et al., 2014).

Once activated, the ATR kinase orchestrates the replication stress response by phosphorylating substrate proteins. Chief among these is its downstream effector kinase CHK1 (checkpoint kinase 1). ATR-CHK1 signalling promotes replication fork stability and restart after stalling, suppresses origin firing, and arrests the cell cycle (Fig. 2) (Saldivar et al., 2017). For example, in response to DNA damage, CHK1 phosphorylates the Cdc25 phosphatases, including Cdc25A (for G1-S transition) and Cdc25C (for mitotic entry), which promotes their degradation by

the proteasome (Mailand et al., 2000; Sørensen et al., 2003). These phosphatases remove inhibitory phosphate groups from CDK1 and CDK2 which form cyclin-CDK complexes required for cell cycle progression. Thus, the ATR-CHK1-mediated inhibition of Cdc25 phosphatase family members suppresses CDK activation and cell cycle progression (Sanchez et al., 1997).

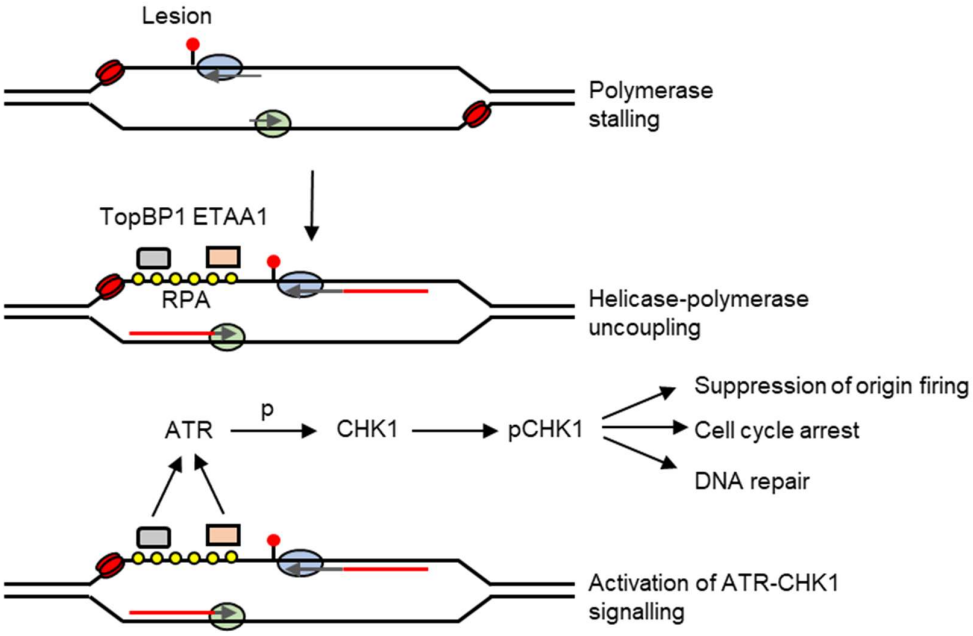



**Figure 2. A simplified model of ATR activation upon replication stress.**

DNA lesions can stall the replicative DNA polymerases while the CMG helicase continues to unwind the DNA duplex. This results in the formation of ssDNA, which is rapidly bound by RPA. The 9-1-1 complex and RHINO recruit TopBP1, while ETAA1 binds directly to RPA. TopBP1 and ETAA1 can independently stimulate ATR kinase activity. ATR phosphorylates CHK1 which phosphorylates hundreds of proteins to control cell cycle progression, origin firing, and DNA repair.

## 1.2 Mechanisms of DNA repair

A major source of DNA replication stress are DNA lesions. DNA lesions can therefore trigger cell cycle arrest and even cell death. In order to cope with DNA lesions, cell have evolved numerous DNA repair and DNA damage tolerance pathways.

### 1.2.1 Base excision repair (BER)

DNA bases can become damaged by oxidation, alkylation, deamination, and spontaneous hydrolysis. Damaged bases can be excised by DNA glycoylases (Lindahl, 1979). Human genomes encode at least 11 DNA glycosylases, each of which acts on a specific type of damaged base. For example, the spontaneous oxidation of guanine generates 8-oxoguanine (8-oxoG), which is removed by the 8-oxogG DNA glycosylase, OGG1 (Bruner et al., 2000). Base excision generates an apurinic/apyrimidinic (AP) site (Fig. 3). An AP site is then cleaved by an AP endonuclease, creating a DNA end that is processed by AP lyase. The resulting DNA gap may only be single nucleotide (short patch) or between 2 and 10 nucleotides (long patch), for reasons which are not entirely clear but appear to be influenced by the type of damaged base and cell cycle stage (Fortini and Dogliotti, 2007). Nucleotides are re-incorporated at short patches by DNA polymerase β and resealed by DNA ligase III (Cappelli et al., 1997). Long patches are synthesised by DNA polymerases β and δ or ε, and resealed by DNA ligase I (Dianov and Hübscher, 2013).

### 1.2.2 Nucleotide excision repair (NER)

NER differs from BER in that it recognises structural distortions in the DNA helix rather than a specific DNA lesion (i.e. damaged bases) (Fig. 3) (Nouspikel, 2009).

NER is perhaps best known for its role in repairing DNA damage caused by UV light, which can induce the covalent dimerization of neighbouring pyrimidines, leading to the formation of cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4 PPs). Helix-distorting DNA lesions can also be induced by endogenous products of cellular metabolism, for instance malendialdehyde produced during lipid peroxidation (pyrimidopurinone adducts) and ROS (cyclopurine dimers) (Melis et al., 2013).

Two distinct NER pathways exist: global genome-NER (GG-NER) and transcription-coupled NER (TC-NER). GG-NER is initiated by the XPC-RAD23B heterodimer which binds the DNA strand opposite the lesion (Sugasawa et al., 1998). DNA lesions that induce relatively little structural distortion, such as CPDs, are recognised by the DDB complex which, in turn, recruits XPC-RAD23B. XPC-RAD23B recruits the transcription factor II H (TFIIH) complex, which is composed of ten proteins, including two helicases: XPB and XPD (Compe and Egly, 2012). These helicases unwind the DNA duplex surrounding the DNA lesion, generating a ‘denaturation bubble’ that acts as a DNA platform for XPA to bind. Next, two endonucleases are recruited: XPF-ERCC1, by binding to XPA, and XPG, by binding to TFIIH. XPF and XPG cleave the DNA 5’ and 3’ of the DNA lesion, respectively. Filling of the resulting DNA gap is carried out by DNA polymerase δ, ε and/or κ and the nick is sealed by DNA ligase I or III. TC-NER differs from GG-NER only in its initiation step. TC-NER is initiated by the stalling of RNA polymerase II (RNAPII) during transcription of a damaged DNA strand. Stalled RNAPII recruits CSB, which in turn bind chromatin remodelling factors, such as p300, to promote formation of the denaturation bubble, independently of XPC (Fousteri et al., 2006).

Mutations in genes encoding NER components cause Xeroderma Pigmentosum, a family of diseases characterised by extreme sensitivity to UV light and a greatly elevated risk of developing skin cancer (Cleaver, 2005).

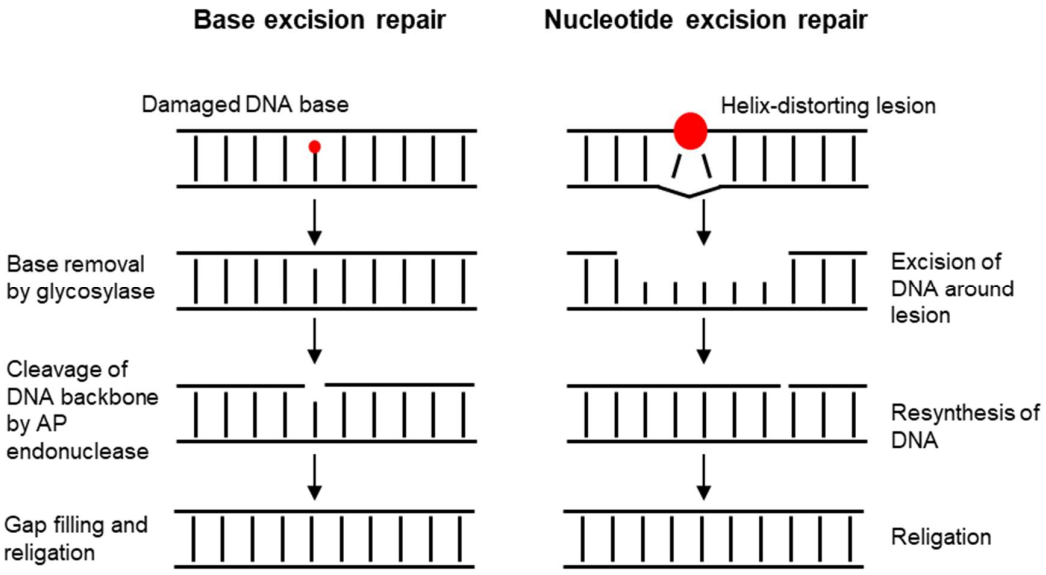



**Figure 3. General principles of excision repair.**

Schematic diagram of base excision repair (left) and nucleotide excision repair (right). Excision repair pathways all involve DNA incision, repair synthesis and religation.

### 1.2.3 Double strand break (DSB) repair

A single DSB can induce cell death, making it one of the most deleterious DNA lesions (Bennett et al., 1993). DSBs can be induced by exogenous sources such as ionising radiation (IR). The most common endogenous source of DSBs is when a replication fork encounters an unrepaired DNA lesion, such as an ssDNA nick (Pfeiffer, 2000).

DSB repair is regulated by a signalling cascade orchestrated by the protein kinase ATM (ataxia telangiectasia mutated). ATM is activated by binding to the NBS1 subunit of the MRN (MRE11-Rad50-NBS1) complex at break sites (Lee and Paull, 2004). Activated ATM phosphorylates hundreds of substrates. Amongst these are p53 and CHK2, both of which can trigger cell cycle arrest, and the histone variant H2AX, which initiates the recruitment of DSB repair factors (Blackford and Jackson, 2017).

H2AX is phosphorylated (i.e. γH2AX; phosphorylated on serine 139) on the DNA flanking the break site and initiates the sequential recruitment of MDC1, RNF8, and RNF168 (Doil et al., 2009; Kolas et al., 2007; Mailand et al., 2007; Rogakou et al., 1998; Stucki et al., 2005). H2AX can also be phosphorylated by the related protein kinases, ATR and DNA-PK (Blackford and Jackson, 2017).

RNF168 is required for the recruitment of BRCA1 and 53BP1 to DSB sites. The competing activities of these two factors guide DSB repair via one of two major pathways: BRCA1 promotes homologous recombination (HR) repair and 53BP1 promotes repair via Non-Homologous End-Joining (NHEJ). DNA repair pathway choice depends on various parameters, including cell cycle stage and the

chromatin structure at break sites (Escribano-Díaz et al., 2013; Goodarzi et al., 2009; Shibata et al., 2011).

#### 1.2.3.1 HR

DSB repair by HR uses a sister chromatid as a homology donor. Accordingly, HR is only active in the S and G2 phases of the cell cycle, when sister chromatids are available. HR is initiated by the binding of the MRN complex to break sites. The MRN complex collaborates with CtIP to generate a 3' ssDNA overhang at the break (Cannavo and Cejka, 2014). The 5'-3' exonuclease activities of EXO1 and DNA2 can extend the length of the ssDNA overhang by up to a few hundred nucleotides (Zhu et al., 2008). ssDNA overhangs are rapidly bound by RPA and thereby shielded from nucleolytic degradation. Rad51 is then loaded onto ssDNA by BRCA2-PALB2, displacing RPA in the process (Zhao et al., 2015). Rad51 can search the genome for sequence homology and catalyse strand invasion. Synthesis from the 3' end of the broken DNA is initiated by replication factor C (RFC), PCNA, and DNA polymerase δ (Maloisel et al., 2008; McVey et al., 2016).

BRCA1 functions in two steps to promote HR: first, it promotes CtIP-mediated DSB end resection; second, it recruits PALB2-BRCA2, to facilitate Rad51 loading (Escribano-Díaz et al., 2013; Zhang et al., 2009). Recent evidence indicates that BRCA1 also directly binds Rad51 to promote strand invasion (Zhao et al., 2017).

#### 1.2.3.2 NHEJ

In contrast to HR, NHEJ does not rely on sequence homology and, ultimately, results in the re-ligation of two broken DNA ends.

53BP1 loss rescues the HR defect in BRCA1-deficient cells by restoring end resection (Bouwman et al., 2010; Bunting et al., 2010). This led to a model being

proposed in which 53BP1 promotes NHEJ by suppressing end resection. 53BP1 has since been shown to cooperate with numerous binding partners (e.g. Rif and REV7) and downstream effectors (the Shieldin complex) to protect DSB ends from resection (Chapman et al., 2013; Dev et al., 2018; Ghezraoui et al., 2018; Gupta et al., 2018; Noordermeer et al., 2018; Xu et al., 2015). Thus, NHEJ only operates on minimally processed DNA ends. Consistent with this, BRCA1 is able to exclude 53BP1 from break sites in S/G2 phase (Chapman et al., 2012).

Classical NHEJ is initiated by the Ku70-Ku80 heterodimer which rapidly binds to exposed DNA ends. Ku70-80 recruit DNA-PKcs, resulting in the formation of the active DNA-PK holoenzyme (Gottlieb and Jackson, 1993). Auto-phosphorylated DNA-PK serves as a recruitment platform for downstream factors, such as XLF-XRCC4 and the nuclease Artemis (Ahnesorg et al., 2006). DNA-PK-dependent phosphorylation of Artemis stimulates the processing of DSB ends with long ssDNA overhangs in order to make them amenable to re-ligation (Goodarzi et al., 2006; Jeggo and Löbrich, 2005). XLF-XRCC4 form a filament structure that physically bridges the two broken DNA ends and PAXX stabilises this filament, allowing DNA ligase IV to ligate the two DNA ends (Ochi et al., 2015; Wang et al., 2018).

While HR is only active in S and G2 phases, NHEJ is active throughout the cell cycle. The rapidity with which the Ku70/Ku80 heterodimer is recruited to all DSB ends suggests that NHEJ makes the first attempt to repair DSBs (Beucher et al., 2009). This begs the question of how HR is initiated in the S and G2 phases. One possibility is that Ku is inhibited or removed from DSB ends to facilitate HR. Ku molecules may be displaced by the activity of the MRN complex, extracted by the

ATPase p97, or directly inhibited from binding DNA in S and G2 phase by CYREN (Arnoult et al., 2017; van den Boom et al., 2016; Shibata et al., 2014).

### 1.2.4 DNA damage tolerance (DDT)

Replicative DNA polymerases have small active sites which cannot accommodate damaged DNA bases. Therefore, during S phase, damaged bases and bulky lesions, can stall elongating DNA polymerases. A failure to restart stalled replication forks can lead to fork collapse, DSB formation, chromosomal rearrangements and, ultimately, cell death. To circumvent these deleterious outcomes, cells have evolved specialised DNA polymerases that can replicate damaged DNA templates by translesion synthesis (TLS) (Friedberg et al., 2002).

While DNA polymerases stall on damaged DNA templates, the replicative helicase continues to unwind the DNA helix, resulting in the formation of long stretches of ssDNA (Fig. 6) (Hedglin and Benkovic, 2015). ssDNA recruits the E3 ubiquitin ligase Rad18 and the ubiquitin-conjugating E2 enzyme, Rad6, which cooperate to attach a single ubiquitin molecule to PCNA. Mono-ubiquitinated PCNA is bound by the error-free TLS polymerase, Pol η (eta), or error-prone polymerases, such as Pol θ (theta), Pol ι (iota), Pol κ (kappa), REV1, and Pol ζ (zeta) (Pilzecker et al., 2019).

In mammalian cells, TLS occurs via a two-step process. First, insertion TLS DNA polymerases (e.g. Pol κ, Pol ι, Pol η, or REV1) introduce a nucleotide opposite the lesion. Second, the newly incorporated nucleotide is extended from its 3' terminus by extension TLS DNA polymerases, such as Pol ζ (Yang and Gao, 2018).

Pol η, is extremely efficient at replicating across CPDs and is therefore crucial for protecting cells against UV-induced genotoxicity (McCulloch et al., 2004). Indeed, mutations in the gene encoding Pol η confer a genetic predisposition to skin cancer (Masutani et al., 1999). However, the major TLS pathway is carried out by the error-prone polymerases REV1 and Pol ζ (Yamanaka et al., 2017). Due to their capacity to incorporate incorrectly base-paired nucleotides opposite template strands, error-prone polymerases are potentially mutagenic (McCulloch and Kunkel, 2008). Accordingly, the upregulation of error-prone polymerases is associated with poor prognosis in various types of cancer (Kawamura et al., 2004).

However, recent work elucidated a critical role for the error-prone polymerase, Pol θ, in protecting against squamous cell carcinoma, the second most common skin cancer caused by UV light (Yoon et al., 2019). Pol θ was found to be remarkably efficient at extending DNA templates across 6-4 PPs. Unexpectedly, Pol θ-mediated TLS, suppressed skin cancer development. Moreover, Pol θ-deficiency exacerbated the development of skin cancers in Pol η-deficient mice. The factors which determine whether error-free or error-prone TLS prevails are not well understood, however, both branches of TLS are clearly critical for suppressing tumorigenesis (Pilzecker et al., 2019).

A second error-free DDT pathway exists, known as template switching, in which a sister chromatid is used as a homology donor (Hishida et al., 2009). Template switching is initiated by Ubc13-Mms2 and Rad5, a ubiquitin ligase complex that extends the ubiquitin molecule on monoubiquitinated PCNA into K63-linked polyubiquitin chain (Hoege et al., 2002). In this way, TLS and template switching are temporally regulated such that TLS is the preferred DDT pathway.

Polyubiquitinated PCNA recruits many of the factors that mediate strand invasion in HR, including Rad51, to promote template switching (Vanoli et al., 2010).

### 1.2.5 Inter-strand crosslink (ICL) repair

Inter-strand crosslinking refers to the covalent crosslinking of the two strands of a DNA helix. ICLs are highly toxic because they prevent DNA unwinding during DNA replication and thereby impede approaching DNA replication forks (Clauson et al., 2013).

ICLs can be induced by aldehydes produced during the metabolism of ethanol and lipids (Crossan et al., 2012; Niedernhofer et al., 2003). ICL-inducing agents, such as platinum compounds and mitomycin C are also commonly used in chemotherapy (Takahara et al., 1995).

Two distinct ICL repair pathways exist in metazoans, both of which operate in a replication-dependent manner (Fig. 4). The first of these to be characterised, the Fanconi Anaemia (FA) pathway, involves the concerted action of numerous proteins (Moldovan and D'Andrea, 2009). The FA pathway is thought to be initiated by the convergence of two replication forks at an ICL (Zhang and Walter, 2014; Zhang et al., 2015). Fork convergence triggers polyubiquitination of MCM7, by the ubiquitin ligase TRAIP (Amunugama et al., 2018; Wu et al., 2019). The ATPase p97 unloads polyubiquitinated MCM7, resulting in the disassembly of the entire CMG helicase. ICLs are then recognised by the FANCM/FAAP24 complex or alternatively by UHRF1 and UHRF2 (Ciccia et al., 2007; Liang et al., 2015; Motnenko et al., 2018). FANCM/FAAP24 recruit the 14 protein subunit FA core complex. The core complex includes the E3 ubiquitin ligase, FANCL, and the E2 ubiquitin-conjugating enzyme, UBE2T, which monoubiquitinate FANCI-FANCD2.

UHRF1 and UHRF2 can bind directly to ICLs and appear to retain FANCD2 at ICLs to enable its mono-ubiquitination. Mono-ubiquitinated FANCD2-FANCI, in turn, is essential to recruit the endonucleases XPF-ERCC1 and SLX4-SLX1 which introduce incisions either side of the ICL. The resulting DSB can then be repaired by HR (Fig. 4) (Zhang and Walter, 2014).

A second, much simpler, ICL pathway was recently shown to act on a subset of ICLs. The NEIL3 glycosylase 'unhooks' one of the two glycosyl bonds (between a DNA base and sugar-phosphate backbone) that comprise an ICL (Fig. 4). This avoids the generation of a DSB, but results in the formation of an abasic site that is bypassed by the TLS polymerases, REV1-Pol ζ (Budzowska et al., 2016).

ICL repair pathway choice appears to be governed by the length of the ubiquitin chains conjugated to MCM7 by TRAIP (Fig. 4). Short ubiquitin chains support NEIL3-mediated ICL repair. Certain ICLs are refractory to N-glycosyl bond unhooking by NEIL3, such as those induced by platinum compounds. This delayed repair presumably allows longer ubiquitin chains to be synthesised on MCM7, which are required for p97-dependent CMG unloading, and activation of the FA pathway (Wu et al., 2019).

The toxicity of endogenous aldehydes and consequent ICL formation is demonstrated in mice that lack ALDH2, the enzyme that detoxifies acetylaldehyde, and FANCD2 (Garaycoechea et al., 2012; Langevin et al., 2011). These mice exhibit developmental abnormalities, anaemia, and mostly die of acute lymphoblastic leukemia within 6 months of birth.

The FA pathway takes its name from the disease that arises when genes that encode components of this pathway are mutated. So far, mutations in the genes of 19 members of the FA pathway have been found to cause Fanconi Anaemia. Almost all FA patients develop early-onset bone marrow failure and are predisposed to developing acute myeloid leukaemia and/or solid tumours (Dong et al., 2015).

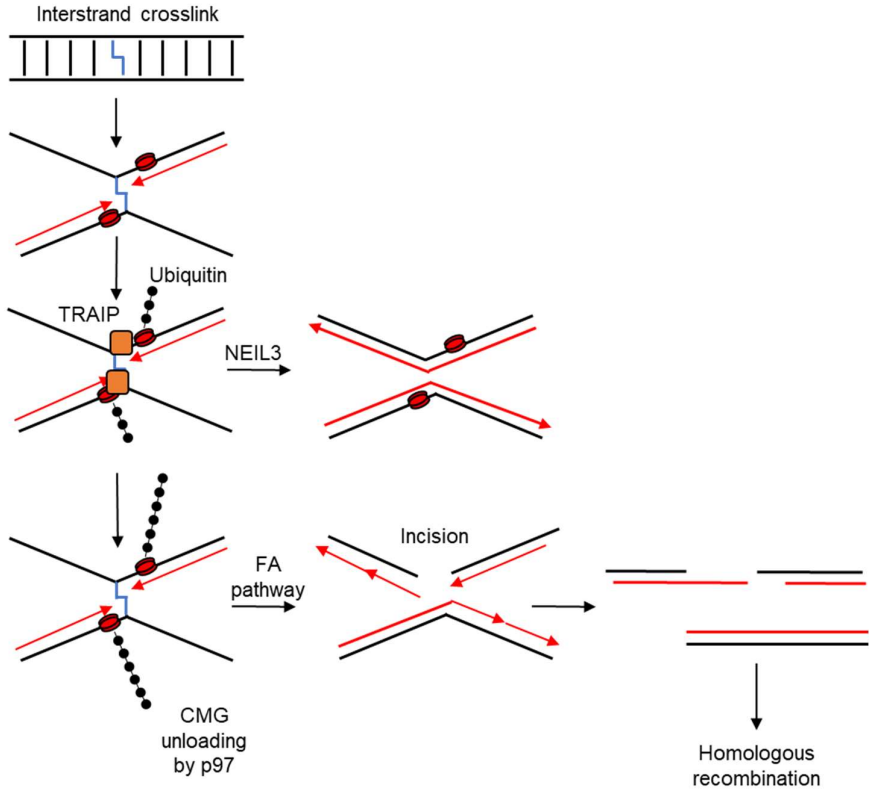



**Figure 4. Pathways involved in ICL repair.**

Schematic diagrams of NEIL3- and Fanconi anaemia (FA) protein-dependent ICL repair pathways. Both pathways are initiated by fork convergence followed by MCM7 ubiquitination by TRAIP. Short ubiquitin chains recruit the NEIL3 glycosylase which 'unhooks' the ICL by hydrolysing the glycosyl bond between DNA base and sugar-phosphate backbone on one side of the ICL. Larger ICLs cannot be repaired by NEIL3, allowing longer ubiquitin chains to be synthesised on MCM7, resulting in CMG unloading by p97. FA pathway proteins then incises the phosphodiester DNA backbone either side of the ICL and the resulting DSB can be repaired by homologous recombination. Adapted from Wu et al., 2019.

### 1.2.6 DNA-protein crosslink repair

DNA-protein crosslinks (DPC) consist of a protein covalently bound to DNA and are broadly classified into two categories: enzymatic and non-enzymatic DPCs (Fig. 5) (Fielden et al., 2018; Ide et al., 2018; Vaz et al., 2017).

Enzymatic DPCs arise due to the permanent trapping of DNA-metabolising proteins that form a, usually transient, covalent intermediate with DNA during their catalytic cycles. Non-enzymatic DPCs form as a result of chemical reactions between DNA and proteins (Vaz et al., 2017). DPCs can disrupt essential processes such as DNA replication, transcription, and mitosis. As they can vary substantially in size and structure, DPCs pose unique obstacles to cells.

#### 1.2.6.1 Enzymatic DPCs

Well-known examples of enzymatic DPCs are topoisomerase 1 and 2 cleavage complexes (TOP1- and TOP2-ccs) (Fig. 5). Topoisomerases are involved in the resolution of DNA topological stress that accumulates during replication, transcription, and mitosis. During their catalytic cycles, topoisomerases introduce DNA strand breaks. TOP1 introduces a DNA break in only one DNA strand, whereas TOP2 is a dimer and cleaves both DNA strands. A tyrosine in the topoisomerase active site covalently binds to a phosphate group on the end of the broken DNA strand(s). These covalent TOP-DNA intermediates are referred to as topoisomerase cleavage complexes (TOP-cc). TOP1 relaxes DNA topological stress by swivelling the broken DNA strand around the unbroken DNA strand, whereas TOP2 enables a double-stranded DNA molecule to pass between a double-stranded DNA break (Pommier et al., 2010).

TOP-ccs are usually transient intermediates that are resolved when the two ends of broken DNA are re-ligated. However, if a topoisomerase cleaves near DNA alterations such as oxidative lesions, base modifications, or DNA breaks, then the re-ligation step and subsequent release of the topoisomerase is disrupted, and a TOP-cc becomes permanent (Pommier et al., 2006). Another example of an enzymatic DPC is DNA polymerase β, which can become covalently trapped to specific oxidative DNA lesions during BER (Demott et al., 2002).

Two tyrosyl-DNA phosphodiesterases, TDP1 and TDP2, exhibit remarkable specificity for TOP1- and TOP2-ccs, respectively (Ledesma et al., 2009; Yang et al., 1996). TDP1 catalyses the hydrolysis of phosphodiester bonds that link proteins to 3' phosphate groups (i.e. TOP1ccs), whereas TDP2 preferentially acts on proteins linked to 5' phosphate groups (i.e. TOP2ccs). In addition to TDP1 and TDP2, nucleases acting on the DNA adjacent to the TOP-cc also contribute to their removal. The ability of the endonuclease, XPF-ERCC1, to cleave 3' single-stranded DNA flaps appears to contribute to TOP1cc repair, albeit to only a modest extent (Zhang et al., 2011). Similarly, MRE11 and the endonuclease activity of CtIP contribute to the resection of TOP2cc-adducted DSB ends (Aparicio et al., 2016; Hoa et al., 2016).

The cytotoxicity of TOP1- and TOP2-ccs is best illustrated by the various neurodegenerative diseases that arise when their repair is defective. Mutations in *TDP1* underlie spinocerebellar ataxia with axonal neuropathy (SCAN1), a severe neurodegenerative disorder (Takashima et al., 2002). Interestingly, the only known mutation in *TDP1* associated with SCAN1 (substitution of histidine at position 493 to arginine) not only diminishes TDP1's enzymatic activity by approximately 25-fold, but also delays the resolution of a covalent TDP1-DNA

reaction intermediate (Interthal et al., 2005a). Defects in TOP1cc repair are also thought to contribute to the neuropathology of ataxia telangiectasia patients (Alagoz et al., 2013; Katyal et al., 2014). TDP2 mutations are associated with a neurological disorder, known as spinocerebellar ataxia, autosomal recessive 23 (SCAR23), characterised by ataxia, seizures, and learning disabilities (Gómez-Herreros et al., 2014; Zagnoli-Vieira et al., 2018).

#### 1.2.6.2 Non-enzymatic DPCs

In addition to enzymatic DPCs, it is possible for any protein in the vicinity of DNA to become permanently crosslinked to DNA (Fig. 5). One way in which such non-enzymatic DPCs can be induced is by chemical by-products of cellular metabolism. Endogenous aldehydes are especially potent DNA-protein crosslinking agents and are produced by a variety of chemical reactions in a cell (Tretyakova et al., 2015). For example, formaldehyde is produced during histone demethylation, a process which regulates gene expression and chromatin structure. Aldehydes may also be formed during lipid peroxidation (e.g. malondialdehyde) or during the oxidation of ethanol by alcohol dehydrogenase (e.g. acetylaldehyde) (Crossan et al., 2012; Niedernhofer et al., 2003).

Aldehydes induce DPCs by reacting with free amino and imino groups on DNA and/or the side chains of amino acids. These, in turn, react with other free amino groups. Given that the concentration of formaldehyde is approximately 100 μM in human plasma and can reach up to 400 μM in some tissues, it is not surprising that DPCs are common (Heck and Casanova, 2004; Tong et al., 2013). Reactive oxygen and nitrogen species (ROS and RNS) can also induce DPCs. By reacting with DNA and/or amino acids, they generate reactive radicals or electrophilic lesions that can, in turn, react with proteins and DNA. Abasic sites, which can

form spontaneously or as a result of BER, also efficiently crosslink histones to DNA (Sczepanski et al., 2010).

Exogenous DPC-inducing agents include IR, UV light, and chemotherapeutic agents, including platinum compounds. Environmental aldehydes are also widespread, for example, in cosmetics (Chválová et al., 2007; Fielden et al., 2018; Vaz et al., 2017). Due to the plethora of ways in which DPCs can arise, mammalian cells are estimated to accumulate approximately 6000 DPCs during exponential growth (Oleinick et al., 1987).

Non-enzymatic DPCs are extremely diverse in terms of the size and structure of the crosslinked protein as well as in the chemistry of the crosslink itself. This implies that their repair cannot rely on factors with a high degree of DNA/protein structure-specific activity. There is evidence that NER and HR counteract formaldehyde-induced cytotoxicity from various studies in bacteria, yeast, and humans (Fig. 5). In bacteria, NER was found to be capable of resolving small DPCs (Nakano et al., 2007). Similarly, XPF-deficient human cells are sensitive to formaldehyde (Kumari et al., 2012). However, *in vitro* reconstitution of DPC repair using mammalian NER components suggested that NER cannot excise DPCs more than 16 kDa in size (Reardon and Sancar, 2006). Genome wide screens in yeast also identified NER as being important in response to acute treatment with a high dose of formaldehyde, whereas HR components were important for survival upon chronic exposure to a low dose of formaldehyde (de Graaf et al., 2009). However, because chronic formaldehyde treatment can induce DSBs as well as other types of DNA lesions (such as ICLs) which rely on HR for their repair, it is difficult to interpret the contribution of HR to DPC repair. Nevertheless, the prevailing consensus was that NER acts on smaller DPCs in the G1 phase of the

cell cycle, whereas HR acts on larger DPCs in the S phase (Duxin and Walter, 2015; Nakano et al., 2007).

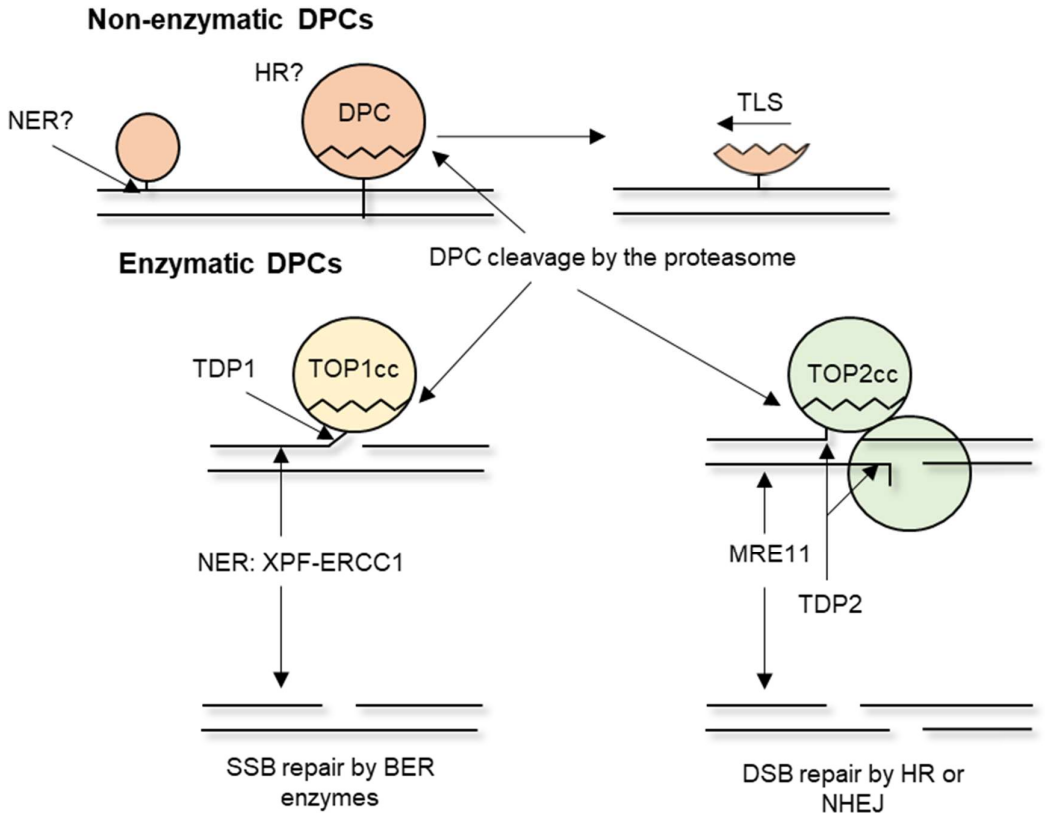



**Figure 5. DPC repair pathways.**

Schematic diagrams of non-enzymatic and enzymatic DPCs and the factors/pathways involved in their repair. Non-enzymatic DPCs can be degraded by the proteasome and the remnant DNA-bound peptide can be bypassed by translesion synthesis (TLS). Smaller DPCs may be repaired by nucleotide excision repair (NER) and larger DPCs may be repaired by homologous recombination (HR) if a DSB forms near a DPC. Enzymatic DPCs, such as TOP1- and TOP2-ccs, can also be degraded by the proteasome after which the remnant DNA-bound peptides are excised by the phosophodiesterases, TDP1 and TDP2. Nucleases, such as MRE11 and XPF-ERCC1, can also resolve DPCs by cleaving DNA adjacent to the DPC. Adapted from Fielden et al., 2018.

#### 1.2.6.3 DPC proteolysis repair

*In vitro,* native, recombinant TOP1- and TOP2-ccs are resistant to hydrolysis by TDP1 and TDP2, respectively (Gao et al., 2014; Interthal et al., 2005b; Reardon et al., 2006). For example, smaller TOP1 peptide fragments of 53 and 108 amino acids are efficiently resolved by TDP1, whereas 333 amino acid TOP1 peptide fragments are resistant to cleavage (Interthal and Champoux, 2011). This led to the conclusion that the phosphotyrosyl substrate bonds of TDP1 and TDP2 are shielded within the large TOP-cc structures. Heat denaturation or trypsin digestion of TOP-ccs enable TDP1 and TDP2 activity *in vitro* (Gao et al., 2014; Interthal et al., 2005b; Reardon et al., 2006). This raised the question of how the large TOP1 (90 kDa) and TOP2 (174 kDa) proteins are processed *in vivo*, to facilitate the access of TDP1 and TDP2 to TOP-DNA phosphotyrosyl bonds. Numerous studies have reported that the proteasome is able to counteract the cytotoxic effects of drugs that stabilise TOP1- and TOP2-ccs (Desai et al., 1997; Lin et al., 2008; Mao et al., 2001). This suggested that the proteasome digests the protein component of TOP1- and TOP2-ccs once they become trapped on DNA (Fig. 5). However, there is no direct evidence of TOP-cc cleavage by the proteasome.

Evidence of the existence of a specialised DPC repair pathway first emerged from yeast studies. Cells lacking the metalloprotease Wss1 and TDP1 exhibited a severe growth defect and sensitivity to the TOP1 poison camptothecin (Stingele et al., 2014). Wss1 is a member of the WLM family of metalloproteases. A feature of this family of proteases is that they possess a HEXXH motif in their active centres, where the glutamic acid constitutes the protease active site and two histidines bind metal ions (Stingele et al., 2015; Vaz et al., 2017). Point mutation of Wss1's predicted protease active centre completely abolished its ability to

complement the growth defect of Wss1/TDP1-deleted cells. This suggested that Wss1 was indeed a protease that might be capable for directly cleaving the TOP1 protein. Wss1 cleaved TOP1 as well as a variety of other DNA-binding proteins *in vitro*, raising the possibility that Wss1's proteolytic activity may extend to other substrates. Indeed, Wss1 was subsequently found to confer cellular resistance to formaldehyde.

Research into the repair of a plasmid-borne DPC in *Xenopus laevis* egg extract, published in the same issue of *Cell*, suggested that a protease-dependent mechanism of DPC repair also exists in metazoans (Duxin et al., 2014). The DPC was proteolytically digested in a DNA replication-coupled manner, yielding a small peptide-DNA remnant that could be bypassed by the TLS polymerases, REV1 and Pol ζ. Importantly, as it did not involve the generation of a DSB intermediate, DPC proteolysis was suggested to be the preferred mechanism of DPC repair in S phase, not HR. While the protease in question was not identified, the proteasome was ruled out as a possible candidate. It was speculated that the protease might be DVC1/SPRTN, a putative metalloprotease which shares the HEXXH protease active centre consensus sequence with Wss1 (Duxin and Walter, 2015).

#### 1.2.6.3.1 DVC1/SPRTN

DVC1/SPRTN (referred to as SPRTN hereafter) was first identified as a regulator of translesion synthesis by several groups (Centore et al., 2012; Davis et al., 2012; Ghosal et al., 2012; Juhasz et al., 2012; Machida et al., 2012; Mosbech et al., 2012). SPRTN possesses several protein domains: a MIU (motif-interacting with ubiquitin), a metalloprotease domain (SprT family), a PCNA-interacting motif

(PIP box), a zinc finger ubiquitin-binding domain (UBZ4), and a p97 interaction motif (SHP box) (Fig. 7A).

It was found that SPRTN interacts with PCNA and TLS polymerases and that SPRTN deficiency sensitised cells to UV light, as well as other replication stress-inducing agents. SPRTN-deficiency did not confer sensitivity to ionising radiation suggesting its function is to protect cells from replication stress and not DSBs. Moreover, SPRTN is only expressed in the S and G2 phases of the cell cycle and is degraded by APC/$C^{Cdh1}$ in G1 (Mosbech et al., 2012).

All these features are consistent with an important role for SPRTN in protecting cells from UV-induced replication stress. However, much of the evidence about its precise function in this process is conflicting. Specifically, three groups concluded that it was recruited to UV-induced DNA lesions by Rad18-dependent PCNA mono-ubiquitination (Centore et al., 2012; Ghosal et al., 2012; Juhasz et al., 2012), whereas two others reported that it was not (Davis et al., 2012; Mosbech et al., 2012). Moreover, two groups suggested that the function of SPRTN is to recruit the ATPase p97 to extract translesion DNA polymerases (Davis et al., 2012; Mosbech et al., 2012) (Fig. 6). In this way, SPRTN was proposed to facilitate the switch from translesion polymerases to replicative DNA polymerases once a UV-induced DNA lesion has been bypassed. However, another group proposed that SPRTN promotes the removal of Pol δ to allow the translation DNA polymerase, Pol η, to be recruited (Ghosal et al., 2012). More recent evidence suggests that SPRTN has no impact on p97 recruitment to UV-induced DNA CPDs (Puumalainen et al., 2014).

Further support for a role of the SPRTN-p97 complex in preserving genome stability came from the finding that it extracts ubiquitinated FANCD2/FANCI from chromatin (Gibbs-Seymour et al., 2015). The removal of FANCD2/FANCI from chromatin may represent a mechanism to suppress the recruitment of nucleases to DNA once repair is complete.

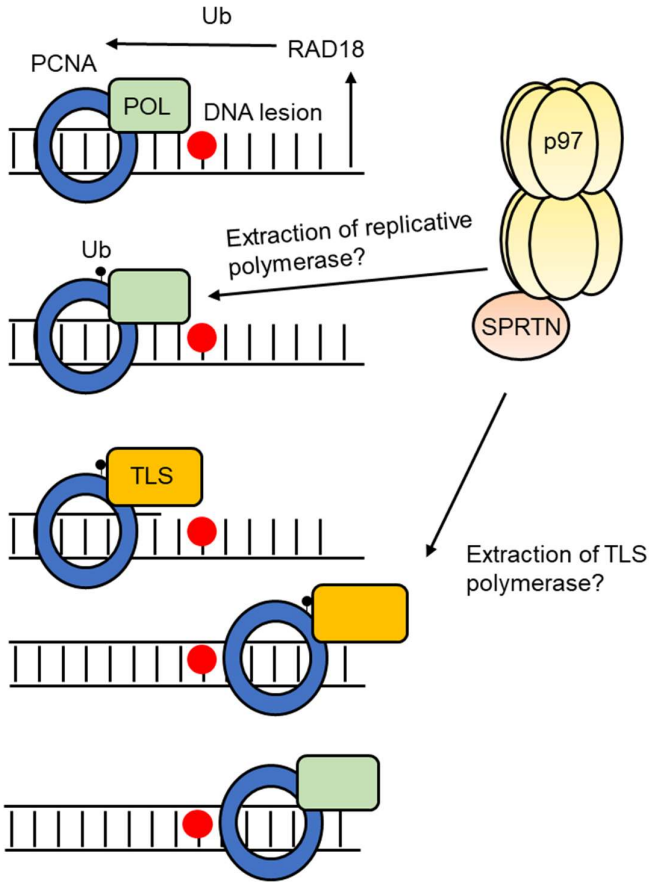



**Figure 6. Schematic diagram of DNA damage tolerance by translesion synthesis (TLS).**

Left: damaged DNA templates can stall DNA polymerases, leading to helicase-polymerase uncoupling and ssDNA formation. ssDNA can trigger activation of the Rad6-Rad18 complex which catalyses the monoubiquitination of PCNA on lysine 164. This triggers the switch from replicative polymerases (POL; green) to TLS polymerases (TLS; orange), which can bypass damaged DNA templates. Once a lesion is bypassed, TLS polymerases are replaced by replicative polymerases. Right: reported roles of the SPRTN-p97 complex in TLS.

Despite the presence of a putative metalloprotease active centre, no evidence that SPRTN possessed proteolytic activity was forthcoming. This is presumably because these groups focused on its potential function as a SUMO isopeptidase based on the HEXXH motif it shares with Wss1 (Mosbech et al., 2012). Wss1 was initially proposed to possess isopeptidase activity, whereby it cleaves the isopeptide bond between the terminal diglycine residues on SUMO and the lysine residue to which it is covalently attached on a substrate (Mullen et al., 2010). However, subsequent work from the Jentsch laboratory was unable to reproduce this activity (Stingele et al., 2014). In addition, some of their observations directly conflicted with those from Mullen et al. For example, Mullen et al reported that Wss1's proteolytic activity was not inhibited by the metal ion chelator, EDTA, whereas Stingele et al found that it was. It would be surprising for Wss1 to be proteolytically active in the presence of EDTA given that it is a predicted zinc metalloprotease and most known metalloproteases need to bind metal ions to be active (Su and Hochstrasser, 2010). Moreover, Wss1 binds DNA and its proteolytic activity depends on the presence of DNA (Stingele et al., 2014). It is not yet known whether SPRTN binds DNA, however, it would be important to address this when assessing whether it possesses protease activity.

Despite conflicting reports over the precise molecular function of SPRTN, its central role in maintaining genome stability was supported by the discovery that monogenic mutations in *SPRTN* cause premature ageing and early-onset hepatocellular carcinoma (Lessel et al., 2014). Three patients from two unrelated families presented with progeroid features. Specifically, all three patients displayed low body weight, micrognathia (undersized jaw), and subsequently developed hepatocellular carcinomas in their teenage years. Cora Aalfs and

Marielle Ruijs described the clinical features of the first patient, the son of consanguineous parents from Morocco, in 2003 (Ruijs et al., 2003). This patient was subsequently found to have biallelic *SPRTN* mutations that resulted in the expression of a C-terminally truncated protein (1-249; Fig. 7B) (Lessel et al., 2014). In two related Australian patients, *SPRTN* was the only gene with rare, biallelic mutations. In both boys, one *SPRTN* allele encodes a protein variant with a tyrosine(Y)-to-cysteine(C) substitution at position 117, located in SPRTN's metalloprotease domain, a few amino acids downstream of its putative active site (E112). The other allele expresses a SPRTN variant truncated at amino acid 246 (Fig. 7B). These syndromes are collectively referred to as Ruijs-Aalfs syndrome (RJALS). An interesting feature of these patients is that they do not display overt signs of sensitivity to UV light which would be expected of SPRTN deficiency, given its reported role in the response to UV-induced replication stress (Davor Lessel, personal communication).

Immunohistochemical staining of the liver tumours from RJALS patients revealed signs of genomic instability, visualised by 53BP1 and γH2AX foci formation (Lessel et al., 2014). Patient-derived fibroblasts exhibited chromosomal rearrangements, proliferation defects, and DNA replication stress, including elevated levels of fork stalling. These phenotypes could be recapitulated by short interfering (si)RNA-mediated depletion of SPRTN in cultured human cell lines. Unlike wild-type SPRTN, ectopic expression of the RJALS patient variant, Y117C, could not rescue DNA replication fork progression in SPRTN-deficient cells. This suggested that the metalloprotease domain of SPRTN is crucial for its function (Lessel et al., 2014).

The progeroid phenotypes and liver tumours observed in RJALS patients were recapitulated in a *SPRTN* hypomorphic mouse model (Maskey et al., 2014). Complete *SPRTN* knockout was found to be embryonic lethal, indicating that SPRTN is essential for life. Genome wide CRISPR-Cas9 studies have since confirmed that complete *SPRTN* knockout is lethal in most of the human cancer cell lines tested (Behan et al., 2019). However, immortalised conditional *SPRTN* knockout mouse embryonic fibroblasts (MEF) could be generated by retroviral expression of SV40 T-antigen (Maskey et al., 2014). *SPRTN* knockout in these cells caused increased cell death and cell cycle defects. *SPRTN* knockout cells accumulated in late S phase and G2/M, suggesting they accumulated DNA damage during S phase. Complementation of *SPRTN* knockout MEFs with retrovirally-expressed wild-type SPRTN could fully reverse their cell cycle defects. SPRTN variants that were defective for p97-, ubiquitin-, and PCNA-binding could also fully rescue these cell cycle defects. Interestingly, a SPRTN variant with a point mutation in its putative metalloprotease active site (E112A) was completely unable to complement the growth defects observed in SPRTN knockout MEFs. These results strongly indicate that the essential function of SPRTN derives from its putative metalloprotease activity. This conclusion is further supported by the fact that RJALS patients are viable, despite expressing C-terminally truncated SPRTN variants that lack PIP, SHP, and UBZ domains (Lessel et al., 2014).

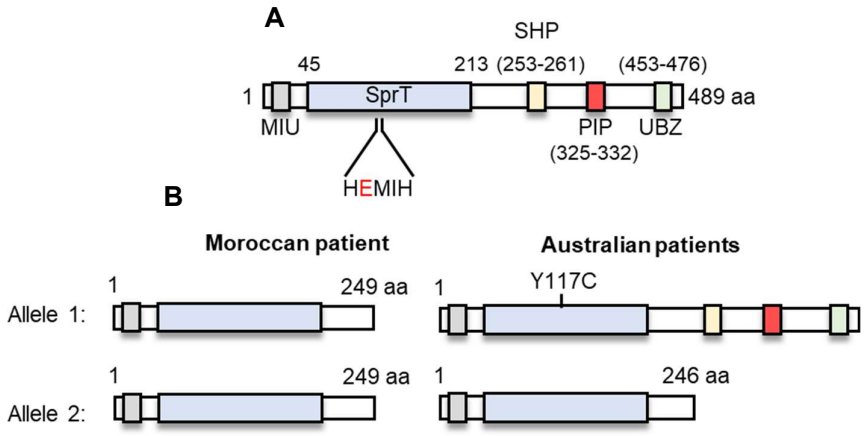



**Figure 7. SPRTN mutations cause early onset ageing and hepatocellular carcinoma.**

(A) Schematic of the SPRTN protein with the location of known protein domains indicated. SPRTN's putative metalloprotease HEXXH active site is shown with the predicted protease active site (E112) coloured in red. (B) Diagrams of the SPRTN variants in RJALS patients.

#### 1.2.6.3.2 VCP/p97

In addition to their protease active centres, another feature that Wss1 and SPRTN share is that they both have motifs that enable them to bind the ATPase p97 (also known as VCP in mammals and Cdc48 in yeast) (Balakirev et al., 2015; Ghosal et al., 2012; Mosbech et al., 2012; Stapf et al., 2011; Stingele et al., 2014).

p97 is an ATPase with a critical role in many cellular processes due to its ability to extract/segregate proteins from macromolecular structures or binding partners (Fig. 8A) (Bodnar and Rapoport, 2017a). The p97 protein is composed of an N-terminal domain, and two ATPase domains (D1 and D2). Six p97 molecules form a homohexameric, ring structure with a central pore. By using energy generated from ATP hydrolysis, p97 unfolds/remodels its substrates by threading them through its central pore (Blythe et al., 2017; Bodnar and Rapoport, 2017b). In this way, its substrates can be physically dissociated from membranes, chromatin, or binding partners. Typically, p97 substrates are modified with K48-linked polyubiquitin chains, which signal for their degradation by the proteasome. Given that the proteasome can only process unstructured substrates, the 'unfoldase' activity of p97 is essential for making proteins with vastly different structural properties amenable to proteasomal degradation (Beskow et al., 2009; Prakash et al., 2004).

p97 acts in many different cellular pathways and subcellular compartments. For example, in the endoplasmic reticulum (ER), p97 pulls misfolded substrates through the ER membrane to enable them to be degraded by the proteasome (Rabinovich et al., 2002; Ye et al., 2003, 2017). More recently, p97 was shown to promote the degradation of chromatin-associated proteins (Vaz et al., 2013). The first of these to be identified was the mitotic kinase, aurora B, which is extracted

from chromatin by p97 to enable nuclear envelope breakdown during mitosis (Meerang et al., 2011). To recognise its diverse array of substrates, p97 acts with adaptor proteins, also known as cofactors. p97 cofactors possess motifs that enable them to bind directly to either the N-terminal domain of p97 or to its C-terminal tail (Fig. 8) (Meyer et al., 2012). Most p97 cofactors also possess ubiquitin-binding domains that enable them to bind poly-ubiquitinated substrates. Distinct p97-cofactor sub-complexes exist in cells. There are thought to be three mutually exclusive cofactor sub-complexes which operate in different cellular processes (Buchberger et al., 2015). These 'core' cofactors are UBXD1, p47, and the Ufd1-Npl4 heterodimer. UBXD1 governs p97 function in endolysosomal sorting, p47 is essential for membrane fusion, and Ufd1-Npl4 cooperate with p97 in the ER as well as in the nucleus and on chromatin (Meyer et al., 2012).

Additional p97-cofactor complexes can be assembled in a hierarchical fashion, whereby the core cofactors can modulate the binding and function of additional p97 cofactors (Fig. 8B) (Buchberger et al., 2015). An example of this is the p97-Ufd1-Npl4-FAF1 complex (Hänzelmann et al., 2011). FAF1 is only able to directly bind p97-Ufd1-Npl4 complexes but not directly to free p97. This suggests that p97 binding by Ufd1-Npl4 induces a conformational change in the N-domain of p97 that enables it to bind FAF1 or other cofactors. The hierarchical binding model is thought to provide p97 core complexes with an additional layer of substrate-specificity control (Buchberger et al., 2015).

As alluded to, p97 plays an important role in numerous DNA repair pathways by processing substrates modified with K48 or K63-linked ubiquitin chains (Meerang et al., 2011).

Some of the substrates it is known to act on include Rad51, RPA, and yeast Rad52 (HR), MCM7 and FANCI-FANCD2 (FA pathway), DDB2, XPC, and Cdt1 (NER), DNA polymerases (TLS), and Ku70-Ku80 (NHEJ) (Amunugama et al., 2018; Bergink et al., 2013; van den Boom et al., 2016; Gibbs-Seymour et al., 2015; Inano et al., 2017; Mosbech et al., 2012; Puumalainen et al., 2014; Raman et al., 2011). p97 thereby orchestrates the timely assembly and disassembly of DNA repair and replication complexes to safeguard genome stability.

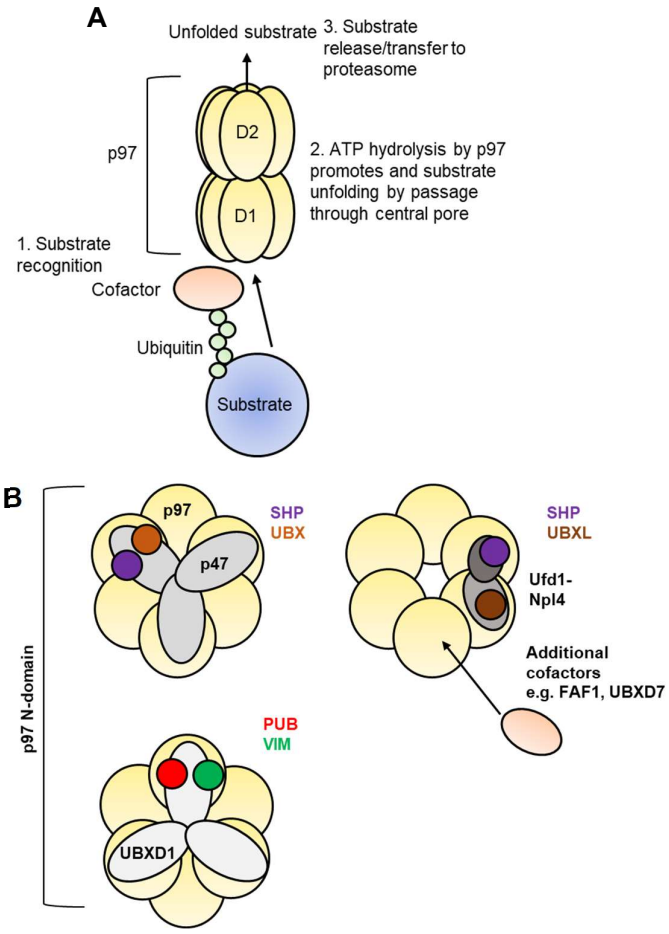



**Figure 8. Models of substrate processing and cofactor binding by p97.**

A) Schematic diagrams of substrate processing by p97. ATP hydrolysis by the D1 ATPase domain of p97 initiates substrate processing. The ATPase activity of p97's D2 ATPase domain drives substrate unfolding. Substrate unfolding occurs as they are translocated through the central pore of p97, starting at the N-domain. (B) Diagram of the p97 N-domain (yellow) and models of mutually exclusive cofactor binding (grey). p97-interaction motifs are highlighted (colours). UBXD1 can also bind C-terminus of p97 via its PUB domain (not depicted).

#### 1.2.6.4 p97 in DPC repair

Cdc48 has also been implicated in the repair of both TOP1- and TOP2-ccs. In the absence of TDP1, Ufd1 was shown to be important for counteracting basal TOP1cc accumulation in yeast (Nie et al., 2012). However, a mechanistic basis for how Cdc48-Ufd1 promotes TOP1cc repair was not described. Cdc48 also counteracts TOP2ccs by facilitating their proteasomal degradation (Wei et al., 2017).

The function of Wss1 in DPC repair depends on its ability to interact with Cdc48 (Balakirev et al., 2015; Stingele et al., 2014). However, the precise role of Cdc48 in this process is unclear. It was speculated that it might remodel substrates to facilitate their cleavage by Wss1 or possibly promote the clearance of remnant DNA-bound peptides generated by Wss1 (Stingele et al., 2014). The latter observation might be preferred given that Wss1 can efficiently cleave the TOP1 protein *in vitro* even in the absence of Cdc48. An alternative explanation was provided by a separate study on the role of Wss1 in TOP1cc repair. Wss1 was found to form a tertiary complex with p97 and its cofactor, Doa1, a protein linked to vesicle transport and autophagy (Balakirev et al., 2015; Ossareh-Nazari et al., 2010; Ren et al., 2008). Wss1 re-localised to autophagic vesicles in response to replication stress, raising the intriguing possibility that the p97 system promotes DPC clearance via autophagy (Balakirev et al., 2015). In this model, certain DPCs may be processed by autophagy because they are poorly recognised by proteasome receptor proteins, akin to protein-protein aggregates (Lu et al., 2017).

A recently solved crystal structure of SPRTN's active centre provided the mechanistic basis for a potential role for p97 in promoting SPRTN-mediated proteolysis (Li et al., 2019). SPRTN's catalytic core is shielded by its zinc-binding

domains within a narrow groove which only small, flexible peptides and not large, globular DPCs would be able to access. Therefore, p97 might be required to remodel/unfold substrates such that they could insert into SPRTN's active site.

#### 1.2.6.5 Post-translational modifications (PTM) in DPC repair

The functions of p97 and the proteasome are intimately linked to the PTM ubiquitin. Post-translational modification refers to the covalent attachment of a molecule or protein to another protein. PTMs can influence a protein's stability, function, localisation, and interactions. A variety of studies from yeast to humans have demonstrated that the PTMs, ubiquitin and the ubiquitin-like protein, SUMO, play prominent roles in DPC repair, connected to p97, the proteasome, and beyond.

##### 1.2.6.5.1 Ubiquitination

The covalent attachment of ubiquitin occurs through a three-step enzymatic cascade. E1 ubiquitin-activating enzymes initiate this cascade by adenylating ubiquitin molecules. The catalytic cysteine residue of the E1 reacts with adenylated ubiquitin, resulting in the formation of a thioester bond that links the E1 and ubiquitin. The E1 then transfers this thioester bond to an E2 ubiquitin-conjugating enzyme. The transfer of ubiquitin to a substrate is typically facilitated by an E3 ubiquitin ligase which either catalyse the conjugation of ubiquitin to a substrate or serve as substrate adaptors that bring an E2 in proximity to a substrate (Dye and Schulman, 2007). Ubiquitin chains can be assembled on substrates when ubiquitin molecules are linked through one of seven lysine residues or via their amino terminus (Met1) (Kulathu and Komander, 2012). Some E3 ligases also possess E4 ligase activity, whereby they can bind single ubiquitin

molecules or polyubiquitin chains generated by other E3 ligases and extend the length of these chains (Koegl et al., 1999).

#### 1.2.6.5.2 SUMOylation

Several ubiquitin-like proteins (UBL), such as SUMO, structurally resemble ubiquitin and can also be serve as covalent protein modifiers. Higher eukaryotes encode 3 types of SUMO - SUMO1, SUMO2, and SUMO3- whereas yeast possess only one SUMO family member, Smt3p. SUMO2 and SUMO3 share 93.6% sequence homology and are typically referred to as SUMO2/3. SUMO molecules are also ligated to target proteins via an enzymatic cascade, but by distinct ligases to ubiquitin. Only one E2 SUMO ligase, UBC9, is known to exist. In some cases, UBC9 can directly ligate SUMO molecules to substrates, without the need for an E3 SUMO ligase, albeit with lower efficiently (Karaca et al., 2011).

#### 1.2.6.5.3 Roles of ubiquitin and SUMO in DPC repair

TOP1- and TOP2-ccs are extensively modified with ubiquitin, SUMO1, and SUMO2/3. Ubiquitination of TOP1- and TOP2-ccs promotes their proteasomal degradation (Desai et al., 1997; Lin et al., 2008; Mao et al., 2001). The ubiquitination of TOP1ccs is driven by the E3 ubiquitin ligase, Cullin 3 (CUL3) (Zhang et al., 2004). The impact of SUMO on TOP1- and TOP2-ccs is less clear (Mao et al., 2000b, 2000a). However, UBC9-deficient cells are sensitive to TOP1cc-inducing agents, suggesting that SUMO is indeed important for TOP1cc resolution (Mao et al., 2000a). In yeast, the SUMO-targeted ubiquitin ligase (STUbL) Slx5-Slx8 is important for the resolution of both TOP1- and TOP2-ccs (Heideker et al., 2011; Sun and Nitiss, 2017). STUbLs are ligases which conjugate ubiquitin chains to SUMO-modified proteins. This suggests that the sequential modification of TOP1- and TOP2-ccs by SUMO, then ubiquitin, may

drive their proteasomal degradation. Accordingly, the human orthologue of Slx5-Slx8, RNF4, is also important for TOP2cc proteolysis (Sun and Nitiss, 2017). In human cells, SUMO1 may therefore act as a platform for the recruitment of RNF4 to TOP1- and TOP2-ccs.

Yeast Ufd1 possesses a SUMO-interacting motif (SIM) that is important for countering the cytotoxic effects of TOP1ccs. Moreover, Ufd1 was proposed to act in the same TOP1cc repair pathway as Slx5-8 (Nie et al., 2012). This indicates that Cdc48-Ufd1 may be recruited to SUMO-modified TOP1ccs to promote their repair by the proteasome. However, no direct evidence of this was demonstrated and the same SIM in yeast Ufd1 is not conserved in humans.

More recently, a proteasome-independent pathway of TOP2cc repair was elucidated, involving the E3 SUMO ligase ZNF451/ZATT (Schellenberg et al., 2017). ZNF451/ZATT is proposed to facilitate TOP2cc resolution via two distinct mechanisms. First, it conjugates SUMO2/3 chains on TOP2ccs. TDP2 has a 'split-SIM' motif that recognises SUMO2/3-modified TOP2. However, it should be noted that a TDP2 variant that lacks its SIM was not defective for TOP2cc repair. The second mechanism of ZNF451-mediated TOP2cc resolution is via a conformational change in the TOP2 protein induced by ZNF451 binding. This may facilitate access of TDP2 to its substrate phosphotyrosyl bond.

Wss1 is also linked to the SUMO system as it possesses two SIMs which are at least partially required for its DPC-processing activity (Balakirev et al., 2015; Stingele et al., 2014). It is possible that SUMO marks specific substrates for binding by Wss1 and/or stimulates Wss1's protease activity. Another possibility is that SUMO is conjugated to another protein near a DPC (such as a replisome

component) to facilitate Wss1 recruitment. The reason Wss1 variants that lack the ability to bind SUMO remain partially functional may be because its binding partner, Cdc48, is able to bind directly to SUMO in yeast (Bergink et al., 2013; Nie et al., 2012).

A second study reported that Wss1 itself possesses E4 SUMO ligase activity due to its ability to elongate SUMO chain *in vitro* (Balakirev et al., 2015). A possible interpretation of this finding is that Wss1 can act as a protein scaffold that brings UBC9 into contact with substrates. It is not clear to what extent this activity is important for Wss1's function *in vivo* since there is a drastic accumulation of SUMOylated proteins, including TOP1, in the absence of Wss1 and TDP1 (Balakirev et al., 2015).

SPRTN variants that cannot bind ubiquitin are not recruited to DNA damage sites. Some groups suggested that this was because SPRTN binds mono-ubiquitinated PCNA, a possibility supported by their observations that SPRTN recruitment is dependent on Rad18 (Centore et al., 2012; Ghosal et al., 2012; Juhasz et al., 2012). While no groups have characterised SPRTN's enzymatic activity, it will be interesting to address whether SPRTN processes DPCs and if this requires ubiquitin.

#### 1.2.6.6 Clinical relevance of DPCs

In the 1950s, the National Cancer Institute (NCI) collected samples of thousands of plant extracts in order to test whether any of them exerted anti-tumour activity. One extract, prepared from the bark of Camptotheca Acuminata, a tree native to China, exhibited potent anti-cancer activity. In 1963, Monroe Wall isolated the active compound, Camptothecin (CPT) (Wall, 1998). CPT, was subsequently

shown to inhibit the completion of TOP1's catalytic cycle, resulting in the stabilisation of TOP1ccs (Hsiang et al., 1985). Two CPT analogues – topotecan and irinotecan - are FDA-approved for use in the treatment of multiple types of cancers (Pommier, 2013). CPT and its derivatives specifically bind the TOP1-DNA interface and prevent the re-ligation of the two ends of a broken strand DNA. They are commonly referred to as poisons, rather than inhibitors, denoting the fact that they stabilise a reaction intermediate rather than inhibit TOP1's catalytic activity.

An interfacial TOP2 poison, etoposide, along with its analogue tenoposide, is front-line therapy for many solid tumours, lymphomas, and leukaemias. Indeed, at one time it was estimated that over 50% of all chemotherapy regimens included TOP2-cc-stabilising drugs (Hande, 1998). The gene encoding the TOP2 isoform, TOP2α, is often co-amplified with the growth factor receptor, HER2, in breast cancers and renders these cancers highly sensitive to etoposide. Another class of drugs, the anthracyclines, stabilise TOP2ccs by intercalating in DNA (Nitiss, 2002).

A drug used to treat myelodysplastic syndromes and some cases of acute myeloid leukaemia, 5-aza-2′-deoxycytidine (5-aza-dC), covalently crosslinks the DNA methyltransferase, DNMT1, to DNA. 5-aza-dC is a cytosine analogue that is incorporated into DNA and acts as a pseudosubstrate of DNMT1. While attempting to methylate 5-aza-dC, DNMT1 becomes covalently trapped, which is thought to kill cancer cells by inducing global loss in DNA methylation (Patel et al., 2010).

DPC induction is being recognised as an important mode of action for numerous other anti-cancer drugs that are already in clinical use. Major examples are the platinum compounds, Cisplatin, Oxaliplatin, and Carboplatin (Ming et al., 2017). Platinum compounds can react with the amino acid side chains of proteins as well as with DNA bases. Platinum-DNA complexes can, in turn, react with DNA bound proteins such as histones, resulting in the formation of a DPC. Notably, there is evidence that the clinical efficacy of platinum compounds positively correlates with the extent to which they induce DPCs (Chválová et al., 2007).

More recently, three PARP inhibitors (PARPi) have been FDA-approved for the treatment of breast and ovarian cancers (Thomas et al., 2018). Interestingly, the three clinically-used PARPis exhibit differing degrees of potency, in a manner that does not correlate with their ability to inhibit PARP1's catalytic activity *in vitro*. It was recently discovered that these drugs not only inhibit PARP1's catalytic activity but also stabilise PARP1- and 2-DNA complexes, resulting in 'trapped PARP' accumulation (Murai et al., 2012). There is evidence that the efficacy of clinically-used PARPis correlates with their ability to trap PARP molecules to DNA (Murai et al., 2012). This model potentially explains why PARP inhibition is more toxic than PARP1 loss to BRCA-deficient cells and may represent the clinically-relevant mode of action of these drugs (Ding et al., 2016). PARP1 was subsequently found to be covalently trapped to abasic sites generated by the alkylating agent, methyl methansulfonate (MMS) (Prasad et al., 2014). A caveat of the PARP trapping model is that trapped PARP1 has only been visualised in the presence of a DNA damaging agent, such as MMS, whereas PARPis are clinically used as a monotherapy (Robson et al., 2017). Nevertheless, this raises the possibility that PARPis could be used in combination with temozolomide, a commonly used anti-

cancer drug that induces abasic sites (Murai et al., 2014). More recently, components of the ubiquitin-proteasome system, such as CUL1, have been implicated in the repair of covalently trapped PARP, suggesting that inhibition of their function may enhance the efficacy of PARPis (Michelena et al., 2018; Prasad et al., 2019).

#### 1.2.6.7 Unanswered questions in DPC repair

A particularly pressing question is to address if SPRTN is indeed a protease and, if so, what are its substrates and how is its activity regulated. There is already evidence that SPRTN deficiency sensitises cells to CPT, hinting that TOP1ccs may be SPRTN substrates (Davis et al., 2012). Moreover, point mutations in SPRTN's putative catalytic active site disrupt SPRTN's function, including its ability to suppress UV-induced mutagenesis and the cell cycle defects in *SPRTN*-deficient cells (Kim et al., 2013; Maskey et al., 2014). As discussed, UV light can also induce the formation of DPCs. Could evidence of a role for SPRTN in TLS, in fact, reflect a DPC repair defect? Indeed, the collision between a replisome and a lagging-strand DPC could conceivably generate ssDNA, trigger Rad18 activation, and initiate SPRTN recruitment. Moreover, what role, if any, does p97 play in DPC repair in human cells?

Answering these questions may help address the fundamental problem of how TOP1- and TOP2-ccs are processed upstream of TDP1 and TDP2. A contribution for the proteasome has been widely reported. The proteasome, however, requires substrates to be partially unfolded before it can cleave them. Could p97 fulfil this role? Moreover, a contribution for the proteasome in TOP1- and TOP2-ccs repair has only ever observed after treating cells with very high doses of camptothecin and etoposide, respectively. Given that TOP1- and TOP2-ccs are very common

endogenous DNA lesions, what mechanisms are employed to repair them in unchallenged cells? SPRTN's function is clearly linked to DNA replication, where such DPCs would present severe obstacles to DNA replication fork progression.

It will be important to address how DPC repair is regulated. PTMs are likely to be important, possibly to distinguish DPCs from other DNA-bound, but non-crosslinked, proteins or to directly activate/regulate DPC repair factors.

DPC repair has attracted a lot of attention in recent years primarily because it is anticipated that targeting DPC repair proteins could enhance the efficacy of the various clinically-used drugs that induce DPCs. It would therefore be very useful to test whether the expression of DPC repair components correlates with clinical response to DPC-based chemotherapy.

## 1.3 Aims of this study

Preliminary results from the Ramadan laboratory indicated that SPRTN is a DNA-dependent protease that counteracts the accumulation of DPCs, including TOP1ccs. The initial aim of this study was to contribute to the characterisation of SPRTN as a DPC repair protease. Specifically, my aim was to assess the role of SPRTN in counteracting DPCs during DNA replication. To address this question, I generated CRISPR-Cas9 *SPRTN* hypomorphic cells and employed two research techniques to study the impact of SPRTN haploinsufficiency on DNA replication: the DNA fibre assay and iPOND. The second aim of this study was to characterise, in detail, the mechanisms by which TOP1ccs are repaired. To this end a mass spectrometry analysis of TOP1-interacting proteins was carried out, resulting in the identification of the p97 ATPase. My aim was to assess whether p97 plays a role in TOP1cc repair in humans and, if so, to understand the mechanism by which it fulfils this role. We sought to identify the cofactor and PTMs that recruit p97 to TOP1ccs and to address if p97 regulates SPRTN's role in TOP1cc repair.

# 2 Materials and Methods

## 2.1 Buffers

### 2.1.1 10 x Phosphate Buffered Saline (PBS):

For 1 litre (L), the following were dissolved in 800 mL double-distilled (dd) $H_2O$:

- 18.1 grams (g) $Na_2HPO_4 \cdot 2H_2O$
- 80 g NaCl
- 2 g KCl
- 2.4 g $KH_2PO_4$

The pH was adjusted to 7.4 with HCl and filled to a volume of 1 L with $ddH_2O$.

1x PBS (working concentration: 137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4) was prepared by diluting the 10x PBS stock 1:10 in $ddH_2O$.

### 2.1.2 10x Tris Buffered Saline (TBS)

For 1 L, the following were dissolved in 900 mL $ddH_2O$:

- 88 g NaCl
- 24 g Trizma Base

The pH was adjusted to 7.6 with HCl and filled to a final volume filled of 1 L with $ddH_2O$.

1x TBS (20 mM Tris, 150 mM NaCl) was prepared by diluting the 10x TBS stock 1:10 in $ddH_2O$.

0.1% TBS-Tween (TBS-T) was prepared by adding 1 mL of Tween-20 to 1 L.

### 2.1.3 10x Tris/Glycine buffer

For 1 L, the following were dissolved in $ddH_2O$:

- 30 g Trizma base

- 144 g Glycine

The 10x stock of Tris/Glycine was diluted 1:10 for use as a running and transfer buffer for Immunoblotting. For running buffer, 1x Tris/Glycine was supplemented with SDS to a final concentration of 0.1%. For transfer buffer, 10x Tris/Glycine was diluted with dd$H_2O$, 20% methanol, and supplemented with SDS to a final concentration of 0.025%.

### 2.1.4 5x Laemmli Buffer

- Bromophenol Blue (SLS, 0.25%)
- Glycerol (Fisher Scientific, 50%)
- SDS (10%)
- Tris-HCl (0.25 M, pH 6.8)

1x LB (50 mM Tris-HCl pH 6.8, 2% SDS, 10% Glycerol, 0.05% bromophenol blue) was prepared in water supplemented with 15% β-mercaptoethanol (βMe).

### 2.1.5 Luria-Bertani (LB) broth and agar

LB broth/agar plates:

- 1 % Tryptone
- 0.5 % yeast extract
- 1 % NaCl
- LB agar plates were supplemented with 1.5 % agar

Where appropriate, LB broth and agar was supplemented with antibiotics (100 µg/ml ampicillin or 50 µg/ml kanamycin).

### 2.1.6 Crystal violet staining solution

For 100 mL, 0.5 g crystal violet powder was dissolved in 80 ml water and 20 ml methanol.

## 2.2 Cell culture

### 2.2.1 Cell line maintenance

Human HEK293, U2OS, RPE-1, and HeLa cells were obtained from ATCC. All cell lines were cultured in DMEM containing 10% FBS and 5% Penicillin/Streptomycin. All cell lines were regularly screened for mycoplasma using a MycoAlert™ Mycoplasma Detection Kit.

### 2.2.2 Generation of cell lines

#### 2.2.2.1 CRISPR-Cas9 *SPRTN* knockout HeLa cells

A plasmid (pX459v2), containing the sgRNA targeting exon 1 of *SPRTN* (CACGCTCCACTTCACCTCGACGG), Cas9, and a puromycin resistance cassette was obtained from the Genome Engineering Oxford (GEO) centre, Sir William Dunn School of Pathology, Oxford, UK. The plasmid was transfected into HeLa cells using Fugene HD (Promega). 24h after transfection, cells were selected with 0.6 µg/ml puromycin for 72h and then seeded as single cells in a 96-well plate (one cell/well) in order to generate a population of cells derived from a single clone (i.e. a genetically homogenous population of cells). SPRTN protein expression in HeLa cell clones (each grown from a single cell) was analysed by immunoblotting. A clone which showed a reduction in SPRTN protein levels was selected for sequencing. Genomic DNA was isolated from these cells using the DNeasy Blood & Tissue Kits (Qiagen). The region containing the *SPRTN* sgRNA

target site was amplified by PCR and subcloned into TOPO TA vectors using the TOPO TA Cloning kit (Thermo Fisher) to achieve allele separation. Plasmids were transformed into bacteria and grown on ampicillin-containing LB agar overnight at 37°C. The following day, 25 individual bacterial colonies were selected and cultured in LB broth overnight at 37°C, with shaking. Plasmid was then isolated from each bacterial colony and analysed by Sanger sequencing. 70% (i.e. 17/25) of the sequences showed Cas9-mediated mutations in exon 1 of *SPRTN*.

#### 2.2.2.2 CRISPR-Cas9 *TEX264* knockout cells,

Two plasmids – a CRISPR/Cas9 KO plasmid containing guide RNA targeting TEX264 (sc-417333), and a homology directed repair plasmid containing a puromycin resistance cassette (sc-417333-HDR) - were purchased from Santa Cruz. 2.5 µg of each plasmid was transfected into early-passage HEK293, HeLa, and U2OS cells using Fugene HD (Promega). After 72 hours, media supplemented with puromycin was added to the cells. The puromycin dose required to kill wild-type cells was determined to be: 1.25 µg/ml for HEK293 cells and 0.6 µg/ml for HeLa cells. After 72 hours, the puromycin-containing media was removed, and single cells were sorted using a cell sorter into 96-well plates. TEX264 expression was analysed by immunoblotting. Multiple clones of each cell line showing loss of all detectable TEX264 were selected for subsequent analysis.

#### 2.2.2.3 Flp-In T-Rex stable cell lines

HEK293 Flp-In T-REx stable cells were purchased from Thermo Fisher Scientific. cDNA encoding either p97 or SPRTN was sub-cloned in a pCDNA5/FRT/TO-cSSH vector. This construct was used as a template from which to generate the $p97^{E578Q}$ or $SPRTN^{SHP*}$ variants by site directed mutagenesis. Constructs were

transfected along with the pOG44 vector into HEK293 Flp-In T-REx cells to enable site-specific integration into the genome. The transformed cells were selected with hygromycin B to generate stable cell lines. To confirm successful integration of the desired cDNA, cells were treated with doxycycline (1µg/ml) for 24 hours, harvested, and protein expression was analysed by immunoblotting.

### 2.2.3 siRNA transfection

All siRNA transfections were carried out using Lipofectamine RNAiMAX (Invitrogen), according to the manufacturer's protocol and assayed after 72 hours.

| **Oligonucleotide** | **Sequence** |
|---|---|
| siLuc | CGUACGCGGAAUACUUCGA |
| sip97 #1 | AAGUAGGGUAUGAUGACAUUG |
| sip97 #2 | AAGAUGGAUCUCAUUGACCUA |
| sip97 #3 | AACAGCCAUUCUCAAACAGAA |
| siTEX264 #1 | CTCATCGACCTCTACCAGAAA |
| siTEX264 #2 | CGGCTGGAGATCTACCAGGAA |
| siTEX264 3'UTR | CUUCCAGGACCCAGAAUAA |
| siTOP1 | GGUCCCUGUUGAGAAACGA |
| siTDP1 | GGAUAUUGCGUUUGGAACA |
| siUBC9 | GUGGCUGUCCCAACAAAAA |
| siSPRTN #1 | AGCCAAUAUAACGGUAUACCA |
| siSPRTN 3'UTR | GUCAGGAAGUUCUGGUUAA |

### 2.2.4 Plasmid DNA transfection

Plasmid DNA transfections were performed using polyethyleneimine (PEI) reagent, Lipofectamine 3000 (Thermo Fisher), or FuGENE HD Transfection Reagent (Promega), following the manufacturer's instructions.

| **Recombinant DNA** |
|---|
| pCDNA5-FRT/TO-cSSH |
| pCDNA5-FRT/TO-SPRTN-wt-cSSH |
| pOTB7 TEX264 cDNA Clone IRAUp969H0427D |
| pGEX-4t-2 TEX264 nGST |
| pCDNA5 FRT/TO FlpIn cSSH |
| pET21a(+) TEX264 ΔLRR cHis |
| pET21a(+) TEX264 ΔLRR ΔSHP cHis |
| pCDNA5 FRT/TO FlpIn TEX264 cSSH |
| pCDNA3.1 TEX264-FLAG |

| |
|---|
| pCDNA5 FRT/TO TEX264-GFP |
| pET21a(+) p97-S•Tag |
| pCDNA5 FRT/TO p97-Myc-Strep |
| pCDNA5-FRT/TO-TEX264-wt-cSSH |
| pCDNA5-FRT/TO-TEX264- ΔLRR-cSSH |
| pNIC-ZB-SPRTN |
| pEGFP-p97 |
| pEGFP-p97 (1-199) |
| YFP-TOP1 |
| pCI-Myc-TDP1 |

## 2.3 Molecular and cellular biology techniques

### 2.3.1 PCR for Site-Directed Mutagenesis

Site-directed mutagenesis was carried out by PCR. Oligonucleotide primers used in PCR reactions were manually designed and purchased from Thermo Fisher Scientific. All reactions were performed using AccuPrime Pfx DNA Polymerase from Thermo Fisher Scientific.

Reactions were performed in a total reaction volume of 50 µl using the following components:

- 1 µl forward primer (10 pM)
- 1 µl reverse primer (10 pM)
- 5 µl 10X AccuPrime Pfx Reaction Mix
- 1 µl AccuPrime Pfx DNA Polymerase
- 200 ng template DNA
- Volume adjusted to 50 µl with sterile water

Reactions were carried out using the following PCR program:

| Temperature | Step | Time |
|---|---|---|
| 95°C | Initial denaturation | 30 seconds |
| 95°C | Denaturation | 10 seconds |
| Adjusted according to primer melting temperature (usually 50-60°C) | Anneal | 30 seconds |
| 68°C | Extend | 1 min/kb |
| 68°C | Final Extension | 8 mins |
| 4°C | Hold | ∞ |

Annealing temperatures were adjusted according to primer melting temperatures. Elongations times were adjusted according to the length of the template DNA according to the AccuPrime Pfx DNA polymerase manufacturer's instructions.

### 2.3.2 SDS-PAGE and immunoblotting

Acrylamide gels were prepared in the laboratory. Separating and stacking buffers were prepared separately, as follows:

A 4x separating buffer composed of 1.5 M Tris, pH 8.8, 0.4% SDS was mixed with 30% acrylamide (to a final concentration of 5 – 18%) acrylamide/Bis solution and ddH2O and set with 0.5% ammonium persulfate and 0.06% Tetramethylethylenediamine. Stacking buffer was prepared as described for separating buffer, except the stacking buffer is composed of 0.5 M Tris, pH 6.8, 0.4% SDS and the final concentration of acrylamide was 4%.

SDS-PAGE was performed using the Mini-PROTEAN Tetra Electrophoresis System (Bio-Rad). Proteins were separated at 150 V in running buffer. SDS-PAGE gels were then either stained using the ProteoSilver Plus Silver Stain Kit (Sigma-Aldrich) or transferred in transfer buffer at 100 V for 75 mins onto a PVDF

membrane (Bio-Rad), which had been activated with 100% methanol. Membranes were then washed once in 1 x TBS-T and blocked in 5% milk diluted in 1x TBS-T for 1 hr at room temperature. Primary antibodies against the protein of interest were diluted in 5% BSA (dissolved in in 1x TBS-T and supplemented with 0.02% sodium azide) and incubated with the membrane overnight 4˚C. The next day, membranes were washed three times in 1 x TBS-T (5 mins/wash) and incubated with the appropriate HRP-conjugated secondary antibody (diluted in 2.5% milk in TBS-T) for 1 hr at room temperature. Membranes were then washed three times in 1 x TBS-T (5 mins/wash), dried, and incubated with enhanced chemiluminescent substrate detection reagent (Thermo Fisher) for 5 mins. Proteins were visualised using the ChemiDoc MP system (Bio-Rad).

### 2.3.3 Preparation of whole cell lysates

Cells were collected in 1x PBS and harvested by centrifugation at 600 x g. Cell pellets were lysed in Radioimmunoprecipitation assay buffer (RIPA buffer), composed of 25 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100). Pellets were lysed for 20 mins at 4°C with rotation. Samples were sonicated using a Bioruptor Plus sonicator (Diagenode) for 30 seconds ON, 30 seconds OFF for 5 cycles to fragment DNA. Samples were centrifuged at 21,800 x g for 15 minutes. The supernatant was collected, and protein concentrations were quantified using the RC DC Protein Assay (Bio-Rad).

### 2.3.4 Cellular fractionations

Cells were resuspended in buffer A (10 mM Hepes pH 7.45, 10 mM KCl, 340 mM Sucrose, 10% Glycerol, Protease and Phosphatase Inhibitors, and 2 mM EDTA). Triton X-100 was added to a final concentration of 0.1% and cells were left on ice

for five mins. Cells were then centrifuged at 350 x g for three mins, the supernatant was collected (i.e. the cytosolic fraction), and the remaining pellet (i.e. nuclei) was washed twice in buffer A without Triton X-100. The nuclei were then resuspended in buffer B (3 mM EDTA, 0.2 mM EGTA, 5 mM Hepes pH 7.9, containing protease and phosphatase inhibitors) and left on ice for 10 mins. NP-40 was then added to a final concentration of 1% to remove membranes and left to incubate on ice for a further five mins. The nuclear fraction was then centrifuged at 1700 x g for five mins, the supernatant was collected (i.e. the nuclear soluble fraction). The remaining pellet was washed twice more in buffer B with 1% NP-40, and then twice in Benzonase buffer (50 mM Tris HCl pH 7.9, 50 mM NaCl, 5 mM KCl, 3 mM $MgCl_2$, containing protease and phosphatase inhibitors). The chromatin pellet was then centrifuged at 5000 x g for five mins and resuspended in Benzonase buffer, supplemented with 125 U of Benzonase (Merk Millipore) and incubated for 1 hour at room temperature or overnight at 4°C with rotation. After Benzonase digestion, samples were centrifuged at 15,000 xg for fifteen mins and the supernatant was collected (i.e. the chromatin soluble).

### 2.3.5 Immunoprecipitation (IP)

Cells were lysed in IP lysis buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100) or IP denaturing buffer (i.e. IP lysis buffer containing 1% SDS) supplemented with protease and phosphatase inhibitors and incubated on ice for 10 min. Denaturing IPs were quenched with Triton X-100 (to a final concentration of 1%). Chromatin was pelleted by centrifugation for 5 mins at 1,000 x g and then digested with Benzonase at room temperature in Benzonase buffer (2 mM $MgCl_2$, 50 mM Tris, pH 7.4, 150 mM NaCl). Lysates were incubated with anti-FLAG M2 agarose (Sigma-Aldrich), Strep-Tactin Sepharose (IBA), or GFP-trap agarose

(Chromotek) for 1-3 hr(s) on a rotator at 4°C, washed three times with IP wash buffer (150 mM NaCl, 50 mM Tris-HCl; 0.5 mM EDTA), and resuspended in 2× Laemmli buffer. For TAP experiments, cells were grown to 80% confluency on 245 x 245 mm cell culture dishes (2 per condition). Immunoprecipitation of TEX264-FLAG was performed as described above. After TEX264-FLAG was captured, samples were boiled for 5 mins in 1% SDS then quenched with 1% Triton-X100 and incubated with anti-HA beads (Thermo Fisher) for 16 hours at 4°C. The next day samples were washed three times in IP wash buffer and eluted in 2x Laemmlli buffer.

### 2.3.6 SUMO binding assay

HEK293 cells transiently expressing TEX264-GFP, TEX264$^{SIM1*}$, or TEX264$^{SIM2*}$ were lysed in a lysis buffer containing 3% Triton X-100 and 0.5 M NaCl. Samples were sonicated using a Bioruptor Plus sonicator (30 seconds ON, 30 seconds OFF for 3 cycles), then diluted (1:3 volume) in IP wash buffer (150 mM NaCl, 50 mM Tris-HCl; 0.5 mM EDTA) and incubated with GFP trap beads (Chromotek). After capture, the beads were washed 3 times in IP wash buffer and incubated with 1 µg of free SUMO1 or SUMO2 (Boston Biochem) for 90 mins at 4°C with rotation. After washing in 3 x in IP wash buffer, the samples were eluted in Laemlli buffer for 5 mins at 95°C.

### 2.3.7 Isolation of Proteins on Nascent DNA (iPOND)

iPOND was performed as described previously (Sirbu et al, 2011). Briefly, HEK293 cells were incubated for 10 mins with 10 µM EdU. In thymidine chase experiments, cells were incubated with normal media supplemented with 10 µM thymidine. Chromatin was fragmented into 50-300 bp fragments by sonication

with a Bioruptor Plus sonicator (30 seconds ON, 30 seconds OFF for 50 cycles). Where indicated, cells were treated with CPT (100 nM) for 1 hour in normal media which was supplemented with 10 μM EdU for the final 10 mins. EdU incubation times were 8 and 16 mins for WT cells and ΔTEX264 cells, respectively, to account for the reduced EdU incorporation rates in ΔTEX264 cells. Biotin-labelled EdU was captured by incubating samples overnight with streptavidin-coupled agarose beads (Merck Millipore).

### 2.3.8 Protein purification & *in vitro* interaction assays

p97 was cloned into a pET21a vector and purified with a C-terminal S-Tag from Rosetta *E. coli* using standard methods. TEX264 was purified without its hydrophobic N-terminus (amino acids 1-25; ΔNT) to render the protein more soluble and enable purification. ΔNT TEX264 was cloned into a pET21a vector (with a C-terminal His-tag) and expressed in Rosetta *E. coli*. Cell pellets were resuspended in denaturing resuspension buffer (800 mM NaCl, 20 mM Tris HCl pH 7.0, 8 M Urea) containing 1% Triton X-100 and PMSF, before adding lysozyme and incubating at room temperature for 30 mins. Cell lysates were then incubated with Ni-NTA agarose beads (Qiagen) for two hours at room temperature. Proteins were refolded by washing three times in 1x denaturing resuspension buffer (without Triton X100), containing decreasing concentrations of Urea (6 M, 4 M, 2 M, 0 M). Proteins were eluted from the beads by rotating at 15 rpm at 4°C in imidazole in 1x native resuspension buffer without Triton X-100.

Protein interaction studies were performed as follows: p97 was coupled to S-protein agarose (Merck Milipore) for 2 hours at 4°C in buffer containing 137 mM NaCl, 20 mM Tris HCl pH 7.0, then incubated TEX264$^{WT}$ or TEX264$^{\Delta SHP}$ in the same buffer, supplemented with 0.5% Triton X-100. TEX264$^{WT}$ was coupled to

HisPur™ Ni-NTA Magnetic Beads (Thermo Fisher), prior to incubation with recombinant Topoisomerase 1 (Inspiralis) in a buffer composed of 137 mM NaCl, 20 mM Tris HCl pH 7.0. Reactions were terminated by the addition of Laemmli sample buffer, and resolved by SDS-PAGE.

### 2.3.9 Rapid Approach to DNA Adduct Recovery (RADAR)

DPCs were isolated using a modified RADAR assay (Kiianitsa and Maizels, 2013). Cells were grown to ~70% confluency then lysed in M buffer (MB), containing 6 M guanidine thiocyanate, 10 mM Tris–HCl (pH 6.8), 20 mM EDTA, 4% Triton X-100, 1% N-lauroylsarcosine and 1% dithiothreitol. DNA was precipitated by adding 100% ethanol, then washed three times in wash buffer (20 mM Tris–HCl pH 7.4, 150 mM NaCl and 50% ethanol), and solubilised in 8 mM NaOH. DNA concentrations were quantified by NanoDrop™ and confirmed by slot blot analysis on a Hybond N+ membrane followed by detection with an anti-dsDNA antibody. For TOP1ccs, samples were digested with Benzonase for 30 mins at 37°C and analysed by slot blot analysis on a Nitrocellulose membrane.

### 2.3.10 Colony forming assay

HeLa cells were seeded in triplicate on 6-well plates and allowed to attach for 16 hours. Cells were then treated with the indicated doses of CPT (Sigma-Aldrich) for 24 hours, after which they were washed with PBS, and released into normal media. 7- 10 days later, colonies were fixed with 100% methanol for 5 mins and stained with crystal violet. The number of colonies was counted using the automated colony counter, GelCount™ (DTI-Biotech). The number of colonies in treated samples is expressed as a percentage of the number of colonies in the untreated samples.

### 2.3.11 *In vitro* TOP1cc repair assay

FlpIn T-REx HEK293 cells were induced with 1 mg/ml doxycycline for 48 hours to deplete TDP1, then treated with 10 µg/ml MG132 for 1 hour and 14 µM CPT for 30 mins. Cells were lysed in a guanidine hydrochloride-based lysis buffer (buffer (8 M guanidine hydrochloride, 30 mM Tris–HCl pH 7.5, 10 mM EDTA, 1% sarkosyl, pH adjusted to 7.5) and separated by CsCl-gradient fractionation for 24 hours at room temperature. TOP1ccs were then precipitated with ice cold 100% acetone at -80°C for 30 mins, washed with 70% ethanol, air dried, and resuspended in dialysis buffer (50 mM HEPES–KOH pH 7.5, 50 mM NaCl, 10 mM MgCl2, 0.5 mM TCEP). Samples were dialysed using Slide-A-Lyzer cassettes overnight at 4°C and sedimented by centrifugation at 15,000 x g for 20 mins. Samples were incubated with recombinant BSA, p97, and/or TEX264 (100 nM/reaction) in reaction buffer (50 mM HEPES pH 7.5, 50 mM NaCl, 10 mM MgCl2, 0.5 mM TCEP, ATP, 0.1 mg/mL protease-free BSA) at 37 °C for 30 min, before addition of the indicated concentrations of recombinant SPRTN for 1 h. Samples were slot blotted onto a nitrocellulose membrane, and probed with a TOP1cc-specific antibody.

### 2.3.12 Immunofluorescence

Visualisation of TOP1ccs by immunofluorescence was performed as described by Patel et al, 2016. Briefly, cells were fixed in 4% formaldehyde for 15 mins at room temperature, then permeabilised in 0.5% Triton X-100 for 15 mins at 4°C. Cells were then treated with 0.5% SDS for 5 mins are room temperature and washed 5 times in a buffer containing 0.1% Triton X-100 and 0.1% BSA diluted in PBS. After blocking in 5% BSA/PBS for 1 hour at room temperature, cells were

incubated with an anti-TOP1cc antibody (Merck) diluted 1:100 in 2.5% BSA/PBS. Following staining with secondary antibodies and DAPI, coverslips were mounted onto slides and imaged using a Nikon 90i microscope. Confocal microscopy was performed using Andor Dragonfly 200 on a Nikon Ti-E microscope. Z-stacks were taken at an interval of 0.1 µM across cells fixed identically to normal brightfield microscopy. ClearView-GPU™ deconvolution was carried out on samples where stated. Images were imported via Bio-Formats (ImageJ plugin) and analysis was carried out using ImageJ. Images displayed are single z-slices unless otherwise stated.

### 2.3.13 DNA fibre assay

HEK293 cells were incubated in media containing 30 µM CldU (Sigma-Aldrich, C6891) for 30 min, washed 3 x in PBS, and then incubated with media containing 250 µM of IdU (Sigma-Aldrich, 17125) for an additional 30 min. Cells were then treated with ice-cold PBS. Cells were lysed in 200 mM Tris-HCl pH 7.4, 50 mM EDTA and 0.5% SDS directly onto glass slides and then fixed with 3:1 methanol and acetic acid overnight at 4C. The next day, the DNA fibres were denatured with 2.5 M HCl, blocked with 2% BSA and stained with antibodies that specifically recognise either CldU (Abcam, Ab6326, dilution 1:500) or IdU (BD-347580, dilution 1:500). Anti-rat Cy3 (dilution 1:300, Jackson Immuno Research, 712-116-153) and anti-mouse Alexa-488 (dilution 1:300, Molecular Probes, A11001) were used as the respective secondary antibodies. Microscopy was performed using a Nikon 90i microscope. The lengths of the IdU-labelled tracts were measured using ImageJ software and converted into microns. Statistical analysis was done by GraphPad Prism software using an unpaired t-test.

### 2.3.14 Alkaline comet assay

Cells were trypsinised, inactivated in 1% FBS-PBS, resuspended in 1% low melting point agarose in PBS (37°C), and embedded on microscope slides pre-coated with 1% normal melting point agarose in dH2O. Cells were lysed for 1 h at 4°C in a buffer containing 2.5 M NaCl, 100 mM EDTA, 10 mM Tris-HCl, 1% (v/v) DMSO and 1% (v/v) Triton X-100, at pH 10.5. Slides were then incubated in 4°C electrophoresis buffer (300 mM NaOH, 1 mM EDTA, 1% (v/v) DMSO, pH>13) for 30 mins to unwind DNA. Electrophoresis was performed in the same buffer for 25 mins at 25 V (300 mA). Cells were then neutralised in 0.5 M Tris-HCl pH 8.1 and stained with SYBR Gold (1:10,000 in dH2O) for 30 mins. Microscopy and analysis were performed using the Nikon NiE microscope and Andor Komet7.1 software. At least 100 cells were quantified per condition. Data are shown as Olive tail moments, calculated as: (Tail mean – Head mean) * Tail%DNA / 100.

## 2.4 Immunohistochemistry and mRNA expression analysis

The study protocol was in accordance with the ethical guidelines of the Helsinki declaration. Tissue samples were prepared from paraffin blocks according to standard histological protocols and immunohistochemical staining was performed using Leica Bond automated staining system. SPRTN (Atlas) and TEX264 (Novus) antibodies were diluted 1:1000. The p97 antibody (Proteintech) was diluted 1:2000. Appropriate IgG isotype–matched negative control antibodies were used in each case. Images were analysed using Aperio ImageScope software. Staining intensity was classified as follows: 0, negative; 1, low; 2, moderate; 3, high. A pathologist estimated the percentage of positively stained

tumour cells. Samples were scored in a blinded fashion. H-scores were calculated by multiplying the percentage of positive cells by the intensity of expression.

Transcriptomic data of FOCUS patients from patients treated with FOLFIRI (folinic acid, fluorouracil and irinotecan) was obtained from the S:CORT consortium. In brief, one section of tumour was cut for H&E staining followed by four unstained sections. H&E slides were re-reviewed by expert gastrointestinal pathologists and tumour tissue annotated to guide RNA extractions with the High Pure RNA Paraffin Kit (Roche). Transcriptional profiles were generated with the Almac Xcel array (Affymetrix). Cel files were pre-processed and normalised using robust multi-array analysis with the R package affy and probes collapsed by median. Data for response to treatment was obtained from the clinical trial database showing 4 categories. Association with relevant genes was analysed with linear regression. Statistical significance was considered at $p<0.05$. The use of patient material for FOCUS cohort as part of the S:CORT program was approved by the ethics commission (REC 15/EE/0241).

Ethics approval was granted by the National Research Ethics Service South Central Oxford—Panel C ethics committee (number 13/SC/0111). Regulatory approval was granted by the UK Medicines and Healthcare products Regulatory Agency (clinical trial authorisation [CTA] number 00316/0245/001-0001). All trial procedures and processes complied with the International Conference on Harmonisation's Good Clinical Practice guidelines. We asked patients to sign a consent form for registration and for analysis of the biomarker panel; additional written informed consent was obtained before randomisation in FOCUS.

## 2.5 Mass spectrometry

Immunoprecipitation of tagged proteins was performed as described in 2.3.4. Proteins samples were prepared by Trichloroacetic Acid (TCA) precipitation and resuspended in 6M urea. Proteins were then reduced with dithiothreitol (5 mM) alkylated with iodoacetamide (20 mM) and precipitated with methanol/chloroform. Proteins were resuspended in 50 ml 6M urea and urea was adjusted to a final concentration of 1 M. Proteins were digested overnight at 37°C with trypsin. Peptides were purified using SEP-C18 purification columns and concentrated using a SpeedVac vacuum. Peptides were dissolved in 15µl LC-MS grade water containing 2% acetonitrile and 0.1% trifluoroacetic acid. Samples were subsequently analysed by nano-liquid chromatography tandem mass spectrometry (LC-MS/MS) using a Dionex Ultimate 3000 UPLC coupled to a Q Exactive hybrid quadrupole orbitrap mass spectrometer (ThermoFisher Scientific). Briefly, peptides were desalted on a PepMapC100 column (100µm x 20mm, 5µm particle size, ThermoFisher Scientific) for 3 mins at a flow rate of 8 µl/min and separated on a directly coupled nEASY column (PepMap C18, 75 µm x 500mm, 2 µm particle, ThermoFisher Scientific) using a multistep gradient starting with 3 mins at 2% Acetonitrile in 5% DMSO with 0.1% Formic acid (buffer B), followed by 60 mins linear gradient up to 35% buffer B at 250nl/min flow rate, 7 mins linear gradient up to 99% buffer B and maintained at 99% buffer B for 5 mins at a flow rate of 250nl/min before reverting to 2% buffer B for 3 mins prior to a last 4 mins at 99% solvent B. Full MS scans were acquired over the m/z range of 380 - 1800 at a resolution of 70,000 at 200m/z (AGC target of 3x106 ions). MS/MS data was acquired in a data dependent manner by selecting the 15 most

abundant precursor ions for HCD fragmentation (CE of 28) and on MS/MS resolution of 17,500.

## 2.6 Statistical analysis

Where statistical analysis was performed, the number of replicates and error bars are indicated in the figures/figure legends. The statistical tests (Student's t-test, ANOVA) and P values (i.e. asterisks and the values they denote) are indicated in the figure legends and/or figure panels and were calculated using GraphPad Prism. Experiments were performed at least two times, except the mass spectrometry analysis of TEX264-SSH and YFP-TOP1, which were each performed once. It should be noted that the Student's t-test is limited in its ability to determine statistical significance where n=2 due to biological and technical variation between experiments which can inflate the standard deviation (e.g. due to outliers) and render the test less sensitive.

## 2.7 Antibodies used in this study

| Antibody | Source | Catalog number |
|---|---|---|
| Phospho-gammaH2AX (Ser139) | Genetex | GTX127342 |
| HA [clone 3F10] (rat monoclonal) | Roche | 3F10 |
| PCNA (mouse monoclonal) | Abcam | ab29 |
| p97 (rabbit polyclonal) | Proteintech | 10736-1-AP |
| p97 (rabbit polyclonal) | In house | N/A |
| Topoisomerase 1 (rabbit polyclonal) | Bethyl Laboratories | A302-589A |

| | | |
|---|---|---|
| Topoisomerase I-DNA Covalent Complexes Antibody, clone 1.1A (mouse monoclonal) | Millipore | MABE1084 |
| TEX264 (rabbit polyclonal) | In house | N/A |
| TEX264 (mouse monoclonal) | Novus | H00051368-M01 |
| Vinculin (mouse monoclonal) | Bethyl Laboratories | ab18058 |
| FLAG (rabbit polyclonal) | Sigma-Aldrich | F7425 |
| Myc (mouse monoclonal) | Cell Signaling Technology | 2276 |
| Double-stranded DNA (mouse monoclonal) | Abcam | ab27156 |
| SUMO1 (rabbit polyclonal) | Abcam | ab11672 |
| SUMO2/3 , clone 18H8 (rabbit monoclonal) | Cell Signaling Technology | 4971 |
| SPRTN N-terminal (rabbit polyclonal) | In house; Lessel *et.al*, 2014 | N/A |
| SPRTN C-terminal (rabbit polyclonal) | Atlas | HPA025073 |
| GFP | Abcam | ab290 |
| POLD1 (mouse monoclonal) | Abcam | ab10362 |
| MCM7 (141.2) (mouse monoclonal) | Santa Cruz Biotechnology | sc-9966 |
| Histone H3 (rabbit polyclonal) | Abcam | ab1791 |
| ds-DNA (mouse monoclonal) | Abcam | ab27156 |

| | | |
|---|---|---|
| TDP1 (rabbit polyclonal) | Abcam | ab4166 |
| Mono- and polyubiquitinylated conjugates, mAb (FK2) | Enzo Life Sciences | BML-PW8810 |
| Tubulin (mouse monoclonal) | Sigma-Aldrich | T6199 |
| Histone H2A.X, phospho (Ser139; clone JBW301; mouse monoclonal) | Millipore | 05-636 |
| UBC9 (rabbit monoclonal) | Abcam | ab75854 |
| Rabbit IgG, HRP-conjugated (goat polyclonal) | Sigma-Aldrich | A9169 |
| Mouse IgG, HRP-conjugated (rabbit polyclonal) | Sigma-Aldrich | A9044 |
| Mouse IgG, Alexa Fluor 488 (goat polyclonal) | Molecular Probes | A-11001 |
| Rabbit IgG, Alexa Fluor 594 (donkey polyclonal) | Thermo Fisher Scientific | R37119 |
| Mouse IgG, Alexa Fluor 488 | Thermo Fisher Scientific | R37114 |
| Rat IgG, Cy3-AffinityPureF(ab')2 Fragment (donkey polyclonal) | Jackson ImmunoResearch Labs | 712-166-153 |

# 3 Results

## 3.1 SPRTN promotes DPC repair and DNA replication

To establish a system in which to characterise the role of SPRTN in DPC repair we used CRISPR/Cas9-mediated genome editing to disrupt the *SPRTN* gene in HeLa cells (Fig. 9A). A single guide (sg)RNA targeting exon 1 in *SPRTN* was used and SPRTN protein expression was screened by immunoblotting in multiple clones each derived from single cells. We did not generate a clone which completely lacked SPRTN protein expression, suggesting that SPRTN is essential for cell viability. SPRTN had previously been identified in genome-wide CRISPR screens as an essential gene in 4 out of 5 human cell lines, and has since been found to be essential in 206 of 324 human cancer cell lines (Behan et al., 2019; Hart et al., 2015). We therefore selected HeLa clones where SPRTN protein expression was reduced by approximately 70% (Fig. 9B). To characterise the CRISPR-Cas9-induced mutations at the *SPRTN* locus, genomic DNA was isolated from one of these HeLa clones. DNA was also isolated from unedited parental HeLa cells for comparison. The DNA encoding exon 1 of *SPRTN* was amplified by PCR, subcloned into TOPO-TA vectors for allele separation, and transformed into bacteria. Plasmid was purified from multiple bacterial colonies and analysed by Sanger sequencing. In the parental cells, the *SPRTN* sequence was the same as the sequence in the human genome reference database. However, in the CRISPR-Cas9-edited cell line, 17 out of 25 (68%) sequences were mutated (Fig. 9C). In total three different sequences identified: wild-type and two different point mutations, resulting in either a premature STOP codon or a frameshift that disrupted the protein sequence (Fig. 9C). As HeLa cells contain

three copies of the chromosome that encodes *SPRTN* (chromosome 1), this indicates that 2 out of 3 *SPRTN* alleles had been disrupted. These SPRTN-hypomorphic HeLa cells are referred to as ΔSPRTN cells hereafter.

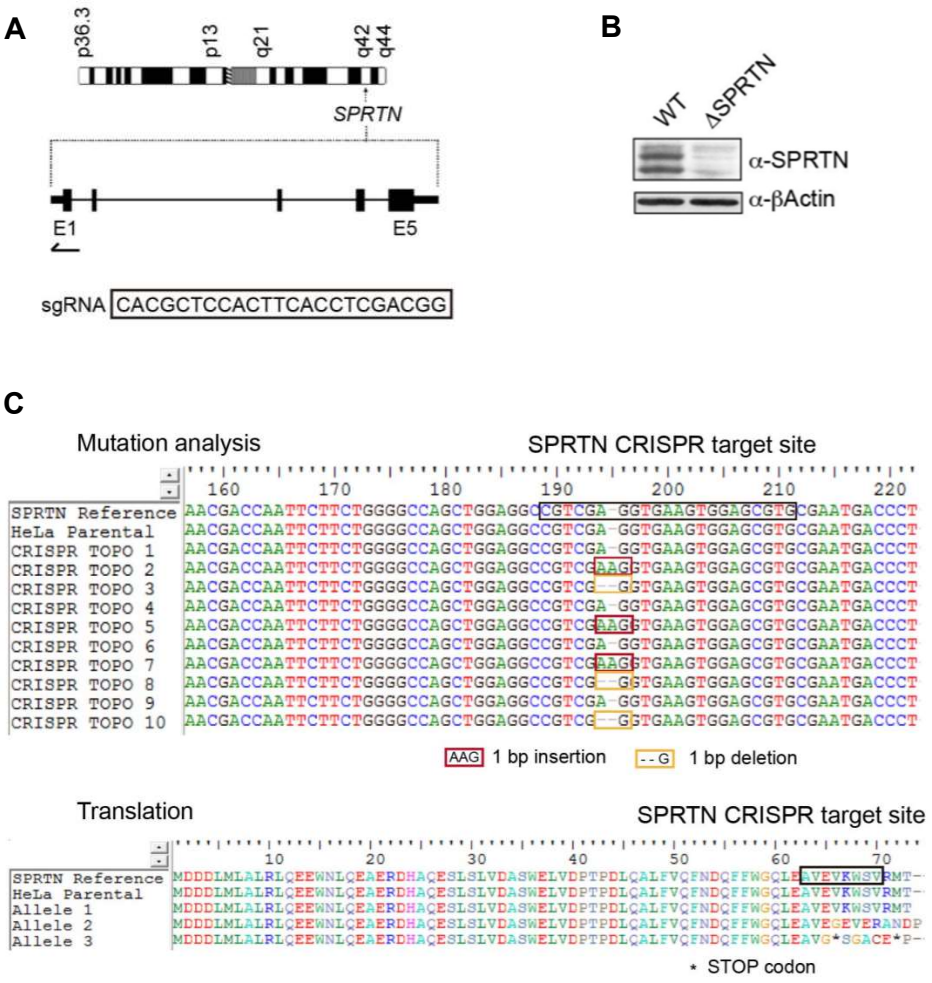



**Figure 9. Generation of ΔSPRTN HeLa cells by CRISPR-Cas9.**

(A) Schematic of sgRNA targeting exon 1 of *SPRTN*. (B) Immunoblot of SPRTN expression in wild-type (WT) and ΔSPRTN HeLa cells. (C) Upper panel: representative DNA sequences of exon 1 of *SPRTN* in parental and ΔSPRTN. Each line represents a different sequence subcloned into a TOPO-TA vector. ~70% of all analysed *SPRTN* sequences were mutated by CRISPR-Cas9, resulting in either a 1 base pair (bp) deletion or 1 bp insertion. Lower panel: predicted protein sequence generated by each sequence. 2/3 *SPRTN* alleles were mutated by CRISPR-Cas9 genome editing, either due to frameshift mutation or the introduction of a premature stop codon. Figs. 9A & B were generated by Bruno Vaz.

To assess whether DPC repair is compromised in ΔSPRTN cells, a version of the recently described RADAR (rapid approach to DNA adducts recovery) assay was used (Kiianitsa and Maizels, 2013). The RADAR assay involves lysing cells in a very stringent buffer, containing chaotropic salts, detergents, and a reducing agent. This ensures that all molecular interactions, other than covalent interactions, are disrupted. DPCs can then be resolved by SDS-PAGE and visualised by silver staining or, alternatively, transferred to a membrane so that specific DPCs can be analysed by immunoblotting.

ΔSPRTN cells displayed a substantial increase in total DPCs as visualised by silver staining (Fig. 10A). This could be reversed by exogenous expression of wild-type (WT) SPRTN but not a variant with a point mutation (E112A) in its putative HEXXH metalloprotease active site. This demonstrates that SPRTN counteracts DPCs in a manner that depends on its putative metalloprotease activity. The accumulation of specific DPCs, such as TOP1ccs, was confirmed by immunoblotting upon siRNA-mediated depletion of SPRTN (Fig. 10B).

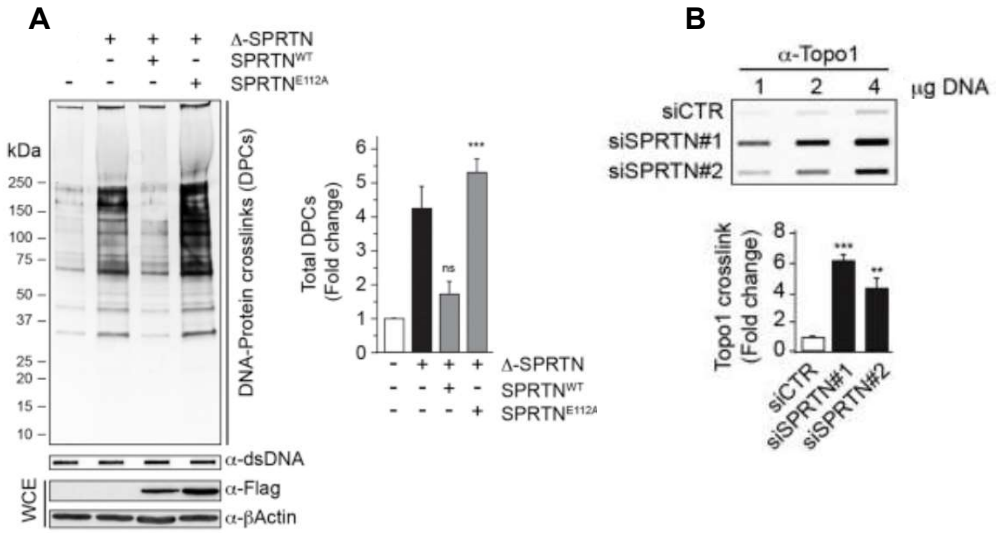



**Figure 10. SPRTN counteracts DPC accumulation.**

(A) Left panel: Analysis of DPCs in WT or ΔSPRTN HeLa cells, transfected with empty vector or cDNA encoding the indicated SPRTN variants. DPCs were isolated by RADAR, resolved by SDS-PAGE, and visualised by silver staining. Right panel: corresponding quantification. Error bars represent mean ± SEM, n = 3, ***$P$ < 0.0001; ns, not significant; Student's t-test. (B) Upper panel: TOP1ccs in HeLa cells transfected with the indicated short-interfering (si)RNA sequences. DPCs isolated by RADAR were transferred to a nitrocellulose membrane and probed with a TOP1-specific antibody. Lower panel: corresponding quantification. Error bars represent mean ± SEM, n = 3, ***$P$ < 0.0001; ** $P$ < 0.001; Student's t-test). All figures generated by Bruno Vaz.

Work performed by others in the laboratory subsequently found that SPRTN can cleave numerous DNA-binding proteins in a DNA-dependent manner *in vitro*, including histones, TOP1, and TOP2 (Vaz et al., 2016). The E112A SPRTN variant was proteolytically inactive. The patient variant SPRTN-Y117C protein displayed a dramatically reduced ability to cleave substrates, suggesting this mutation disrupts SPRTN's active site. A C-terminally truncated patient variant, SPRTN-ΔC (amino acids 1-240), could cleave histones, but not other substrates, such as TOP1 and TOP2, suggesting the C-terminus may be important for enabling SPRTN to bind and/or process certain substrates (Vaz et al., 2016).

A variety of evidence indicates that SPRTN plays a role in DNA replication. It is only expressed in the S and G2 phases of the cell cycle, possesses a PCNA-interacting motif, and forms nuclear foci that co-localise with PCNA (Davis et al., 2012; Mosbech et al., 2012). Fibroblasts derived from RJALS patients also exhibit DNA replication stress, including slower DNA replication fork progression and elevated levels of fork stalling (Lessel et al., 2014).

To test whether SPRTN activity is important for DNA replication in the presence of DPCs, we employed the DNA fibre assay. The velocity of DNA replication fork progression can be measured based on the rate at which cells incorporate nucleotide analogues. DNA fibres can then be visualised by microscopy using fluorescent antibodies that recognise these nucleotide analogues. In the presence of the DPC-inducing agent, formaldehyde, the rate of DNA replication fork progression in ΔSPRTN cells was approximately two-fold slower than in parental HeLa cells (Fig. 11A). Exogenous expression of WT SPRTN could fully correct this defect (Fig. 11A). This suggests that SPRTN activity is required for DNA replication in the presence of DPCs.

The identification and study of proteins that act at or near DNA replication forks has been greatly aided by the development of the iPOND (isolation of proteins on nascent DNA) technique (Sirbu et al., 2011). iPOND is a reverse chromatin immunoprecipitation procedure in which newly synthesised DNA is purified along with its associated proteins which, in turn, can be analysed by immunoblotting. Based on its role in influencing DNA replication fork progression, we asked whether SPRTN is enriched at replication forks (click sample) compared with mature chromatin behind replication forks (chase sample) in Human Embryonic Kidney 293 (HEK293) cells. We found that SPRTN was present at DNA replication forks along with PCNA, Pol δ, and MCM3, but absent on mature chromatin behind forks (Fig. 11B). This strongly supports the conclusion that SPRTN acts at DNA replication forks. In ΔSPRTN cells, replication proteins were retained in the 'chase' samples, indicating that replisome progression is impeded in ΔSPRTN cells (Fig. 11C). Since SPRTN's enzymatic activity promotes DNA replication fork progression and because some of its substrates, such as TOP1ccs, can stall fork progression, we assessed whether TOP1 accumulates at DNA replication forks in ΔSPRTN cells. We observed that the amount of TOP1 protein at DNA replication forks was much higher in ΔSPRTN cells than in control cells (Fig. 11D). This likely reflects the accumulation of TOP1ccs at DNA replication forks. Altogether these results support the conclusion that SPRTN's promotes DNA replication by counteracting DPCs that impede the replisome (Vaz et al., 2016).

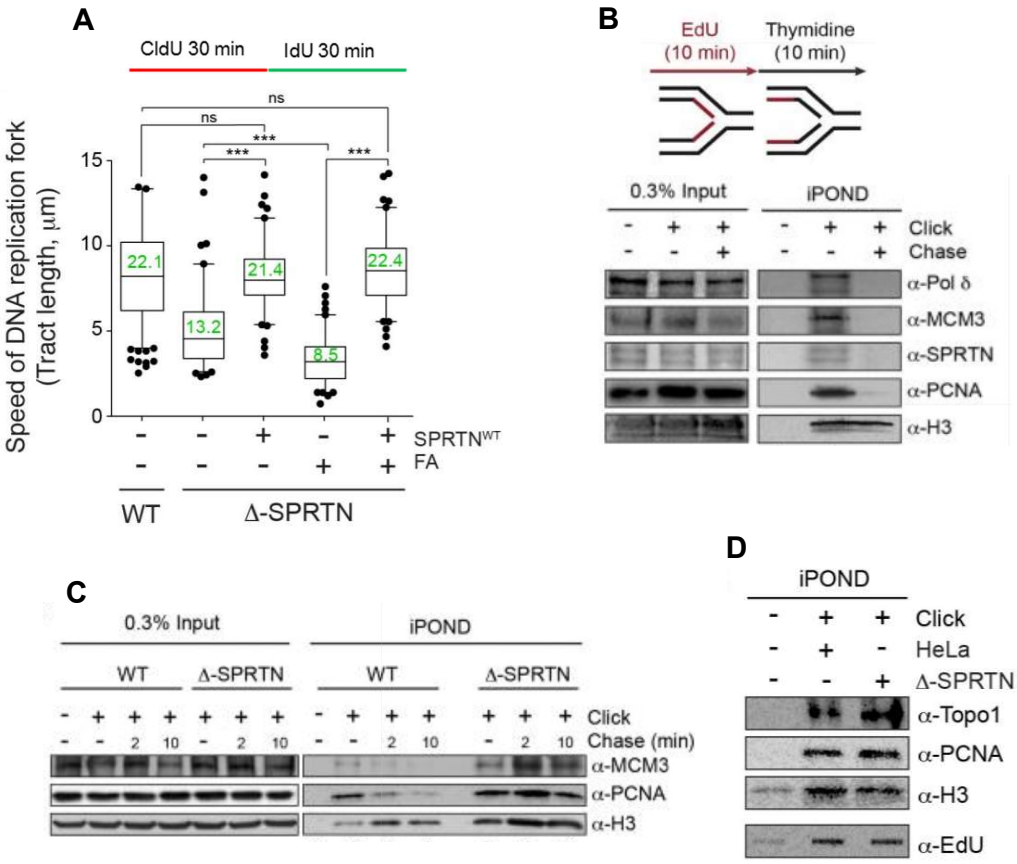



**Figure 11. SPRTN regulates DNA replication fork progression.**

(A) Upper panel: Schematic of DNA fibre assay. Lower panel: DNA fibre analysis of replication fork velocity in WT and ΔSPRTN HeLa cells. Where indicated, cells were co-treated with 50 μM formaldehyde (FA) for 30 mins along with IdU. Measurements of IdU track lengths are shown. At least 100 fibres were measured per condition. Whisker box plots show median values and data within the 10–90 percentile (***$P < 0.001$; two-tailed Mann-Whitney test). Green numbers represent mean tract lengths as kilobases. (B) Upper panel: schematic representation of iPOND approach. Lower panel: immunoblots of proteins isolated by iPOND in HEK293 cells. (C) Immunoblots of proteins isolated by iPOND from WT and ΔSPRTN HeLa cells. (D) Same as C. An anti-Streptavidin HRP antibody that recognises biotinylated EdU was used to ensure equal amount of nascent DNA was isolated from each sample.

## 3.2 Identification of new TOP1cc repair proteins

During our analysis of SPRTN's role in DPC repair, we noted that SPRTN preferentially cleaves disordered protein regions (Vaz et al., 2016). For example, it cleaves histones in their unstructured N-terminal tails. However, SPRTN was not active on recombinant, intact TOP1; it could only cleave TOP1 isolated from cells under denaturing conditions. Unlike histones, the TOP1 protein is large, globular, and lacking in unstructured protein regions (Li et al., 2019). We therefore reasoned that additional factors, such as PTMs or binding partners, may be necessary to promote TOP1cc proteolysis by SPRTN in cells, likely by making the TOP1 protein more amenable to cleavage.

To identify new potential modulators of TOP1cc repair, we took an unbiased proteomic approach. TOP1ccs are very common endogenous DNA lesions so we reasoned that factors that promote TOP1cc repair should interact with the TOP1 protein even in the absence of CPT. Chromatin was isolated from HEK293 cells expressing exogenous TOP1 tagged with YFP at its amino (N) terminus. Protein complexes containing YFP-TOP1 were purified by co-immunoprecipitation using anti-GFP beads and subjected to tandem liquid chromatography mass spectrometry (LC-MS/MS).

In total, we identified 1015 TOP1-interacting proteins and ranked them by the number of significant peptide sequences detected. As expected, we identified numerous chromatin remodelling factors as well as proteins involved in RNA splicing and transcription elongation (Fig. 12A). We proceeded to search this list for candidate regulators of TOP1cc repair. We identified peptides from five proteasome subunits, the most abundant of which was PSMD3, the 714$^{th}$ ranked

TOP1-interacting protein. We did not detect any SPRTN peptides, probably because SPRTN is present at very low abundance in cells (Beck et al., 2011). The ATPase p97 was the 154$^{th}$ most abundant TOP1-interacting protein. Since Cdc48 is implicated in TOP1cc repair and because p97 interacts with SPRTN, we asked whether p97 contributes to TOP1cc repair in human cells (Ghosal et al., 2012; Mosbech et al., 2012; Nie et al., 2012).

We first validated that endogenous p97 interacts with YFP-TOP1 in co-immunoprecipitation experiments (Fig. 12B). We then immunoprecipitated either wild-type p97 (p97$^{WT}$) or an ATPase-defective variant (p97$^{EQ}$; E578Q), tagged with myc-Strep epitopes at the carboxy (C) terminus, and monitored their interactions with endogenous TOP1. We found that endogenous TOP1 interacts much more strongly with p97$^{EQ}$ than with p97$^{WT}$ (Fig. 12C, D). The E578Q mutation disrupts p97's D2 ATPase domain and almost completely abolishes p97's ability to hydrolyse ATP (Dalal et al., 2004). Thus, p97$^{EQ}$ can interact with substrates but cannot efficiently process and release them, resulting in a 'trapping effect'. This result suggests that the TOP1 protein may be a substrate of p97.

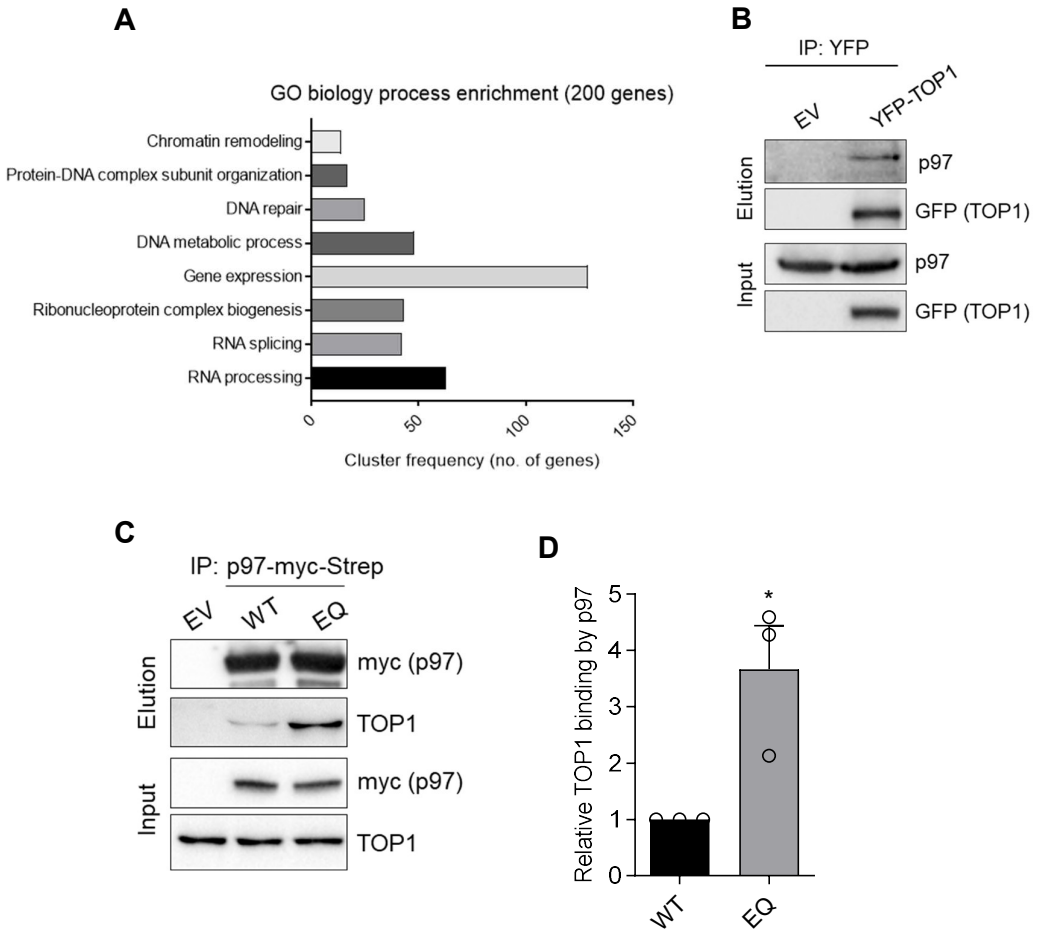



**Figure 12. p97 interacts with TOP1.**

(A) Analysis of gene ontology (GO) terms enriched among peptides identified in YFP-TOP1 LC-MS/MS. (B) Immunoblots of chromatin-associated anti-GFP immunoprecipitates prepared from HEK293 cells expressing either empty vector (EV) or YFP-TOP1. (C) Immunoblots of anti-Strep immunoprecipitates prepared from HEK293 cells transfected with cDNA encoding either WT p97 or the E578Q (EQ) variant. (D) Quantification of C. Error bars represent mean ± SEM, n = 3.

To assess if p97 deficiency causes TOP1cc accumulation we used the RADAR assay. Cells in which p97 was depleted with either of two siRNA sequences exhibited higher levels of TOP1ccs (Fig. 13A-C). TOP1cc accumulation resulting from p97 depletion was similar to that induced by a high dose of CPT (Fig. 13B). This suggested that p97 is important for TOP1cc repair. To test if this requires the ATPase activity of p97, we monitored TOP1cc levels in stable, doxycycline-inducible HEK293 cells, expressing either $p97^{WT}$- or $p97^{EQ}$-myc-Strep at near endogenous levels (Fig. 14B). E578Q is a dominant negative mutation meaning $p97^{EQ}$ antagonises endogenous p97. We found that TOP1ccs only accumulated in cells expressing $p97^{EQ}$, suggesting that p97's ATPase activity is important for TOP1cc repair (Fig. 14A, C). However, abrogating p97's function can have pleotropic effects, such as depleting the cellular pool of ubiquitin. Therefore, in parallel, we monitored total ubiquitin levels in these cells. Short-term induction of $p97^{EQ}$ did not result in depletion of the free pool of ubiquitin (Fig. 14B). Thus, defective TOP1cc repair resulting from impaired p97 function is not an indirect consequence of ubiquitin depletion.

TOP1ccs can also be visualised by immunofluorescence using a recently developed antibody that specifically recognises TOP1ccs (Fig. 14D) (Patel et al., 2016). This antibody recognises 15 amino acids in the active site of TOP1 as well as the phosphate group on DNA to which TOP1 is covalently linked. We used a specific inhibitor of p97's D2 ATPase activity, CB-5083, to further test our conclusion that this activity is important for TOP1cc repair (Anderson et al., 2015). A short treatment with CB-5083, but not its solvent DMSO, caused significant TOP1cc accumulation in Retinal Pigmented Epithelial (RPE-1) cells (Fig. 14E, F). Overall, these results establish p97 as a TOP1-interacting protein with a role in

TOP1cc repair. Importantly, p97 inhibition or depletion caused TOP1ccs to accumulate even in the absence of CPT treatment, confirming that TOP1ccs are common endogenous DNA lesions that are counteracted by p97.

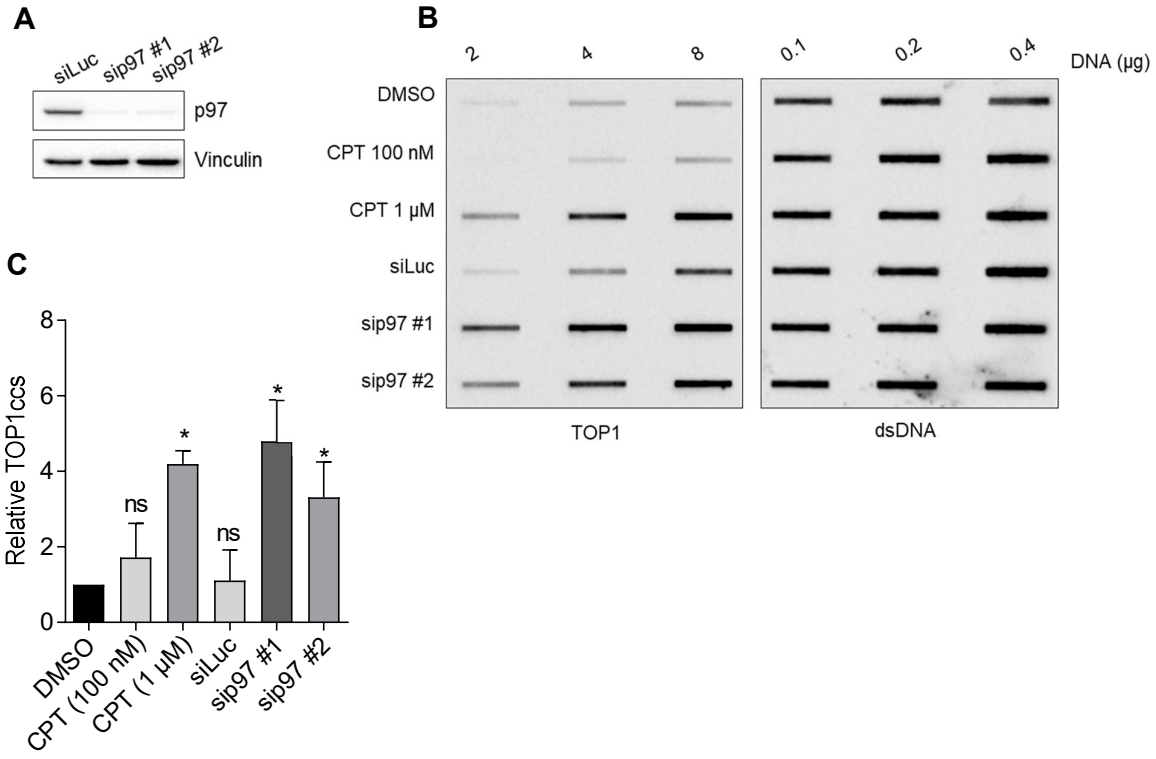



**Figure 13. p97 counteracts TOP1ccs.**

(A) Immunoblot to confirm siRNA-mediated depletion of p97 in HEK293 cells. (B) RADAR assay to assess TOP1cc accumulation after siRNA-mediated depletion of p97. Where indicated, cells were treated with CPT for 1 hour. Double-stranded (ds)DNA is used as a loading control. (C) Quantification of B (error bars represent mean ± SEM; $n = 3$; $^*P < 0.05$; ns, not significant; Student's t-test).

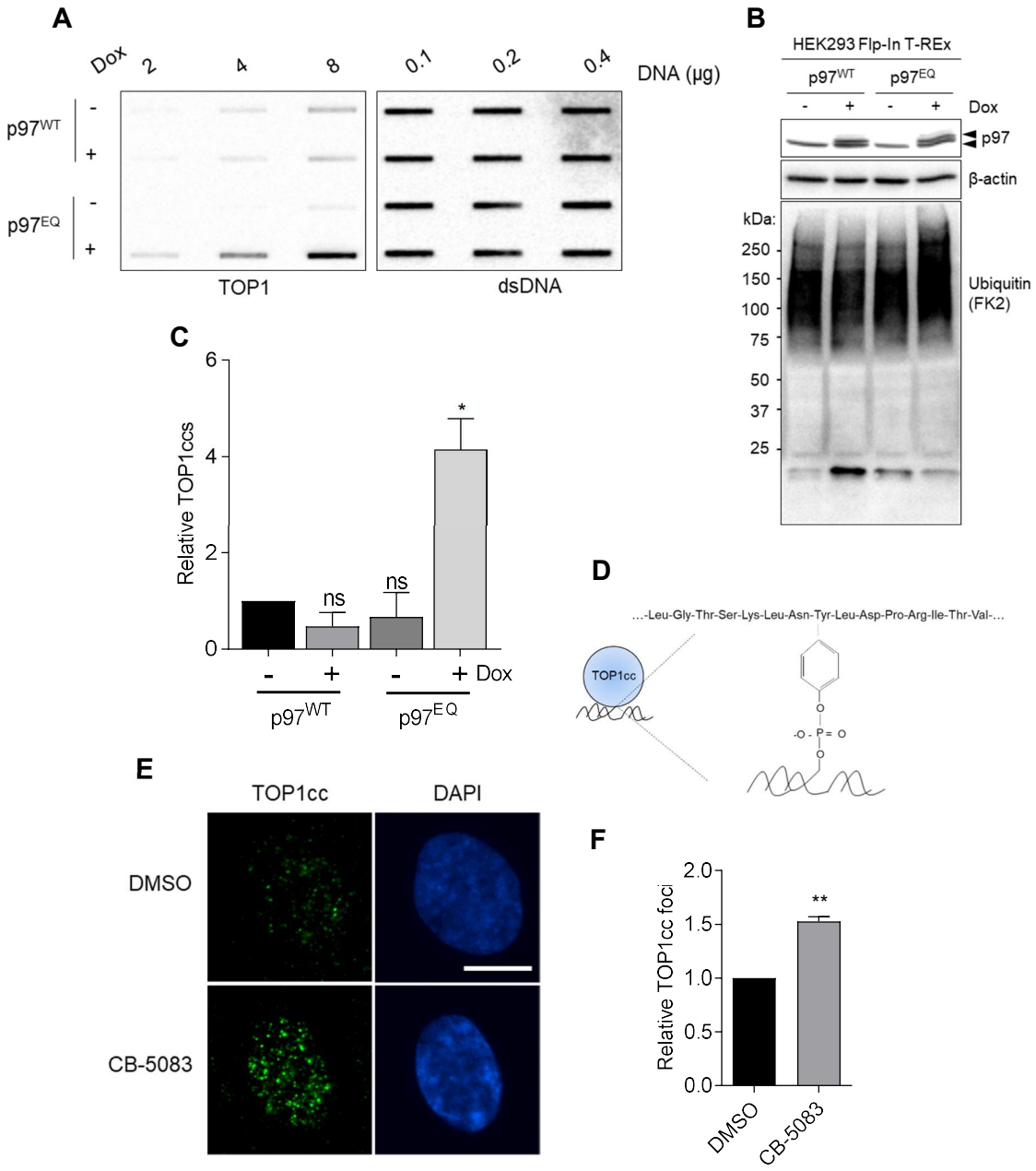



**Figure 14. p97 ATPase activity is required for TOP1cc repair.**

(A) RADAR assay monitoring TOP1ccs in doxycycline (Dox)-inducible HEK293 Flp-In T-REx cells expressing the indicated p97 variants. Where indicated, cells were treated with Dox for 36 hours. (B) Immunoblots of whole cell extracts (WCE) prepared from the same cells as those analysed by RADAR in A. Arrowheads indicate endogenous p97 (lower band) and induced p97-myc-Strep (upper band). (C) Quantification of A (error bars represent mean ± SD; n = 2; **P* < 0.05; Student's t-test). (D) Schematic representation of antigen recognised by TOP1cc-specific antibody. (E) Immunofluorescence detection of TOP1ccs (green) in RPE-1 cells treated with DMSO or CB-5083 (10 µM) for 4 hours. Nuclear DNA is stained with DAPI (4′,6-diamidino-2-phenylindole; blue). Scale bar, 10 µm. (F) Quantification of E (error bars represent mean ± SD; *n* = 2; ***P* < 0.01; Student's t-test).

## 3.3 Identification of TEX264

p97 is recruited to its substrate by its cofactors, which directly bind to p97 via conserved p97 interaction motifs, and typically recognise ubiquitin chains on substrates (Meyer et al., 2012). We therefore sought to identify the cofactor which targets p97 to TOP1ccs. An ongoing aim of the laboratory has been to characterise the p97 interactome before and after DNA damage induction. To this end, a Stable Isotope Labelling by/with Amino acids in Cell culture (SILAC)-mass spectrometry was performed in stable HEK293 cells expressing doxycycline-inducible myc-Strep-tagged $p97^{EQ}$ (Fig. 15A). Cells were left untreated or irradiated at 10 Gy and allowed to recover for either 30 minutes or 8 hours. Cytoplasmic, nuclear, and chromatin fractions were isolated from these cells and anti-Strep immunoprecipitates prepared from each fraction were analysed by SILAC-MS.

Numerous ER-related p97 cofactors were found to interact with p97 in the cytoplasm, including Ufd1-Npl4, UBXD8, UBXD9, and AMFR. The Ufd1-Npl4 heterodimer also interacted with p97 inside the nucleus and its interaction with p97 was slightly enhanced upon irradiation (Fig. 15B). In addition to detecting known p97 cofactors, the SILAC-MS analysis also detected an increase in the interaction between p97 and ubiquitin as well as potential substrates of p97 after irradiation. Among the DNA repair factors that exhibited an increased interaction with p97 upon irradiation were the Ku complex and the MRN complex component Rad50. An independent study has since demonstrated that p97 extracts the Ku complex from DSB ends to facilitate their repair by HR, supporting the validity of our SILAC-MS analysis (van den Boom et al., 2016).

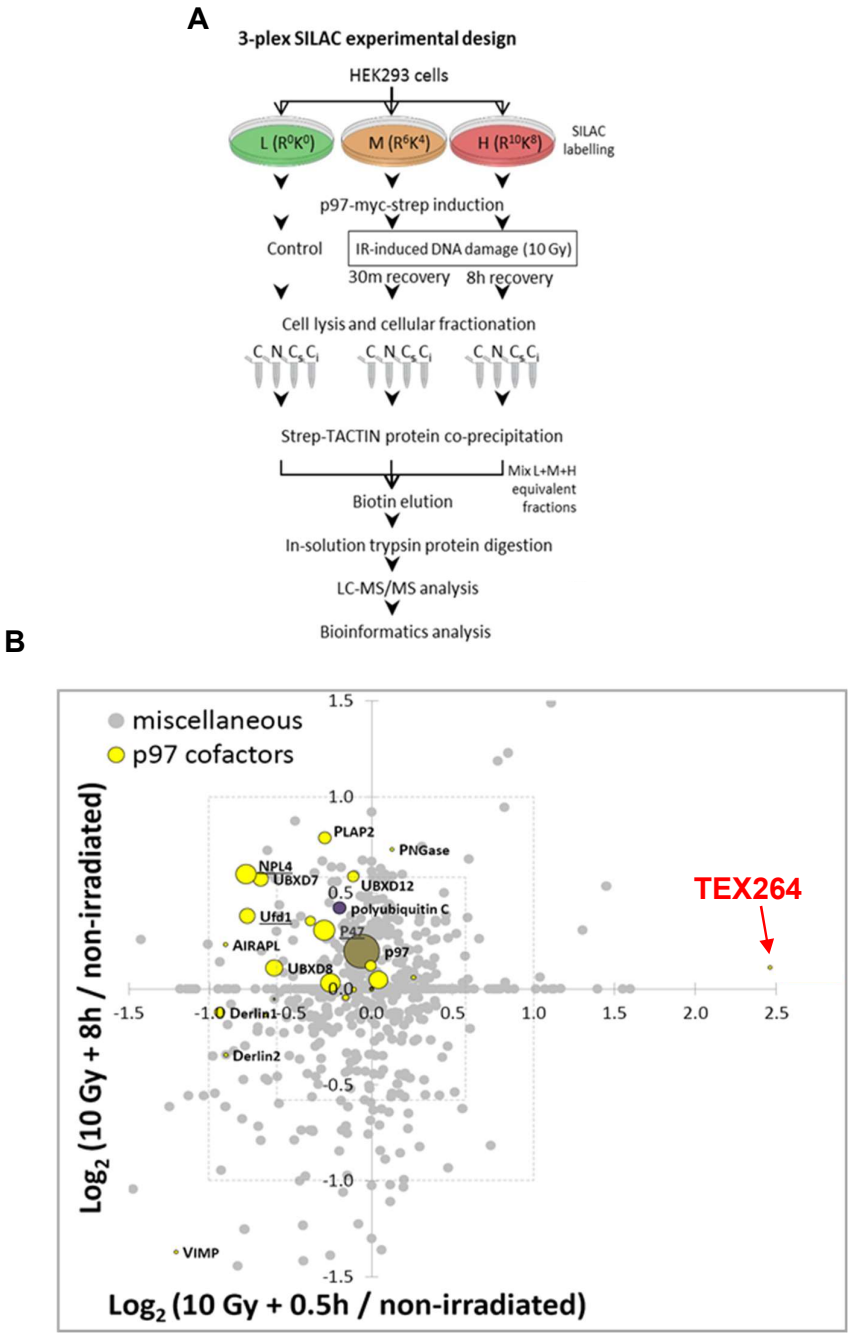



**Figure 15. Identification of TEX264.**

(A) Schematic of SILAC-MS approach. Stable, doxycycline-inducible HEK293 Flp-In T-Rex cells expressing $p97^{EQ}$-myc-Strep were grown in medium containing light (L), medium (M), or heavy (H) amino acids. Cells were exposed to 10 Gy ionising radiation (IR) and allowed to recover for either 30 mins or 8 hrs. Cells were then fractionated. Corresponding light, medium, and heavy fractions were pooled. Samples were subjected to immunoprecipitation with anti-Strep-tag beads and analysed by mass spectrometry. (B) Graphical representation of the p97 nuclear interactome 30 mins (x axis) and 8 hrs (y axis) after IR.

A protein which stood out was the previously uncharacterised protein, Testes-expressed 264 (TEX264). An interaction between p97 and TEX264 was only detected in the nucleus and was ~2.5-fold enhanced 30 minutes after IR (Fig. 15B). TEX264 is a gyrase inhibitory-like protein (Gyrl-like) (Fig. 16A). In *E. coli*, Gyrl proteins inhibit the activity of the bacterial topoisomerase, DNA gyrase, by preventing its binding to DNA (Nakanishi et al., 1998; Sengupta and Nagaraja, 2008). This led us to speculate that TEX264 may bind human topoisomerases. A bioinformatic analysis of the TEX264 protein sequence identified a potential SHP box, located in its unstructured C-terminus (Fig. 16A). The SHP box is a p97-interaction motif possessed by numerous p97 cofactors which enables them to bind directly to the N-terminus of p97. SHP boxes are defined by the consensus sequence FxGxGx2h (x, any amino acid; h, hydrophobic amino acid) and are usually found in the unstructured C-terminal tails of p97 cofactors (Hänzelmann and Schindelin, 2016). This raised the possibility that TEX264 might be a p97 cofactor. TEX264 also possesses an N-terminal leucine-rich repeat (LRR), which has since been shown to be a single-pass transmembrane domain (An et al., 2019). Homologs of TEX264 are present in vertebrates, including teleost fish such as *Danio rerio*, but are absent in established model organisms such as *Saccharomyces cerevisiae*, *Schizosaccharomyces Pombe,* and *Caenorhabditis elegans* (Fig. 16B).

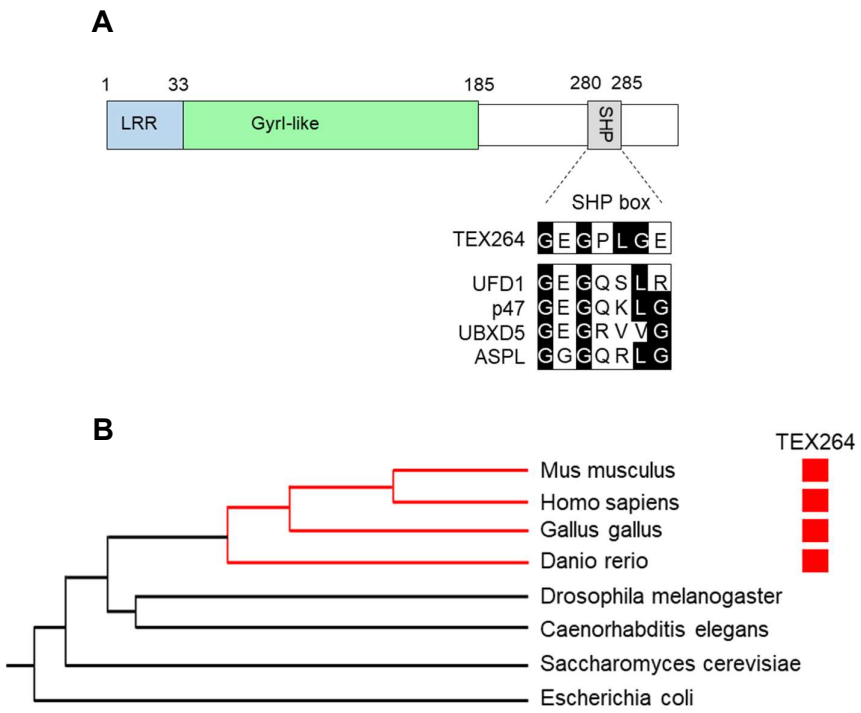



**Figure 16. TEX264 is a vertebrate-specific Gyrase inhibitory-like protein.**

(A) Above: schematic diagram of TEX264. LRR denotes leucine-rich repeat; Gyrl-like, Gyrase inhibitory-like domain; SHP, SHP box. Below: sequence alignment of TEX264$^{280-286}$ with the SHP box of other p97 cofactors. Conserved residues are highlighted in black. (B) Taxonomic tree showing the presence of TEX264 homologs across 8 selected species.

## 3.4 The role of TEX264 in TOP1cc repair

### 3.4.1 TEX264 physically bridges p97 and TOP1

We first used biochemical fractionations to assess the subcellular localisation of TEX264 in HEK293 cells. We found that TEX264 is present in the cytoplasm, nucleus, and on chromatin (Fig. 17A). HEK293 cells in which the *TEX264* gene was disrupted by CRISPR-Cas9 genome editing (ΔTEX264) were used to control for the specificity of the anti-TEX264 antibody.

We then monitored the expression of TEX264 throughout the cell cycle using the T24 urinary bladder carcinoma cell line. T24 cells arrest in G0 at 100% confluency and enter the cell cycle when passaged to lower densities. The expression of TEX264 and cell cycle markers was monitored by immunoblotting. We found that TEX264 is expressed throughout the cell cycle (Fig. 17B).

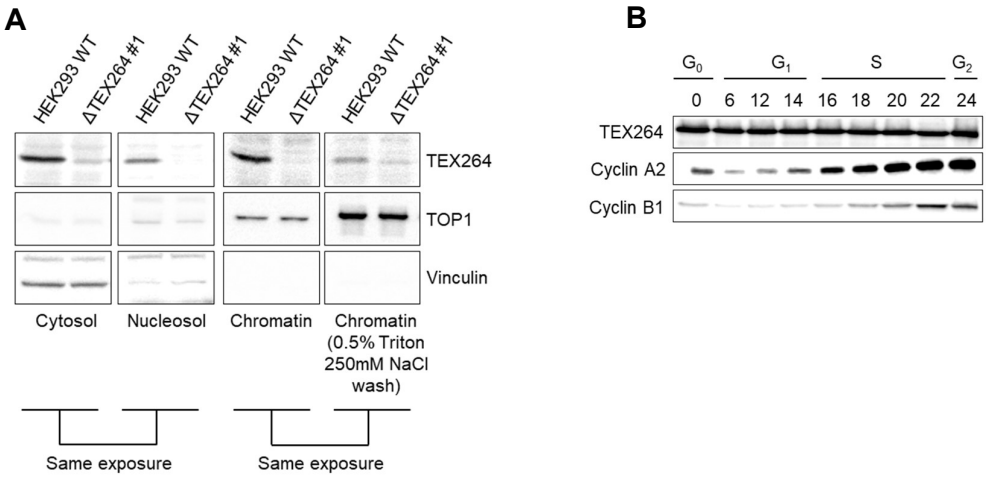



**Figure 17. TEX264 subcellular localisation and expression through the cell cycle.**

(A) Cytosolic, nuclear soluble (nucleosol), and chromatin fractions prepared from WT and *TEX264* knockout (ΔTEX264) HEK293 cells were immunoblotted with the indicated antibodies. Chromatin was washed with 0.5% Triton X-100 and 250 mM NaCl to separate loosely-bound and tightly-bound protein fractions. (B) Immunoblots of WCEs prepared from T24 cells. T24 cells were arrested in G0 by contact inhibition and passaged to lower densities for the indicated numbers of hours (hrs).

We next used co-immunoprecipitation experiments to assess whether TEX264 forms a complex with TOP1 and p97 in cells. Endogenous TEX264 and TOP1 were detected in anti-Strep immunoprecipitates prepared from stable cells expressing either doxycycline-inducible $p97^{WT}$- or $p97^{EQ}$-myc-Strep (Fig. 18A). Endogenous TOP1 and p97 also co-immunoprecipitated with FLAG-tagged TEX264 (Fig. 18B). Overall this indicates that these proteins form a functional complex within cells. In response to treatment with a low dose of CPT, the interaction between TEX264 and TOP1 was strongly enhanced, suggesting that TEX264 is recruited to TOP1ccs, along with p97 (Fig. 18B).

If TEX264 is responsible for recruiting p97 to TOP1, we would expect that p97 would no longer interact with TOP1 in the absence of TEX264. To test this possibility, we co-immunoprecipitated myc-Strep-tagged p97 from wild-type and ΔTEX264 HEK293 cells. In wild-type cells, p97 interacts with both TEX264 and TOP1. However, in ΔTEX264 cells, the interaction between p97 and TOP1 was almost completely abolished (Fig. 18C). Thus, TEX264 is required to recruit p97 to TOP1.

We noted that the interaction between p97 and TOP1, unlike the interaction between TEX264 and TOP1, was not notably enhanced by CPT treatment. This is possibly because p97-cofactor-substrate interactions are very dynamic, and often transient, partly because p97 processes substrates very rapidly (Papadopoulos et al., 2017; Xue et al., 2016). We therefore pre-treated cells with CB-5083 before CPT to slow the rate at which p97 processes TOP1ccs. We observed that the interaction between p97 and TOP1 was enhanced upon CPT treatment only when p97's ATPase activity was inhibited (Fig. 18D). This

interaction was also dependent on TEX264, demonstrating that TEX264 recruits p97 to TOP1ccs.

It was noteworthy that TEX264 and p97 interact with TOP1 in untreated cells and in those treated with a low dose of CPT (25 – 100 nM), but that neither interacts with TOP1 in cell treated with a high dose of CPT (1 μM). High and low doses of CPT elicit drastically different cellular responses. For example, high doses induce DSBs and replication fork collapse, whereas low doses do not (Zellweger et al., 2015). This was confirmed by the fact that treatment with 1 μM CPT, but not 25 or 100 nM, induced the appearance of ubiquitinated forms of γH2AX, which are generated by the E3 ubiquitin ligases, RNF8 and RNF168, upon DSB formation (Fig. 18E) (Pinder et al., 2013). This suggested that p97 and TEX264 may only interact with TOP1 when replication complexes are stable, as is explored later.

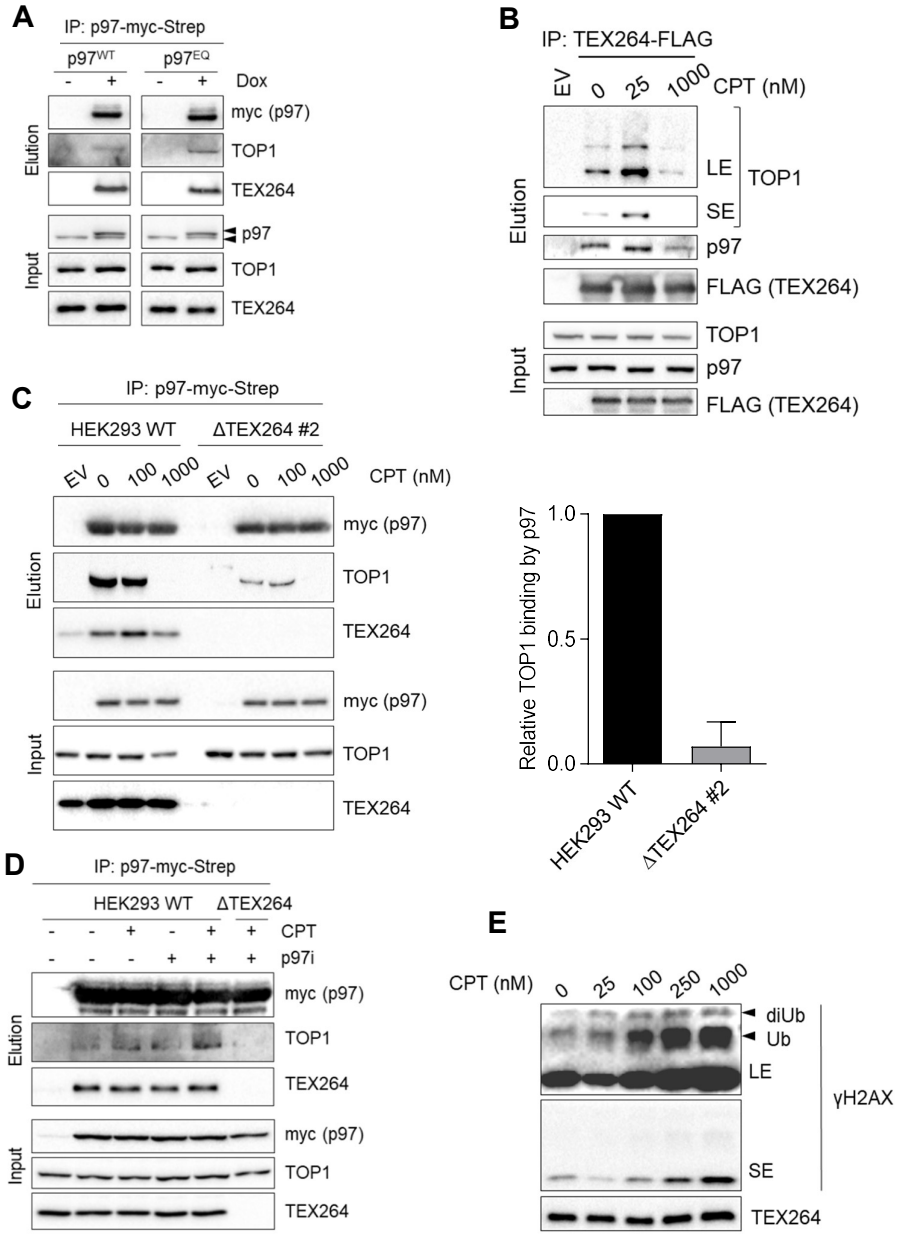



**Figure 18. TEX264 bridges p97 and TOP1.**

(A) Immunoblots of anti-Strep-tag immunoprecipitates prepared from doxycycline (Dox)-inducible HEK293 cells expressing p97$^{WT}$- or $^{EQ}$-myc-Strep. Arrowheads indicate endogenous p97 (lower band) and p97-myc-Strep (upper band). (B) Analysis of anti-FLAG immunoprecipitates prepared from HEK293 cells transfected with cDNA encoding TEX264-FLAG, treated with the indicated doses of CPT for 1 hour. (C) Left: immunoblots of anti-Strep-tag immunoprecipitates prepared from wild-type (WT) or ΔTEX264 HEK293 cells expressing p97-Strep-Myc. Right: corresponding quantification (mean ± SEM; n = 3; $^{**}P < 0.01$; Student's t-test). (D) Immunoblots of anti-Strep immunoprecipitates prepared from WT or ΔTEX264 HEK293 cells expressing p97-Strep-Myc. Cells were treated with the p97 inhibitor (p97i) CB-5083 (10 μM), for 90 mins, CPT (50 nM) for 60 mins, or both, as indicated. (E) Immunoblots prepared from HEK293 cells treated with indicated doses of CPT for 1 hr. LE, long exposure; SE, short exposure.

To address whether these interactions are direct, human TEX264 and p97 were purified from *E. coli* and recombinant TOP1 purified from insect cells was purchased. We could not purify full-length TEX264 as it was not soluble, however, this was overcome by deleting TEX264's hydrophobic N-terminal LRR (amino acids 1-33). TEX264 could directly interact TOP1, whereas a direct interaction between p97 was either weak or not detected (Fig. 19A-D). When TEX264 was pre-incubated with p97 before adding TOP1 to the reaction, the amount of TOP1 that co-immunoprecipitated with p97 was significantly higher (Fig. 19C, D). These results confirm that TEX264 can interact directly and simultaneously with both p97 and TOP1.

TEX264 interacts with the N-terminus of p97 (amino acids 1-199), meaning the p97-TEX264 interaction is not affected by the nucleotide-bound status of p97, as is the case with most p97 cofactors (Fig. 19E) (Buchberger et al., 2015). To address if the p97-TEX264 interaction is mediated by TEX264's putative SHP box, a TEX264 variant lacking its SHP box (amino acids 273-285) was purified from bacteria. This variant bound much less readily to p97 than wild-type TEX264 (Fig. 19F). Overall, we concluded that TEX264 binds to the N-terminus of p97 via its SHP box, bridges p97 to TOP1 *in vitro* and *in vivo*, and recruits p97 to TOP1ccs.

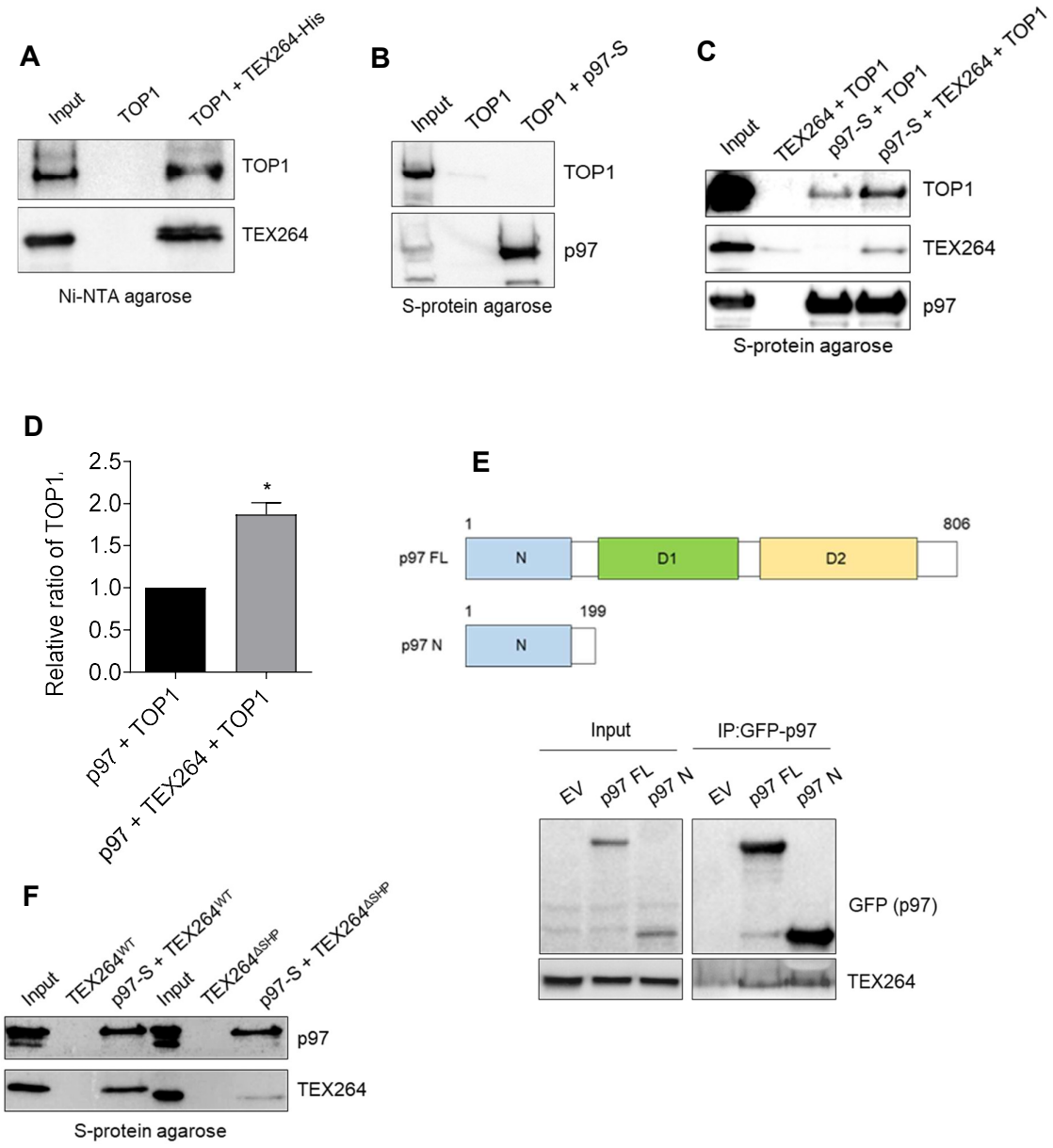



**Figure 19. TEX264 binds directly to the N terminus of p97 via its SHP box and bridges p97 to TOP1 *in vitro.***

(A) *In vitro* interaction (pull-down) of recombinant TOP1 incubated with or without His-tagged TEX264 (amino acids 34-313). (B) *In vitro* pull-down of recombinant TOP1 incubated with or without S-tagged p97. (C) Pull-down of recombinant p97-S after incubation with TOP1, with or without TEX264. (D) Quantification of C (error bars represent mean ± SD; $n = 2$; $^*P < 0.05$; Student's t-test). (E) Above: schematic diagrams of p97 protein constructs used below. Below: immunoblots of anti-GFP immunoprecipitates prepared from HEK293 cells expressing the indicated GFP-tagged p97 constructs. (F) *In vitro* p97 pulldown experiments after incubation of recombinant p97-S tag with His-tagged $TEX264^{WT}$ or $TEX264^{\Delta SHP}$. Figs. 19A-C & 19F generated by Katherine Wiseman.

### 3.4.2 TEX264 counteracts TOP1ccs

Since p97 is important for TOP1cc repair and is recruited to TOP1ccs by TEX264, TOP1ccs should accumulate in cells when TEX264 function is disrupted. Indeed, TEX264 depletion resulted in elevated numbers of TOP1cc foci in RPE-1 cells (Fig. 20A-C). Moreover, in either of two ΔTEX264 HEK293 clones, basal TOP1cc levels were significantly higher (Fig. 20D-F). This increase could be reversed by exogenous TEX264 but not when ΔTEX264 were depleted of p97 (Fig. 20G, H). Thus, TEX264 counteracts TOP1ccs in a p97-dependent manner. When cells were treated with a low dose of CPT and allowed to recover in CPT-free media, TOP1ccs persisted far longer in ΔTEX264 cells than in control cells (Fig. 20I, J). These data demonstrate that TEX264 promotes TOP1cc repair.

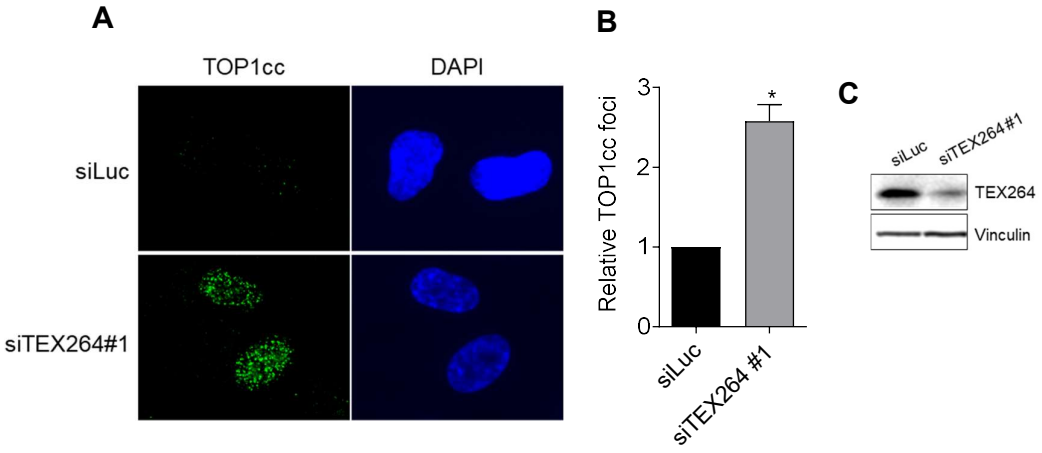



**Figure 20. TEX264 counteracts TOP1cc formation.**

(A) Immunofluorescent detection of TOP1ccs using a TOP1cc-specific antibody in RPE-1 cells transfected with the indicated siRNAs. Scale bar, 10 μm. (B) Quantification of A (error bars represent mean ± SD; $n = 2$; $^*P < 0.05$; Student's t-test). (C) Immunoblots confirming efficacy of TEX264 depletion. Figure continued on page 89.

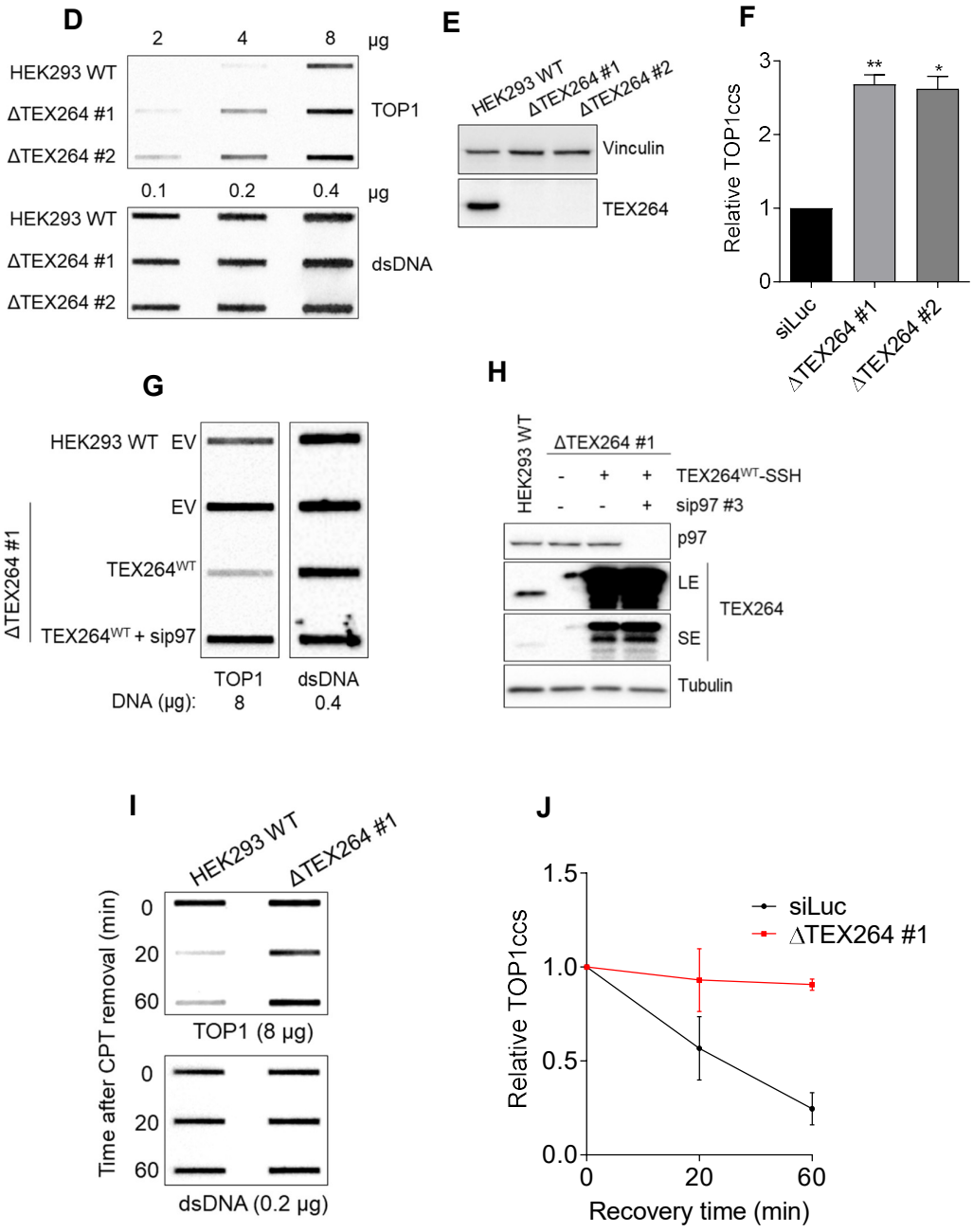



**Figure 20 continued. TEX264 counteracts TOP1cc formation.**

(D) RADAR analysis of TOP1ccs isolated from WT or ΔTEX264 HEK293 cells. (E) Immunoblot demonstrating absence of TEX264 protein expression in ΔTEX264 cells. (F) Quantification of D. (error bars represent mean ± SEM; n = 2; $^{**}P < 0.01$, $^{*}P < 0.05$; Student's t-test). (G) RADAR analysis of TOP1ccs isolated from WT or ΔTEX264 HEK293 cells, transfected with the indicated cDNAs and/or siRNA. (H) Control immunoblot for G. LE, long exposure; SE, short exposure. (I) RADAR analysis of TOP1ccs isolated from WT or ΔTEX264 HEK293 cells treated for 1 hr with CPT (50 nM), then released into normal media for 20 or 60 mins. (J) Quantification of I (error bars represent mean ± SEM; n = 3).

TOP1ccs result in ssDNA breaks that can also be converted into DSBs during DNA replication (Strumberg et al., 2000). To test if elevated TOP1ccs levels lead to DNA strand break accumulation in TEX264-depleted cells, we used an alkaline comet assay. This assay is a highly sensitive method for detecting both ss- and ds-DNA breaks (Singh et al., 1988). RPE-1 cells were transfected with siRNA targeting TEX264, treated with a low dose of CPT (100 nM) for 1 hour, and either collected or allowed to recover for 3 or 6 hours in normal media. A continuous 6 hour-long treatment with CPT was also included. TDP1-depleted cells were used as a positive control of defective DNA damage repair after CPT treatment. In unchallenged conditions, the levels of DNA damage in TEX264-depleted cells were significantly increased (Fig. 21A). Treatment with CPT induced DNA damage in all samples (Fig. 21A, B). In control cells, this damage was fully repaired after 6 hours of recovery, whereas it remained significantly elevated in TEX264- and TDP1-depleted cells (Fig. 21A, B).

Strikingly, the DNA damage repair kinetics were not significantly different between TEX264- and TDP1-depleted cells (Fig. 21A). When these two proteins were co-depleted, CPT-induced DNA damage was not significantly higher, nor were the DNA damage kinetics further delayed, compared with either of the single depletions (Fig. 21A). This strongly indicates that TDP1 and TEX264 act in the same TOP1cc repair pathway.

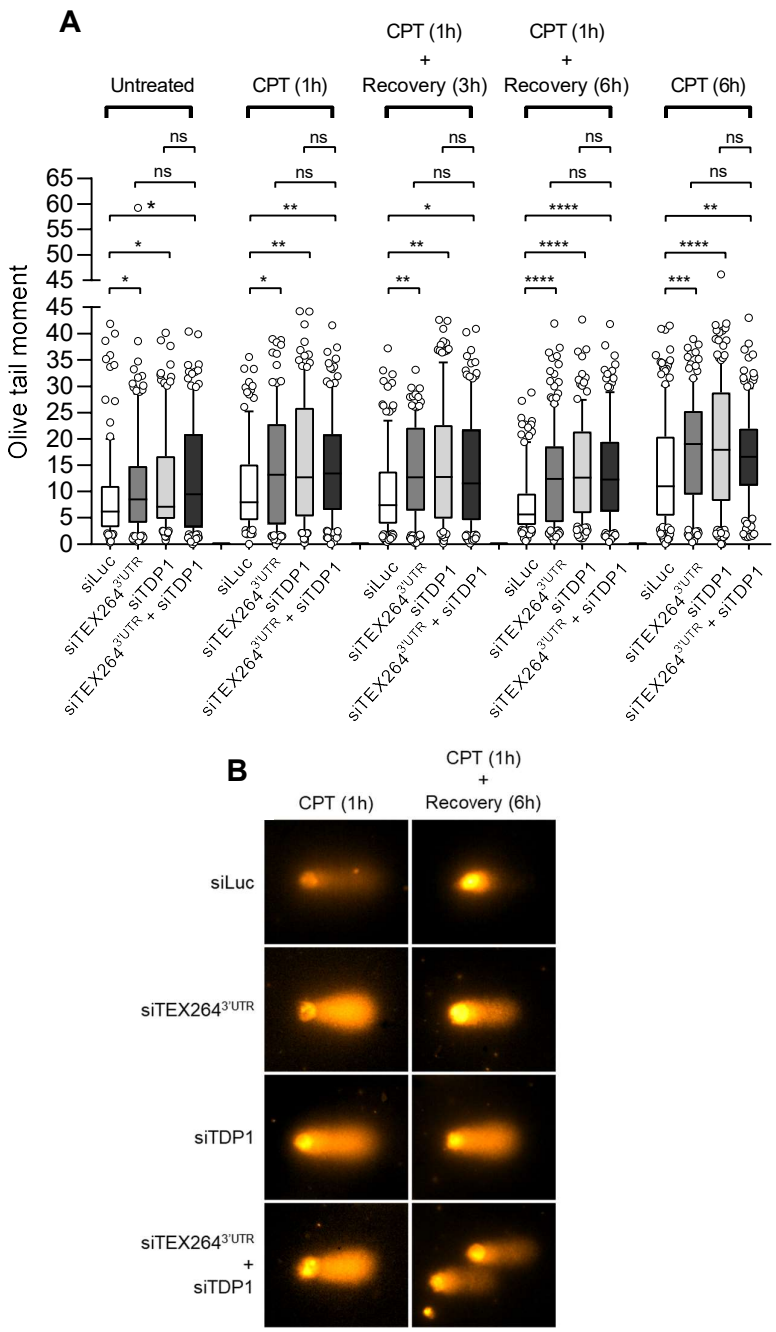



**Figure 21. TEX264- and TDP1-deficient cells exhibit delayed DNA damage repair kinetics after CPT treatment.**

(A) Alkaline comet assay performed in RPE-1 cells treated with the indicated combinations of siRNAs. Cells were treated with CPT (100 nM) for 1 or 6 hr(s) and allowed to recover in normal media where indicated. Quantification of >100 cells per condition from one representative experiment is shown (n = 3). Whisker box plots show median values and data within the 10–90 percentile. Statistical analysis was performed using Kruskal-Wallis ANOVA with multiple comparisons, with Benjamini Hochberg post-test. (B) Representative images of cells analysed by alkaline comet assay in A. Experiments were performed by Samuel Hume.

To further test this possibility, we monitored TOP1cc levels by RADAR. TOP1ccs accumulated to the same extent in cells that were co-depleted of p97, TEX264, or TDP1 in pairs as they did when these proteins were depleted individually (Fig. 22A-C). We next assessed if this TOP1cc repair defect resulted in cellular sensitivity to CPT using a colony forming assay. TEX264- and TDP1-depleted HeLa cells were sensitive to low doses of CPT and this was not further enhanced by their co-depletion (Fig. 22D). ΔTEX264 HeLa cells exhibited a similar degree of sensitivity to CPT as TEX264-depleted HeLa cells (Fig. 22E). The sensitivity of cells depleted of TEX264 with an siRNA sequence targeting its 3'UTR (untranslated region) could be rescued by exogenously-expressed TEX264 but not TDP1 (Fig. 22F). This indicates that TDP1 alone is not sufficient for TOP1cc repair and suggests that it requires TEX264 to fulfil its role in TOP1cc repair.

Compared to *TDP1*-knockout cells, TEX264-depleted cells were much less sensitive to micromolar doses of CPT (Fig. 22G). This fits with our observation that TEX264 does not interact with TOP1 at such high doses and supports the conclusions that TEX264 acts when DNA replication complexes are intact and that TEX264 does not regulate DSB repair.

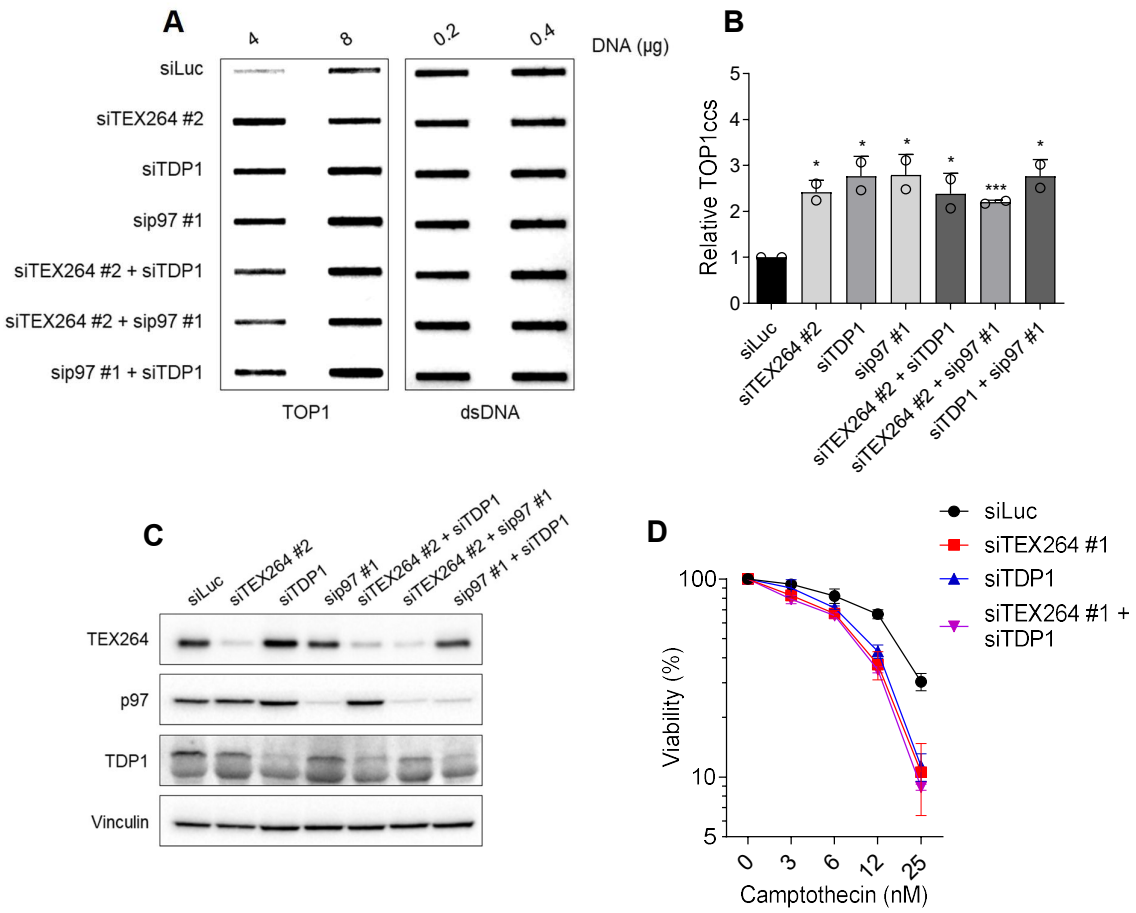



**Figure 22. TEX264, p97, and TDP1 act in the same TOP1cc repair pathway.**

(A) RADAR analysis of TOP1ccs isolated from HEK293 cell treated with the indicated siRNAs. (B) Quantification of A (error bars represent mean ± SD; $n = 2$; *$P < 0.05$; ***$P < 0.001$; Student's t-test). (C) Immunoblots to confirm efficacy of depletions for experiment in A. (D) Colony forming assay in HeLa cells transfected with the indicated siRNAs. Cells were treated with the indicated doses of CPT for 24 hrs and allowed to recover in normal media for 7 days. Viability is expressed as a percentage of the corresponding untreated cells. Figure continued on page 94.

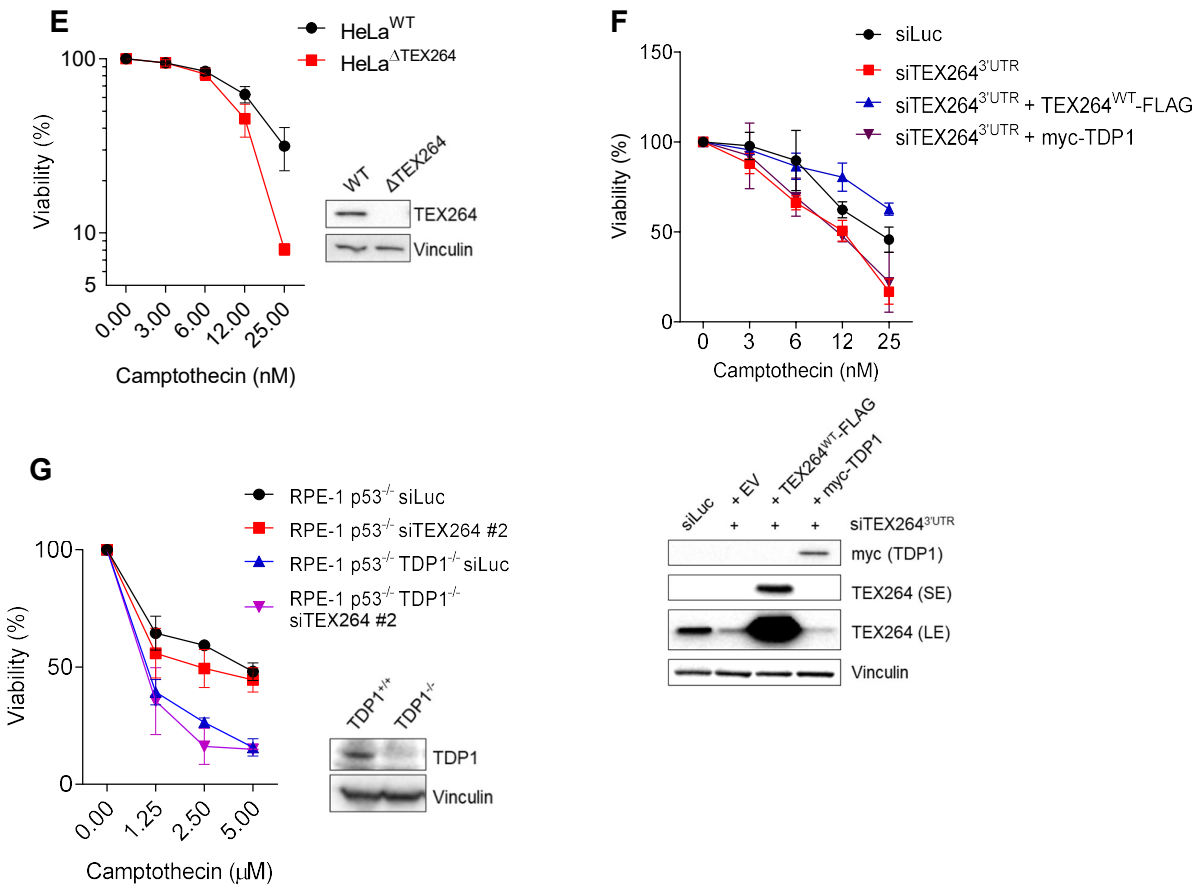



**Figure 22 continued. TEX264, p97, and TDP1 act in the same TOP1cc repair pathway**.

(E) Left panel: colony forming assay in the indicated HeLa cell lines. Cells were treated with indicated doses of CPT for 24 hrs. Right panel: corresponding immunoblots. (F) Left panel: colony forming assay in HeLa transfected with the indicated siRNAs and cDNAs. Cells were treated with indicated doses of CPT for 24 hrs. Right panel: corresponding immunoblots. (G) Left panel: colony forming assay using RPE-1 $p53^{-/-}$ or RPE-1 $p53^{-/-}$ $TDP1^{-/-}$ cells, transfected with the indicated siRNAs. Cells were treated with the indicated doses of CPT for 1 hr. Right panel: corresponding immunoblot confirming TDP1 knockout.

### 3.4.3 The GyrI-like domain of TEX264 binds TOP1

We next focused on addressing how TEX264 recognises the TOP1 protein. The crystal structure of the *E. coli* GyrI protein, sbmC, shows that the proteins forms a solvent-exposed surface, which may mediate protein-protein interactions (Romanowski et al., 2002). We generated a crystallographic-based homology structural model of the GyrI-like domain of TEX264 (residues 43-193) in Phyre2 using sbmC and other GyrI-like proteins as templates (Fig. 23A). We identified residues on the predicted surface of the GyrI-like domain of TEX264 (highlighted in pink in Fig. 23A). In addition, residues 193-210, which are immediately adjacent to TEX264's GyrI-like domain, are predicted to form an α-helix that appears to be unique among GyrI-like proteins (Paul Elliott, personal communication). We generated a series of TEX264 variants with single point mutations between residues 154 and 199 and monitored their interactions with TOP1 by co-immunoprecipitation. Each variant bound TOP1 less efficiently than wild-type TEX264 ($TEX264^{WT}$), indicating that these residues comprise a binding surface which enables TEX264 to bind TOP1 (Fig. 23B). Each variant retained their interaction with p97, indicating that the TEX264 protein structure was not significantly altered by any of these point mutations (Fig. 23B). We then assessed whether these mutations impact the ability of TEX264 to promote TOP1cc repair. In Human Bone Osteosarcoma Epithelial (U2OS) cells, TOP1cc foci accumulation resulting from TEX264 depletion was completely rescued by exogenous $TEX264^{WT}$ but not by $TEX264^{E194A}$ (Fig. 23C, D). This demonstrates that the interaction between TEX264 and TOP1 is important for its role in TOP1cc repair.

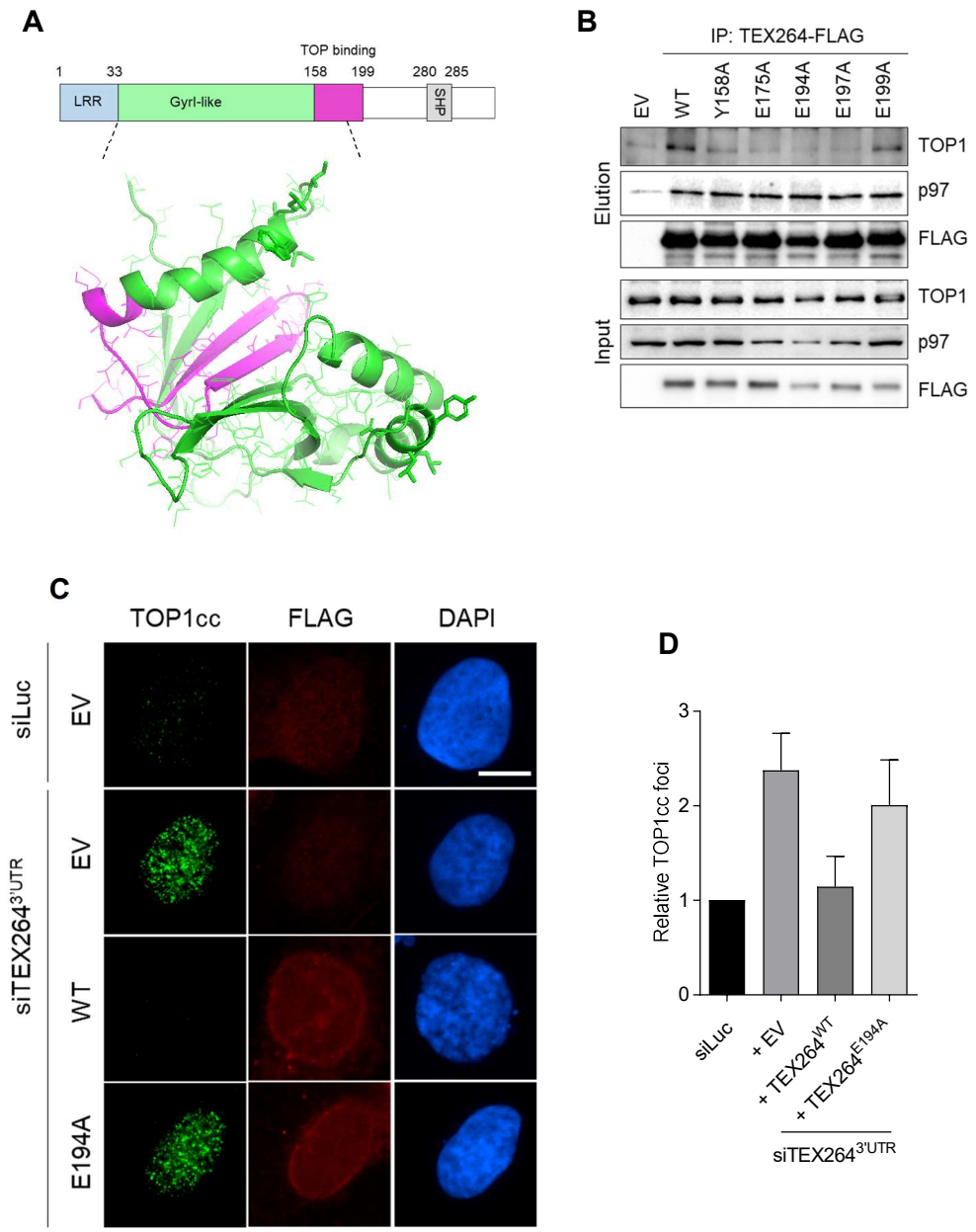



**Figure 23. Identification of TOP1-binding residues in TEX264.**

(A) Upper panel: schematic diagram of TEX264 protein with the location of TOP1-binding residues highlighted in pink. Lower panel: crystallography-based homology model of the Gyrl-like domain of TEX264. Residues 154 – 185 are highlighted in pink. (B) Immunoblots of anti-FLAG immunoprecipitates prepared from ΔTEX264 HEK293 cells expressing the indicated FLAG-tagged TEX264 variants. (C) Representative immunofluorescence images of U2OS cell nuclei. Cells were transfected with indicated siRNA and cDNA and stained with antibodies against TOP1ccs (green) and FLAG (red). Scale bar, 10 μm. (D) Quantification of C (error bars represent mean ± SEM, n = 3).

### 3.4.4 TEX264 binds SUMO1

Most p97 cofactors have motifs that enable them to bind ubiquitin chains. This enables them to recognise substrates that have been marked for p97-dependent turnover by ubiquitin ligases (Meyer et al., 2012). However, TEX264 does not appear to contain any canonical ubiquitin-binding motifs nor did we detect a direct interaction between TEX264 and ubiquitin chains (data not shown). The ubiquitin-like protein, SUMO1, has previously been implicated in TOP1cc repair (Mao et al., 2000a). Cdc48 and Ufd1 are also thought to regulate the SUMO-dependent repair of TOP1ccs in yeast (Nie et al., 2012). Bioinformatic analysis subsequently identified two potential SIMs in TEX264, both of which are in its GyrI-like domain (Fig. 24A). We therefore asked whether TEX264 is linked to SUMO1-mediated TOP1cc repair. We did not detect a direct interaction with either free SUMO2 or poly-SUMO2 chains (Fig. 24B). However, we found that TEX264 can bind to SUMO1 (Fig. 24C). Point mutating either of TEX264's SIMs reduced its binding to SUMO1, suggesting both SIMs contribute to SUMO1 binding to some extent (Fig. 24C).

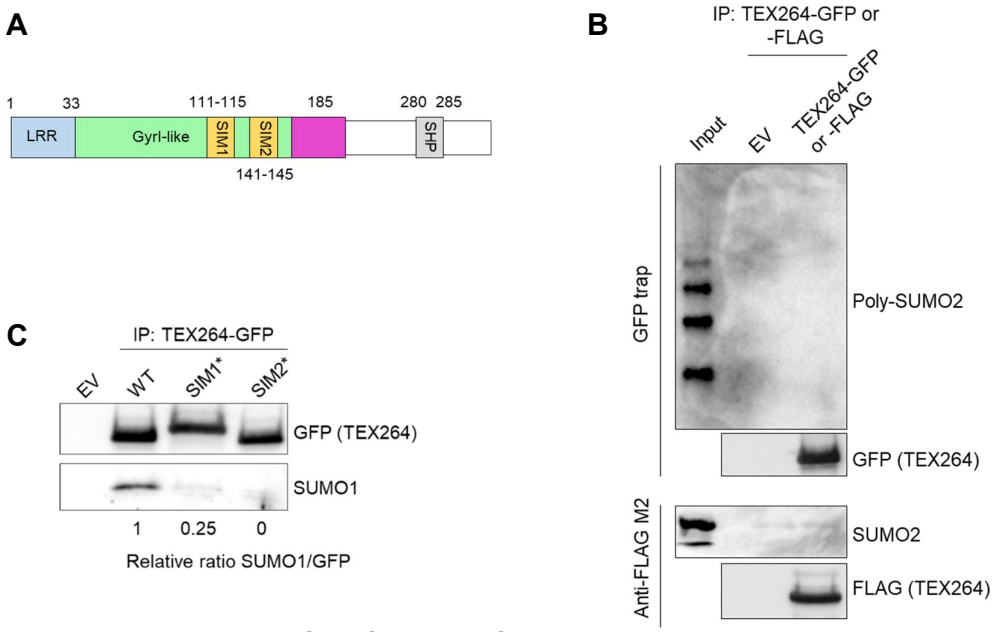



**Figure 24. TEX264 binds SUMO1 via its SIMs.**

(A) Schematic of TEX264 protein with location of putative SIMs highlighted in yellow. (B) *In vitro* SUMO binding assay. Immunoblots of GFP, FLAG, & SUMO2/3 after incubation of free SUMO2 or poly-SUMO2 chains with either GFP- or FLAG-tagged TEX264, purified from ΔTEX264 HEK293 cells. (C) Immunoblots of SUMO1 and GFP after incubation of GFP-tagged TEX264 variants purified from ΔTEX264 HEK293 with recombinant SUMO1.

When YFP-TOP1 was isolated by immunoprecipitation under denaturing conditions, we found that CPT treatment induced a dose-dependent increase in the amount of SUMO1-modified TOP1 (Fig. 25A). These results suggest that TOP1ccs are modified with SUMO1. To assess if SUMOylation promotes TOP1cc repair, we monitored TOP1cc accumulation in UBC9-depleted cells before and after CPT treatment. In unchallenged conditions, basal TOP1cc levels were lower in UBC9-depleted cells (Fig. 25B, C). This has been previously reported and likely reflects the pleiotropic roles of SUMOylation in regulating TOP1 function (Horie et al., 2002; Li et al., 2015). However, TOP1ccs were drastically higher in UBC9-depleted HEK293 cells after CPT treatment, demonstrating that SUMOylation is important for the repair of TOP1ccs (Fig. 25B, C). This increase was similar in extent to that observed in p97-depleted and ΔTEX264 cells (Fig. 25B, C). If TEX264 and p97 are involved in the repair of SUMO1-modified TOP1ccs, then we would expect SUMO1-modified TOP1 to accumulate when TEX264 and p97 are absent. As anticipated, depletion of TEX264 or p97 caused SUMO1-modified YFP-TOP1 to accumulate, even in unchallenged cells (Fig. 25D, E).

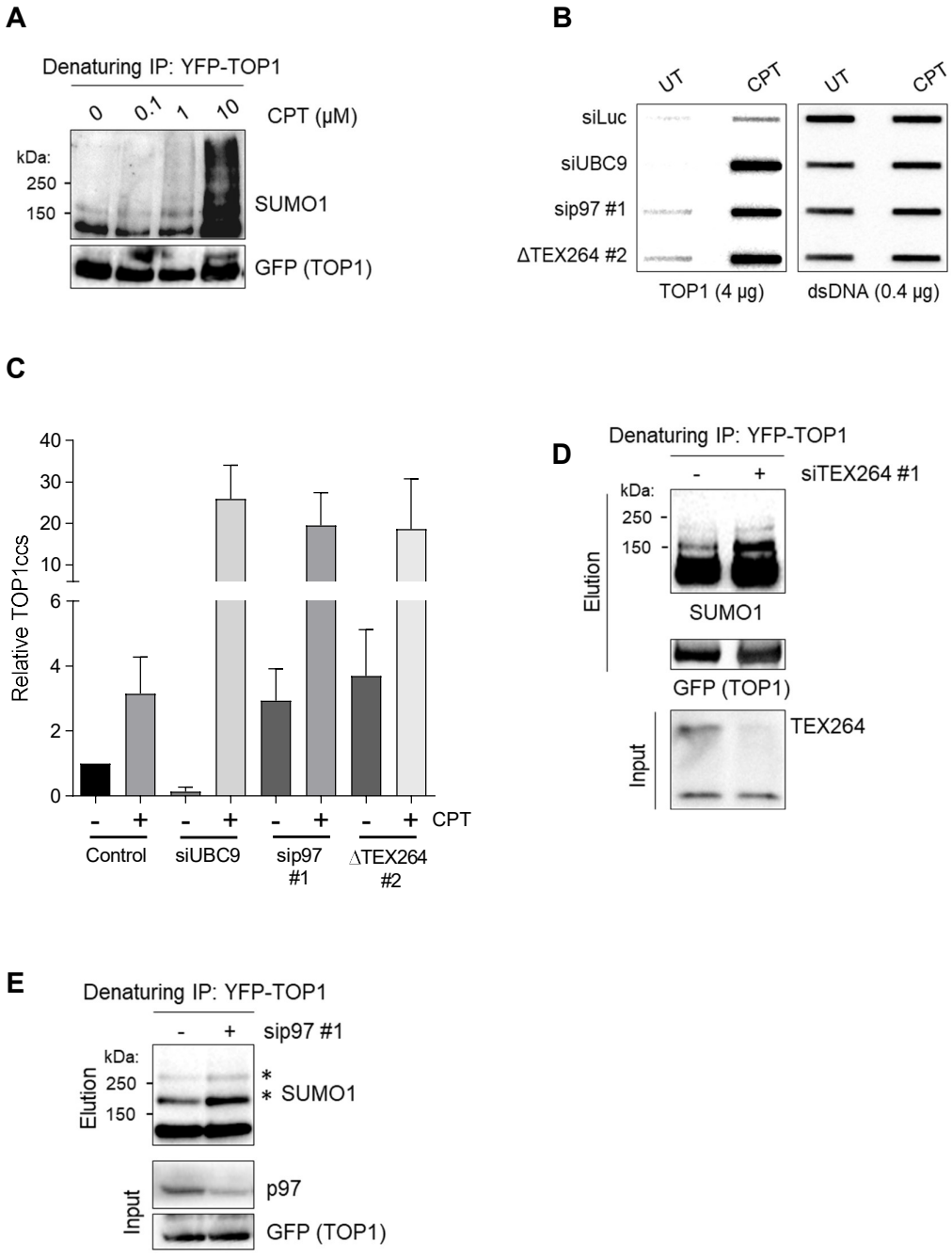



**Figure 25. SUMOylation promotes TOP1cc repair.**

(A) Immunoblots of SUMO1 and GFP. HEK293 cells expressing exogenous YFP-TOP1 were treated with the indicated doses of CPT for 1 hr. YFP-TOP1 was purified under denaturing conditions. (B) TOP1ccs analysed by RADAR from the HEK293 cells treated with indicated siRNAs Where indicated cells were treated with 1 μM CPT for 30 mins. (C) Quantification of B (error bars represent mean ± SD; *n* = 2). (D) and (E). Immunoblots of SUMO1 and GFP after denaturing immunoprecipitation of YFP-TOP1 from HEK293 cells transfected with the indicated siRNAs.

TEX264 can bind directly to unmodified TOP1 *in vitro* and co-immunoprecipitates with unmodified TOP1 in cells (Figs. 18C & 19A). However, we also observe that TEX264 co-immunoprecipitates with modified species of TOP1 (Fig. 26A). This interaction was substantially diminished by mutation of SIM2 (TEX264$^{SIM2*}$: I141A, S143A, W145A) and to a much lesser extent by mutation of SIM1 (TEX264$^{SIM1*}$: L111A, I112A, D113A, L114A, Y115A) (Fig. 26A). Using our structural model of TEX264 we found that SIM2 likely folds into closer proximity to the TOP1-binding region of TEX264 than SIM1 (Fig. 26B). This may explain why SIM2 appears to be more important than SIM1 for binding to modified TOP1.

To conclusively test if TEX264 binds SUMO1-modified TOP1 in a SIM2-dependent manner we performed tandem-affinity purifications (TAP). FLAG-tagged TEX264 variants and HA-tagged SUMO1 were co-expressed in HEK293 cells (Fig. 26C). We first purified TEX264-containing proteins complexes under native conditions by co-immunoprecipitation using anti-FLAG beads. To identify SUMO1-modified TEX264-interacting proteins, we denatured TEX264-containing protein complexes in 1% SDS with boiling (Fig. 26C). SUMO1-modified proteins that interact with TEX264-FLAG were then purified by immunoprecipitation with anti-HA beads. When this extract was immunoblotted with an anti-SUMO1 antibody we observed that TEX264 interacts with numerous SUMO1-modified proteins. TEX264$^{SIM2*}$ bound slightly less readily to SUMO1-modified proteins (Fig. 26C). When we probed for TOP1, we found that approximately 70% less SUMO1-modified TOP1 co-immunoprecipitated with TEX264$^{SIM2*}$ than TEX264$^{WT}$ (Fig. 26C). Thus, TEX264 can bind SUMO1-modified TOP1 in a manner that largely depends on SIM2. We speculate that the SUMOylation of TOP1ccs may act as a signal for the recruitment of TEX264. Alternatively, SUMO1 may stabilise

the binding between TEX264 and TOP1ccs. Formal proof of these hypotheses will require in-depth studies with purified proteins and would be aided by a crystal structure of TEX264 bound to SUMO1-modified TOP1. The interaction between TEX264 and SUMO1-modified TOP1 is clearly important for TOP1cc repair since TEX264$^{SIM2*}$ displayed a significantly reduced ability to reverse TOP1cc accumulation in TEX264-depleted cells (Fig. 27A, B).

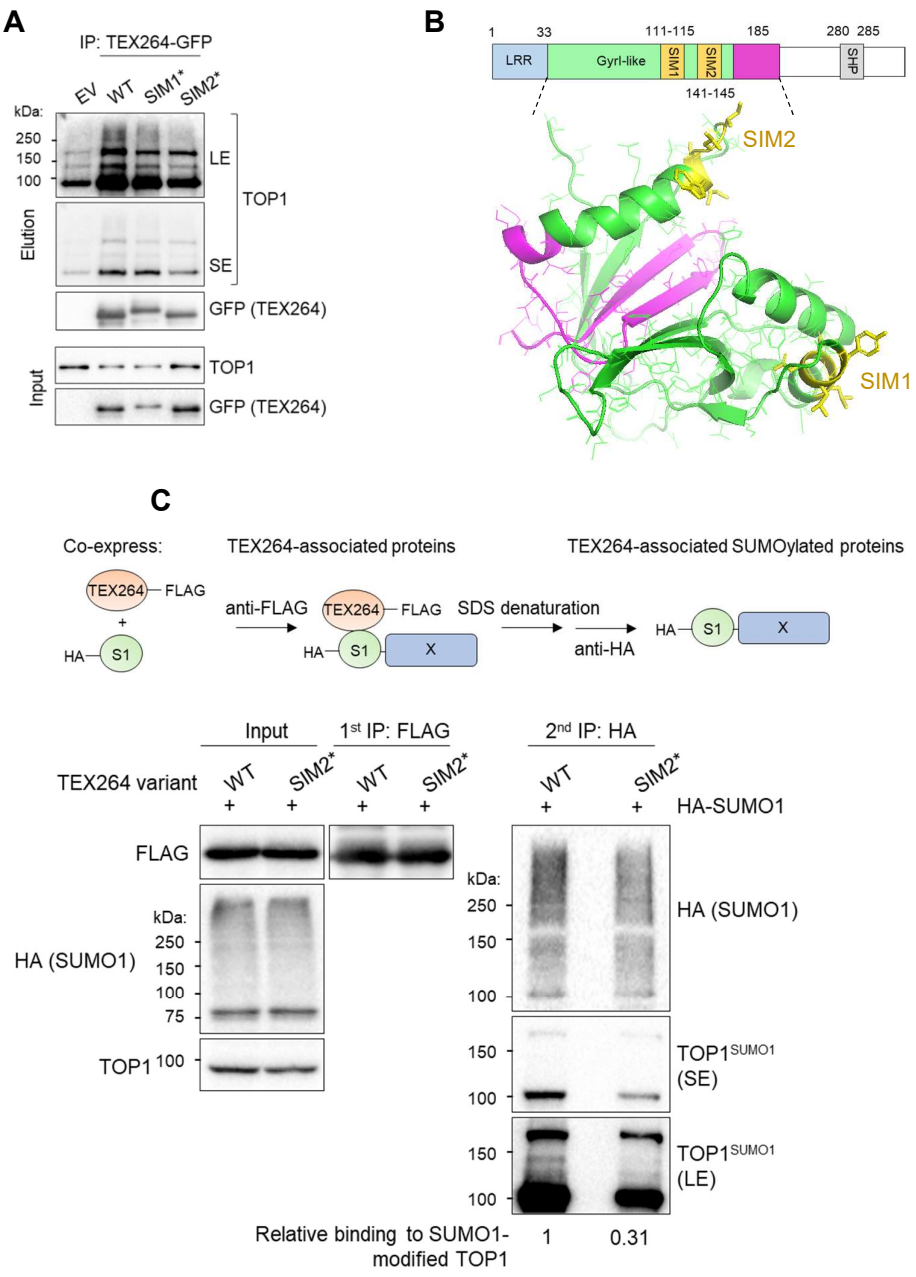



**Figure 26. TEX264 binds SUMO1-modified TOP1.**

(A) Immunoblots of anti-GFP immunoprecipitates prepared from ΔTEX264 HEK293 cells expressing indicated cDNAs. (B) Upper panel: schematic of TEX264 protein with location of putative SIMs highlighted in yellow. Lower panel: homology-based structural model of Gyrl-like domain of TEX264 with location of SIMs indicated. (C) Upper panel: schematic of tandem-affinity purification (TAP) procedure. Lower panel: Immunoblots of TAP extracts prepared from ΔTEX264 HEK293 cells expressing the indicated FLAG-tagged TEX264 variants and HA-tagged SUMO1.

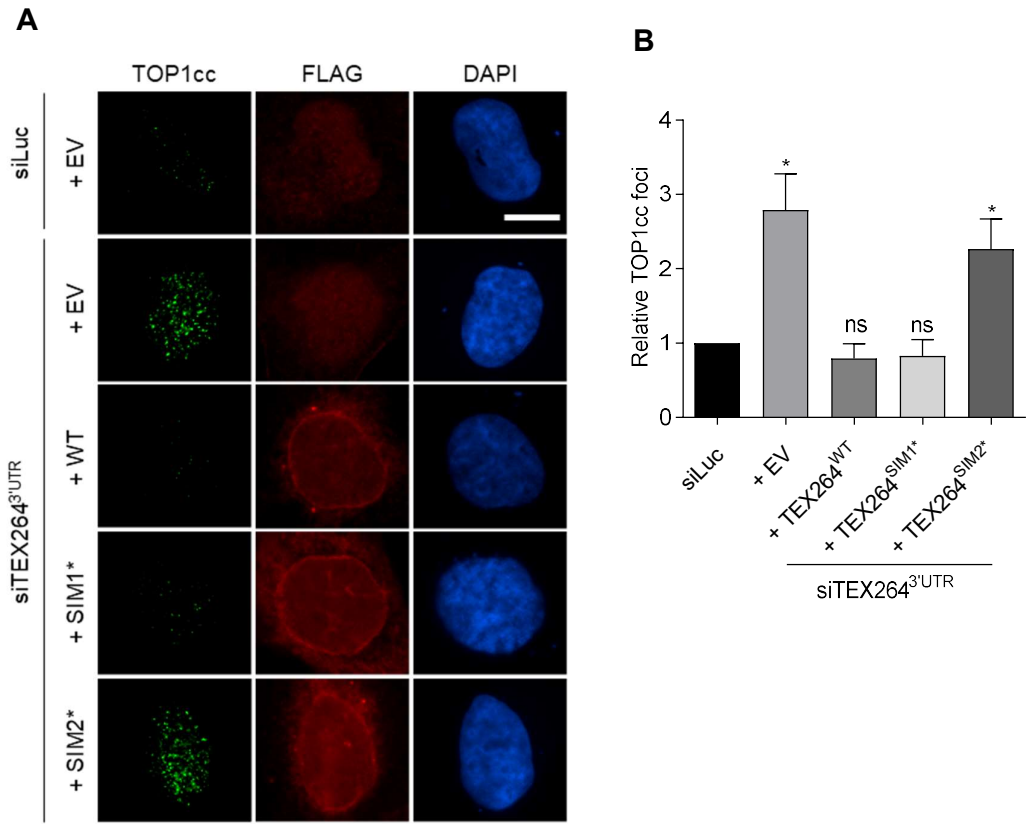



**Figure 27. SIM2 in TEX264 is required for TOP1cc repair.**

(A) Representative immunofluorescence images of U2OS cell nuclei. Cells were transfected with indicated siRNA and cDNA and stained with antibodies against TOP1ccs (green) and FLAG (red). Scale bar, 10 μm. (B) Quantification of A (error bars represent mean ± SEM, n = 3; **P* < 0.05; ns, not significant; Student's t-test).

### 3.4.5 TEX264 cooperates with SPRTN in TOP1cc repair

As demonstrated in Fig. 10B, SPRTN is another p97 cofactor involved in TOP1cc repair (Vaz et al., 2016). As p97 is a homohexameric complex, it can bind multiple cofactors simultaneously (Hänzelmann et al., 2011). We therefore asked whether TEX264 regulates SPRTN-dependent TOP1cc repair.

We first assessed whether these two proteins act within the same TOP1cc repair pathway by colony forming assay. SPRTN depletion sensitised HeLa cells to low doses of CPT, albeit to a lesser extent than TEX264 depletion (Fig. 28A, B). Importantly, there was no additive effect on cellular sensitivity to CPT when SPRTN and TEX264 were co-depleted (Fig. 28A, B). The likeliest interpretation of this result is that SPRTN's role in TOP1cc repair is dependent on TEX264 whereas TEX264 has SPRTN-independent roles in TOP1cc repair.

Given that TEX264 recruits p97 to TOP1, we hypothesised that TEX264 may recruit SPRTN-containing p97 sub-complexes to TOP1ccs. To address this, we assessed whether SPRTN's interaction with TOP1 was affected by TEX264 depletion. Endogenous TOP1, TEX264, and p97 all co-immunoprecipitate with Strep-Strep-HA (SSH)-tagged SPRTN, indicating that they form a functional complex in cells (Fig. 28C). TEX264 depletion almost completely abolished the interaction between SPRTN and TOP1 but did not affect SPRTN's interaction with p97 (Fig. 28C). This suggests that p97-SPRTN sub-complexes are recruited to TOP1 by TEX264.

A prediction of this model is that SPRTN would require its SHP box, and therefore its interaction with p97, to counteract TOP1ccs. To test this hypothesis, stable doxycycline-inducible HEK293 cells that mildly overexpress either wild-type

SPRTN ($SPRTN^{WT}$) or a p97 binding-defective variant ($SPRTN^{SHP*}$: G255A, G257A, L260A, G261A) were generated (Fig. 28D). Cells in which endogenous SPRTN was depleted with a sequence targeting its 3'UTR accumulated TOP1ccs (Fig. 28E, F). Expression of $SPRTN^{WT}$ could completely counteract this, whereas $SPRTN^{SHP*}$ failed to do so (Fig. 28E, F).

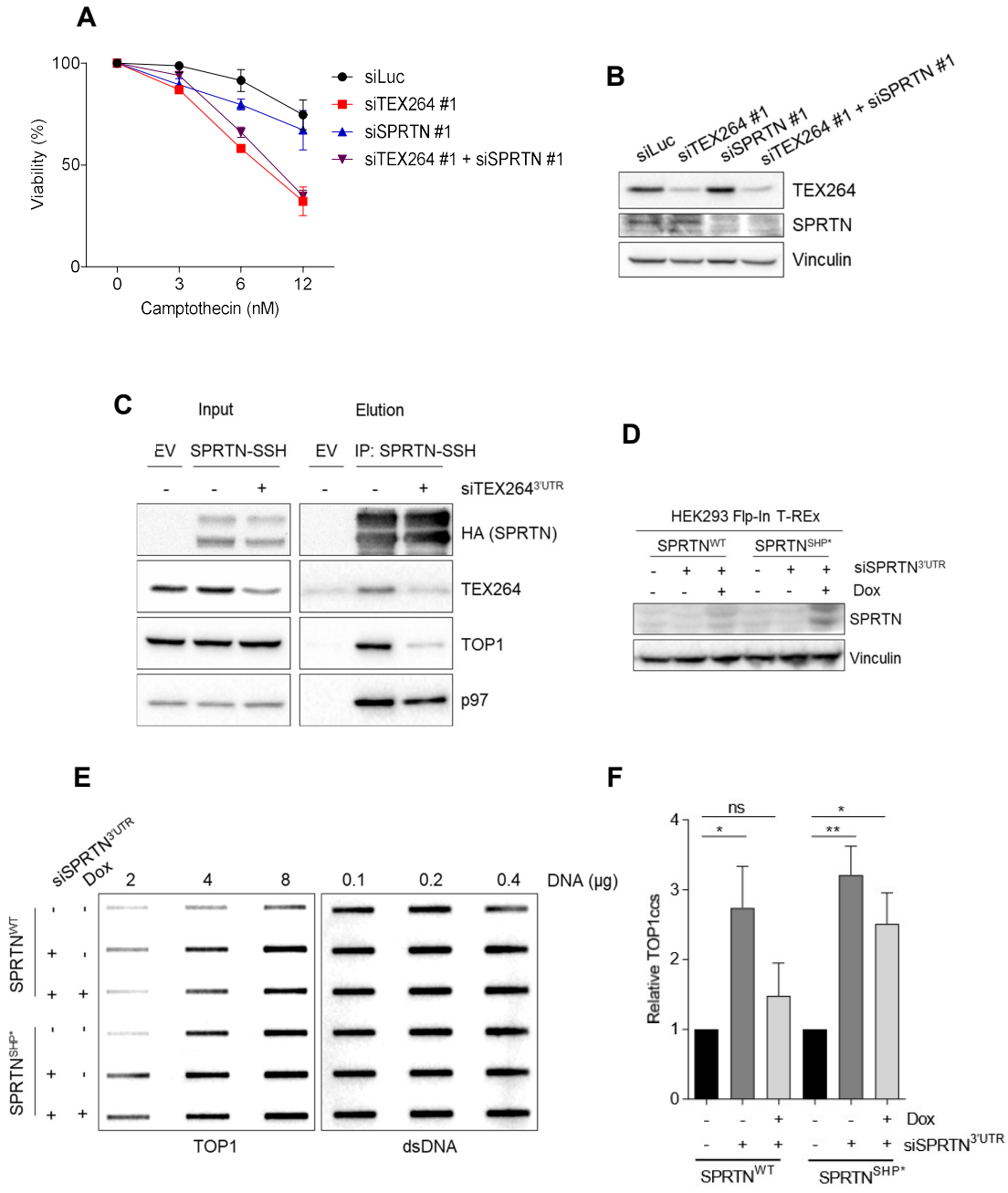



**Figure 28. TEX264 cooperates with SPRTN in TOP1cc repair.**

(A) Colony forming assay in HeLa cells transfected with the indicated siRNAs and treated with the indicated doses of CPT for 24 hrs. (B) Corresponding immunoblots for A. (C) Immunoblots of anti-Strep immunoprecipitates prepared from HEK293 cells transfected with SSH-tagged SPRTN and the indicated siRNAs. (D) control immunoblots for E and F. (E) RADAR analysis of TOP1ccs prepared from the indicated cells lines, transfected with the indicated siRNA. 'Dox' denotes where cells were treated with doxycycline for 24 hrs before harvesting. (F) Quantification of E (error bars represent mean ± SEM, n = 3, $^{**}P < 0.001$; $^{*}P < 0.05$; Student's t-test).

A recent structural analysis of SPRTN's SprT domain demonstrated that its protease active site is shielded within a narrow groove by SPRTN's zinc-binding sub-domain (Li et al., 2019). Thus, SPRTN's active site is only accessible to small flexible peptides and not large proteins. This fits with our earlier observation that SPRTN could cleave the disordered N-terminal tails of histones but not recombinant TOP1 (Vaz et al., 2016). TOP1ccs also lack flexible linker regions between the TOP1 protein and DNA that could be accommodated by SPRTN's active site. This implies that larger DPCs, such as TOP1ccs, need to undergo remodelling or unfolding to be rendered amenable to proteolysis by SPRTN. We suspected that p97 may fulfil this role this since both p97's ATPase activity and SPRTN's SHP box are needed to counteract TOP1cc accumulation. To test this, we isolated TOP1ccs from cells by CsCl-gradient fractionation and assessed the ability of recombinant SPRTN to process TOP1ccs *in vitro.* SPRTN could not process TOP1ccs alone (Fig. 29A, B). However, its activity was significantly enhanced when TOP1ccs were pre-incubated with p97 or TEX264 and p97 (Fig. 29A, B). TEX264 only slightly enhanced the effects of p97 *in vitro*, indicating its role as a recruiting factor is more important *in vivo* (Fig. 29A, B). These results demonstrate that p97 and TEX264 actively promote TOP1cc cleavage by SPRTN *in vitro.*

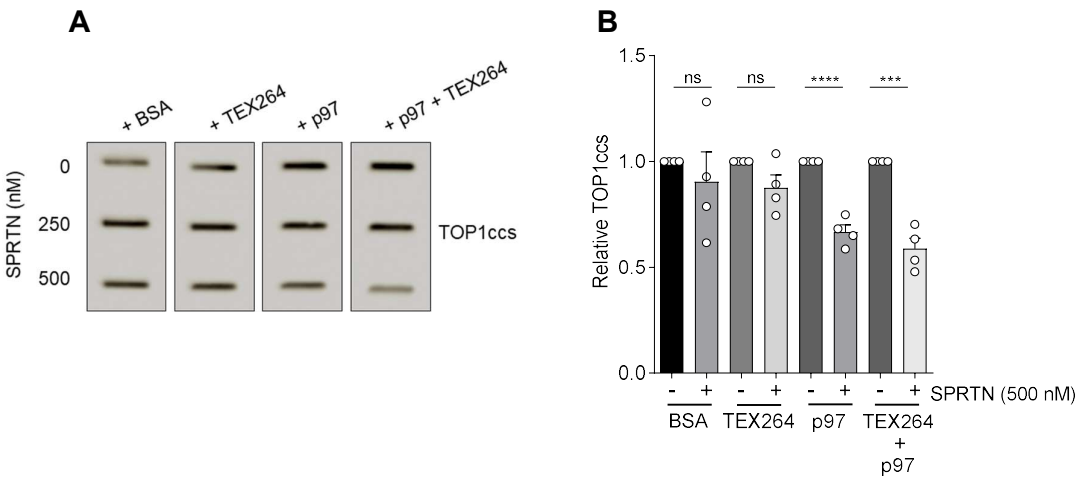



**Figure 29. p97 and TEX264 promote SPRTN-dependent TOP1cc repair *in vitro*.**

(A) Slot analysis of TOP1ccs. TOP1ccs were purified by CsCl-gradient centrifugation from HEK293 cells, incubated with the indicated purified proteins (100 nM/reaction), transferred to a nitrocellulose membrane and probed with a TOP1cc-specific antibody. (B) Quantification of A (mean ± SEM; *n* = 4). TOP1cc levels were normalised to the corresponding control sample (i.e. 0 nM SPRTN/reaction) in each condition. Experiments were performed by Shih-Chieh Chiang (University of Sheffield).

In addition to TOP1ccs, SPRTN processes numerous other DPCs. We therefore sought to address whether TEX264 also plays a role in general DPC repair. To this end, we isolated DPCs by RADAR from TEX264- or SPRTN-depleted HEK293 cells and visualised them by silver staining. As previously observed, SPRTN depletion caused total DPCs to accumulate (Fig. 30A, B). By contrast, TEX264 depletion did not result in total DPC accumulation (Fig. 30A, B). This is consistent with a model in which TEX264 forms a sub-complex with p97 and SPRTN that is important for the repair of TOP1ccs but not all DPC substrates of SPRTN.

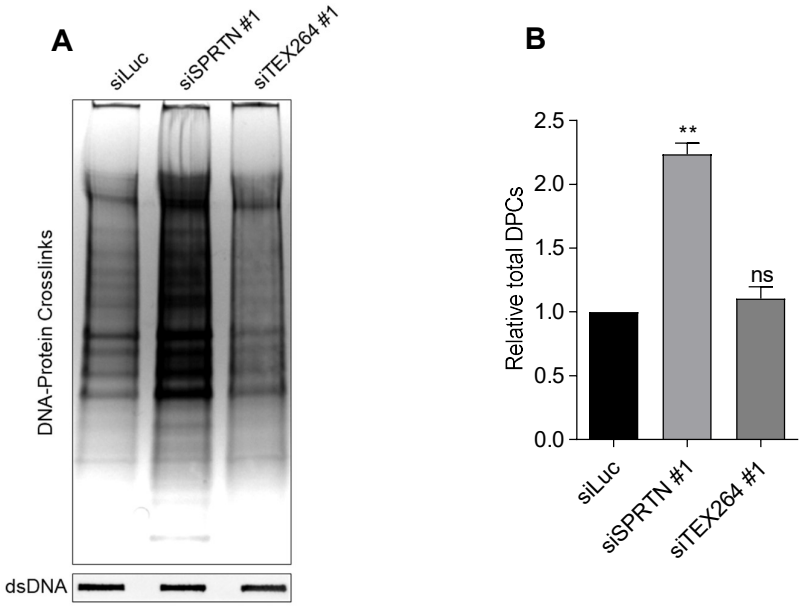



**Figure 30. TEX264 does not regulate general DPC repair.**

(A) RADAR analysis of DNA-protein crosslinks isolated from HEK293 cells transfected with the indicated siRNAs. DPCs were resolved by SDS-PAGE and visualised by silver staining. (B) Quantification of A (error bars represent mean ± SD; $n = 2$; $^{**}P < 0.01$; ns, not significant; Student's t-test). Experiments performed by Annamaria Ruggiano.

### 3.4.6 TEX264 acts at DNA replication forks

Given that SPRTN's function is tightly coupled to DNA replication, we next asked whether TEX264 regulates TOP1cc repair during DNA replication (Davis et al., 2012; Larsen et al., 2019; Mórocz et al., 2016; Mosbech et al., 2012; Vaz et al., 2016).

To identify proteins that interact with TEX264 on chromatin we isolated chromatin from stable, doxycycline-inducible HEK293 cells expressing SSH-tagged TEX264 and analysed anti-HA immunoprecipitates by LC-MS/MS (Fig. 31A). In addition to p97 and TOP1, this analysis detected peptides from many replication factors such as the MCM complex and PCNA. We confirmed these interactions in co-immunoprecipitation experiments (Fig. 31B).

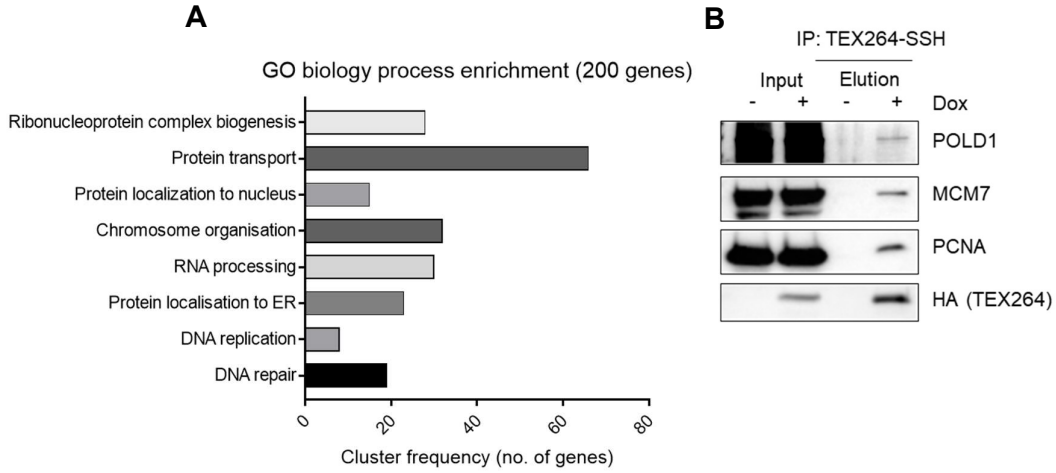



**Figure 31. TEX264 interacts with DNA replication factors.**

(A) Gene ontology (GO) analysis of terms enriched in LC-MS/MS of anti-Strep immunoprecipitates prepared from (Dox) doxycycline-inducible stable HEK293 expressing SSH-tagged TEX264. (B) Immunoblots of same cells used in A. Where indicated cells were treated with Dox for 24 hrs. Experiments performed by Katherine Wiseman.

While this project was ongoing, two papers reported a role for TEX264 in the ER (An et al., 2019; Chino et al., 2019). In one of these papers, the N-terminal LRR of TEX264 was shown to be a single-pass transmembrane domain that inserts into the membrane of the ER (An et al., 2019). However, in their two proteomic analyses of TEX264-interacting proteins, peptides from many nuclear proteins, including TOP1 and PARP1, were among the most abundant of those identified (An et al., 2019).

We therefore analysed the localisation of TEX264 in cells by immunofluorescence confocal microscopy. We observed that TEX264 exhibited a very striking 'rim-like' staining at the nuclear periphery (Fig. 32A). By analysing individual optical sections of cell nuclei, we could exclude out-of-focus fluorescence. This would explain why TEX264 staining differs from that observed by epifluorescence microscopy (see Figs. 23 & 27), where we cannot exclude TEX264 staining deriving from the top and bottom surfaces of nuclei.

Our biochemical fractionations had previously identified subpopulations of TEX264 in the cytoplasm, nucleus, and on chromatin (Fig. 17A). Based on this, our immunofluorescence analysis, and data showing that TEX264's LRR is a transmembrane segment, we speculated that TEX264 may be tethered to the inner nuclear membrane (INM) (An et al., 2019). In support of this, our LC-MS/MS analysis of chromatin-bound TEX264 also identified numerous INM proteins, including, POM121, LEMD3, and components of the nuclear pore complex. We therefore analysed the subcellular localisation of a TEX264 variant that lacks its N-terminus ($TEX264^{\Delta LRR}$; amino acids 1-33). This TEX264 variant did not display the same 'rim-like' staining as $TEX264^{WT}$ (Fig. 32A). Rather, it was distributed

throughout the cytoplasm and nucleoplasm. This is consistent with TEX264 being localised at both the ER membrane and INM.

At the nuclear periphery, TEX264$^{WT}$ co-localised with newly synthesised DNA (i.e. EdU-labelled DNA), TOP1cc foci, and SPRTN (Fig. 32B-D). The nuclear pool of TEX264$^{\Delta LRR}$ formed foci-like structures that also partially co-localised with EdU-labelled DNA, TOP1cc foci, and SPRTN foci (Fig. 32B-D). This indicates that the TEX264 protein has properties that enable it to interact with TOP1 and associate with newly synthesised DNA and SPRTN, regardless of whether it is tethered to a membrane.

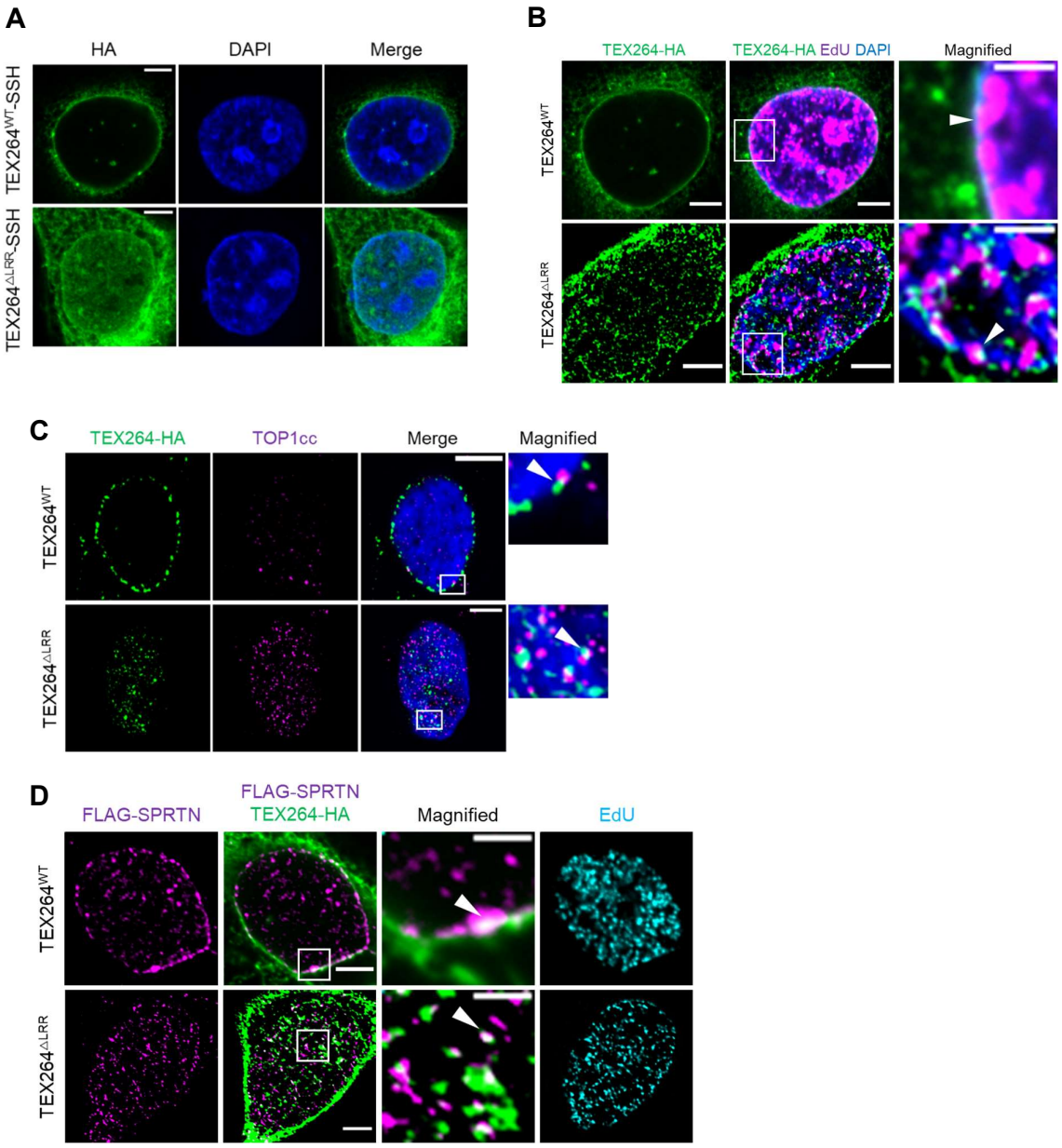



**Figure 32. Immunofluorescence confocal microscopy analysis of TEX264.**

(A) Immunofluorescence images of U2OS cell nuclei. Cells were transfected with cDNA encoding either SSH-tagged TEX264$^{WT}$ or TEX264$^{\Delta LRR}$ and stained with the indicated antibodies. (B) Same as in A, except cells were treated with EdU (10 μM; 30 mins). (C) Same as A. (D) Same as in A and B. Scale bars, 5 μm (large panels) and 2 μm (magnified panels). White arrowheads indicate examples of co-localisation. Experiments were performed by Shudong Li.

Consistent with a role in DNA replication, depletion of either TEX264 or p97 caused DNA replication forks to progress approximately two-fold more slowly as measured by DNA fibre assay (Fig. 33A). As TOP1ccs can stall DNA replication forks, we tested whether TOP1ccs accumulate at DNA replication forks in the absence of TEX264 (Koster et al., 2007). When ΔTEX264 HEK293 cells were treated with CPT (100 nM, 1 hour), we observed a striking accumulation of the TOP1 protein at DNA replication forks, which is likely the result of a failure to clear TOP1ccs (Fig. 33B). Interestingly, we observed that SPRTN was less abundant at DNA replication forks in ΔTEX264 cells, both before and after CPT treatment (Fig. 33B). As TEX264 deficiency does not impair total DPC repair, this result suggests that TOP1ccs may be among the more abundant SPRTN substrates in human cells. We did not observe a reduction in p97 levels at DNA replication forks in ΔTEX264 cells. This is probably because p97 has pleiotropic roles in DNA replication and can be recruited to DNA replication forks by other cofactors (Fig. 33B) (Franz et al., 2016b, 2016a, Maric et al., 2014, 2017; Moreno et al., 2014).

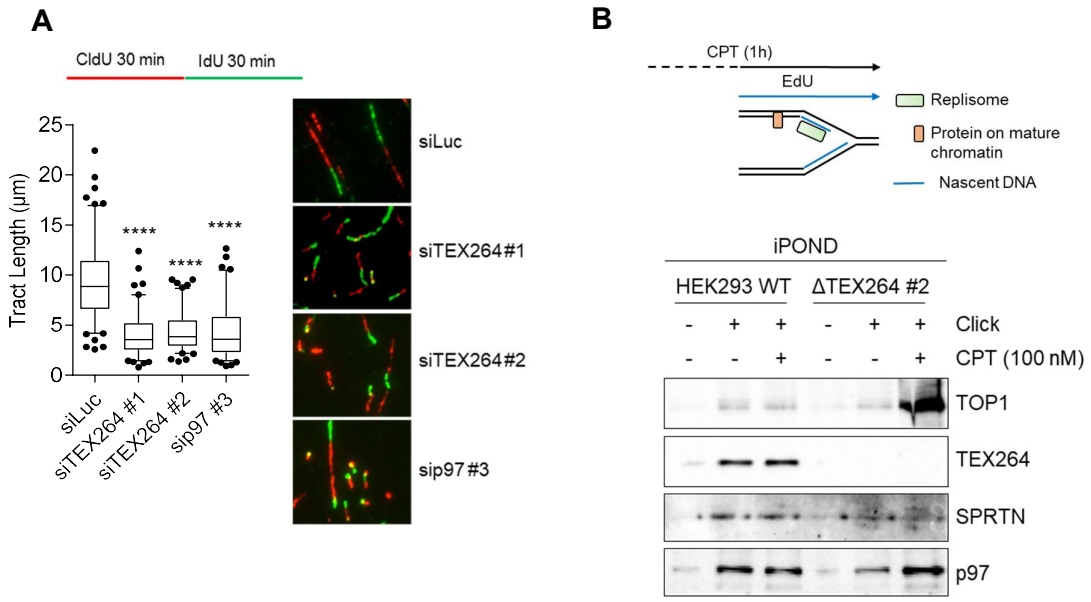



**Figure 33. TEX264 acts at DNA replication forks.**

(A) Left panel: DNA fibre analysis of replication fork velocity in HEK293 cells treated with the indicated siRNAs. IdU track lengths are shown. At least 100 fibres were measured per condition (****$P < 0.0001$; two-tailed Mann-Whitney test). Right panel: representative DNA fibre images. (B) Upper panel: schematic of iPOND experiment performed below. Lower panel: immunoblots of iPOND experiment performed in WT and ΔTEX264 HEK293 cells. Where indicated cells were treated with CPT (100 nM) for a total of 1 hr. Fig. 33B generated by Ignacio Torrecilla.

TOP1ccs can stall DNA replication and exacerbate DNA damage (Strumberg et al., 2000). We therefore reasoned that reducing the occurrence of TOP1ccs would alleviate the DNA damage observed in TEX264-deficient cells. When TEX264 is depleted, the intensity of nuclear γH2AX is increased, consistent with these cells suffering from elevated numbers of DNA strand breaks (Fig. 34A-C). TOP1 depletion did not result in increased DNA damage in interphase nuclei (Fig. 34A-C). This suggests that TOP1ccs induce DNA damage to a greater extent than the DNA topological stress that arises when TOP1 function is impaired due to TOP1cc formation. Strikingly, when TOP1 was co-depleted in TEX264-depleted cells, the nuclear intensity of γH2AX was significantly decreased (Fig. 34A-C). This indicates that the TOP1 protein drives a large proportion of the DNA damage observed in TEX264-depleted cells via the formation of TOP1ccs.

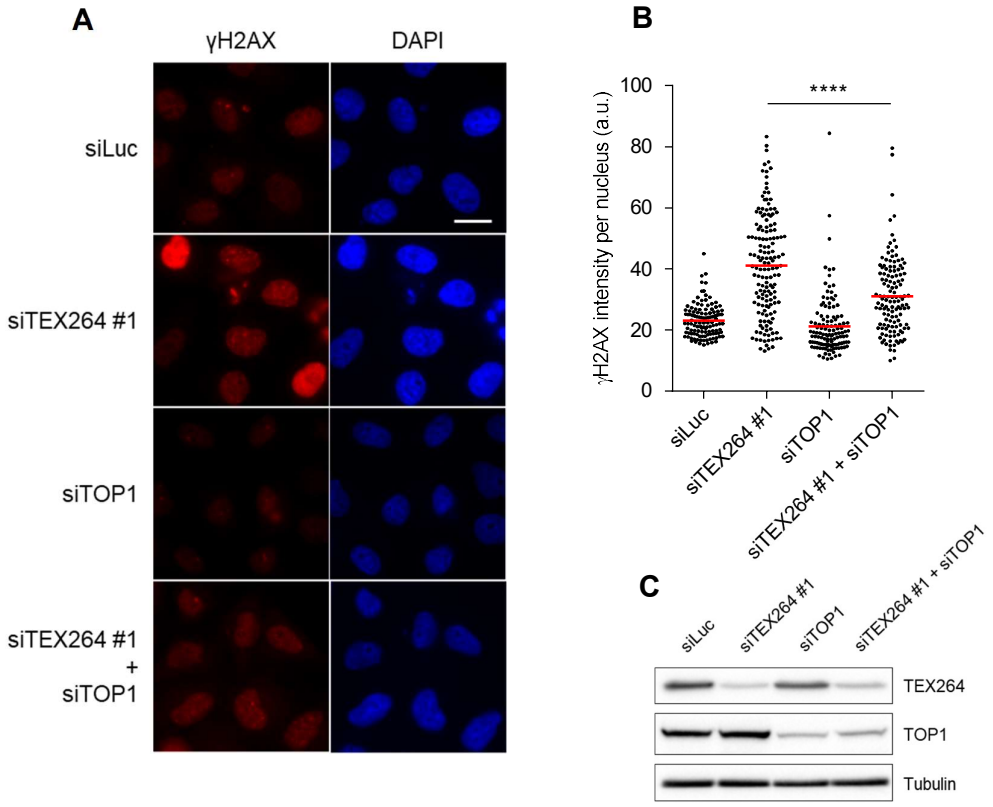



**Figure 34. TOP1ccs drive DNA damage in TEX264-deficient cells.**

(A) Representative images of HeLa cell nuclei. Cells were transfected with the indicated siRNA and stained with an antibody against phosphorylated (Ser139) γH2AX. (B) Quantification of the mean nuclear γH2AX intensity for experiment in E (****$P$ < 0.0001; two-tailed Mann-Whitney test). At least 100 nuclei were measured per condition and experiment. Scale bar, 20 μm. (C) Immunoblots to confirm depletions for A.

Altogether, we have identified a critical role for the GyrI-like domain-containing protein, TEX264, in TOP1cc repair. TEX264 is localised to the nuclear periphery via its N-terminal LRR and recognises both unmodified and SUMO1-modified TOP1 and promotes TOP1cc repair by recruiting p97 and SPRTN. We propose that p97 uses its ATPase activity to unfold the TOP1 protein to enable it to be efficiently cleaved by SPRTN. We propose that the remnant DNA-bound peptide can then be excised by TDP1 (Fig. 35).

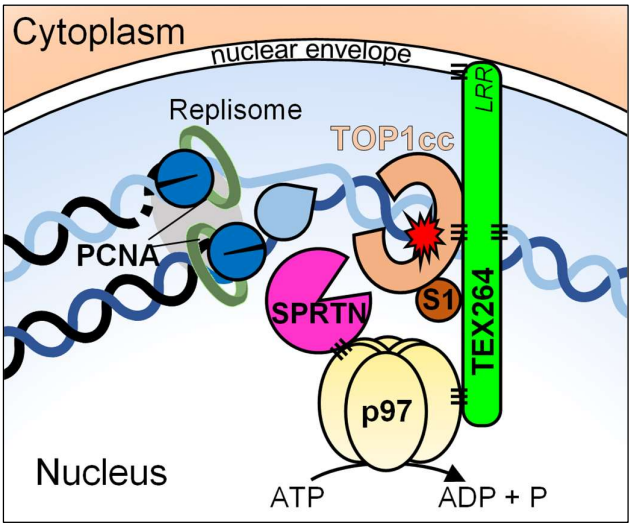



**Figure 35. Proposed model.**

TEX264 is localised to the nuclear periphery via its N-terminal LRR. TEX264 binds SUMO1-modified and unmodified TOP1 and promotes their removal by recruiting p97-SPRTN sub-complexes to TOP1ccs during DNA replication. ATP hydrolysis by p97 may promote remodelling of the TOP1 protein to make it amenable to degradation by SPRTN.

## 3.5 SPRTN as a biomarker

### 3.5.1 High SPRTN expression correlates with poor prognosis in metastatic colorectal cancer

The camptothecin derivative, irinotecan, is frontline therapy for many types of cancers, including colorectal cancer. The MRC FOCUS trial was initiated to assess whether patients with metastatic colorectal cancer (mCRC) had higher response rates when treated with irinotecan plus 5-fluorouracil (5FU) and folinic acid (LV) (FOLFIRI) compared with standard 5FU/LV therapy (Seymour et al., 2007). Patients response rates were 10-15% when treated with 5FU/LV and 40-50% when treated with FOLFIRI. Nevertheless, up to 60% of patients did not respond to therapy, underscoring the need to identify molecular determinants of resistance and biomarkers that can predict patient responses.

We have identified a protein complex, composed of TEX264, SPRTN, and p97, that is important for TOP1cc repair. These factors are critical for TOP1cc repair in cells treated with clinically relevant doses of CPT, which are in the nanomolar range (Slichenmyer et al., 1993; Takimoto and Arbuck, 1997). We therefore speculated that expression of these components may correlate with clinical responses to camptothecin derivatives. Specifically, we hypothesised that high expression of any of these components would promote clinical resistance to irinotecan.

We obtained access to primary tumour tissue samples and mRNA expression data from 78 mCRC patients who went on to receive FOLFIRI in the MRC FOCUS trial. The mRNA expression of TEX264, SPRTN, and p97 was analysed, as was the expression of TDP1 and TOP1, whose expression may also influence tumour

response to FOLFIRI (Fig. 36A) (Gilbert et al., 2012). The correlation between the RNA expression of each factor and response to therapy was analysed by linear regression (Fig. 36B). Only SPRTN expression was significantly different in responders vs. non-responders. Specifically, SPRTN expression was significantly higher in the patients who did not respond to FOLFIRI (Fig. 36B).

Using the trial data, we selected 6 of these patients who did not respond to therapy (5 with progressive disease; 1 with stable disease) and 6 patients who did (6 complete responders) for further analysis. The expression of p97, TEX264, and SPRTN in the tumours of these patients was assessed by immunohistochemistry (IHC) in a blinded fashion. The intensity of expression and the percentage of positively stained tumour cells for each protein were estimated by a pathologist. A H-score for each protein was calculated by multiplying the percentage of positively stained tumour cells by the intensity of protein expression. TEX264 and p97 were highly expressed in all tumour samples (Figs. 38 & 39). However, in line with our mRNA expression analysis, SPRTN expression was significantly higher in the patients who ultimately did not respond to FOLFIRI (Fig. 37A-D). These results suggest that SPRTN may be a driver of irinotecan resistance, that its expression levels can predict response to therapy, and inhibition of its function may improve patient outcomes.

**A**

| Response | Number of patients |
|---|---|
| Complete | 5 |
| Partial | 30 |
| Stable Disease | 27 |
| Progressive Disease | 16 |
| Total | 78 |

**B**

| Gene | *P* value |
|---|---|
| SPRTN | 0.0139 * |
| TDP1 | 0.43307 |
| TEX264 | 0.779788 |
| TOP1 | 0.578 |
| VCP/p97 | 0.68454 |

**Figure 36. SPRTN expression correlates with resistance to irinotecan in mCRC.**

(A) Table showing the clinical responses of 78 mCRC patients in whom mRNA expression data was analysed in B. (B) Table of genes whose mRNA expression was analysed in patients shown in A. The association between gene expression and clinical response was analysed with linear regression. Statistical significance was considered at $*P < 0.05$. mRNA expression analysis was performed by Sylvana Hassanieh and Enric Domingo.

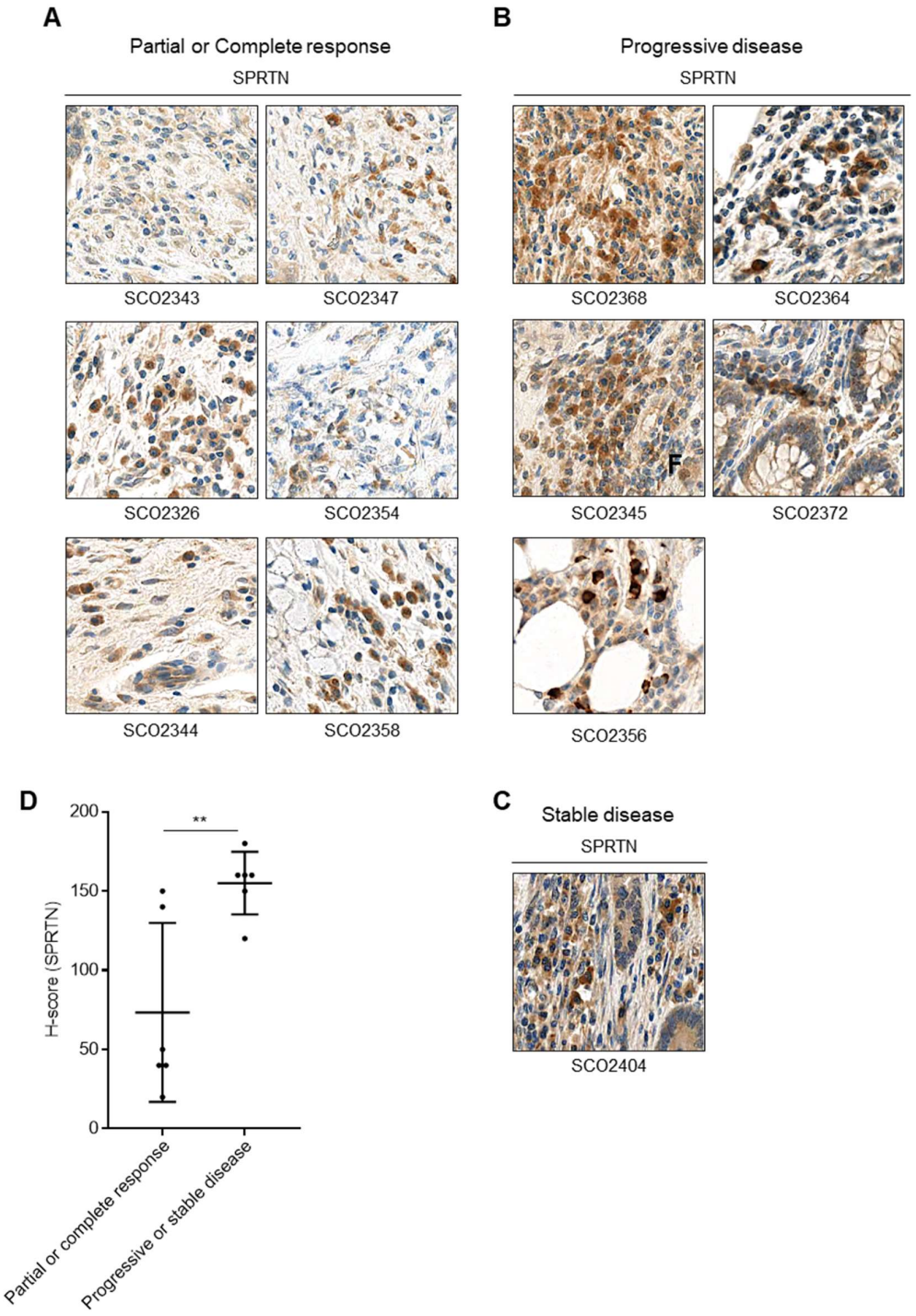



**Figure 37. SPRTN protein expression correlates with resistance to irinotecan in mCRC patients.**

Histological analysis of SPRTN expression in primary mCRC tumours before FOLFIRI. SPRTN expression in patients who exhibited a partial or complete response to therapy (A), patients with progressive disease (B), and in a patient with stable disease after therapy (C). (D) H-scores for SPRTN, calculated by multiplying the percentage of positively stained tumour cells by the intensity of SPRTN staining (**$P$ < 0.01; Student's t-test). IHC samples were scored by Robert Goldin.

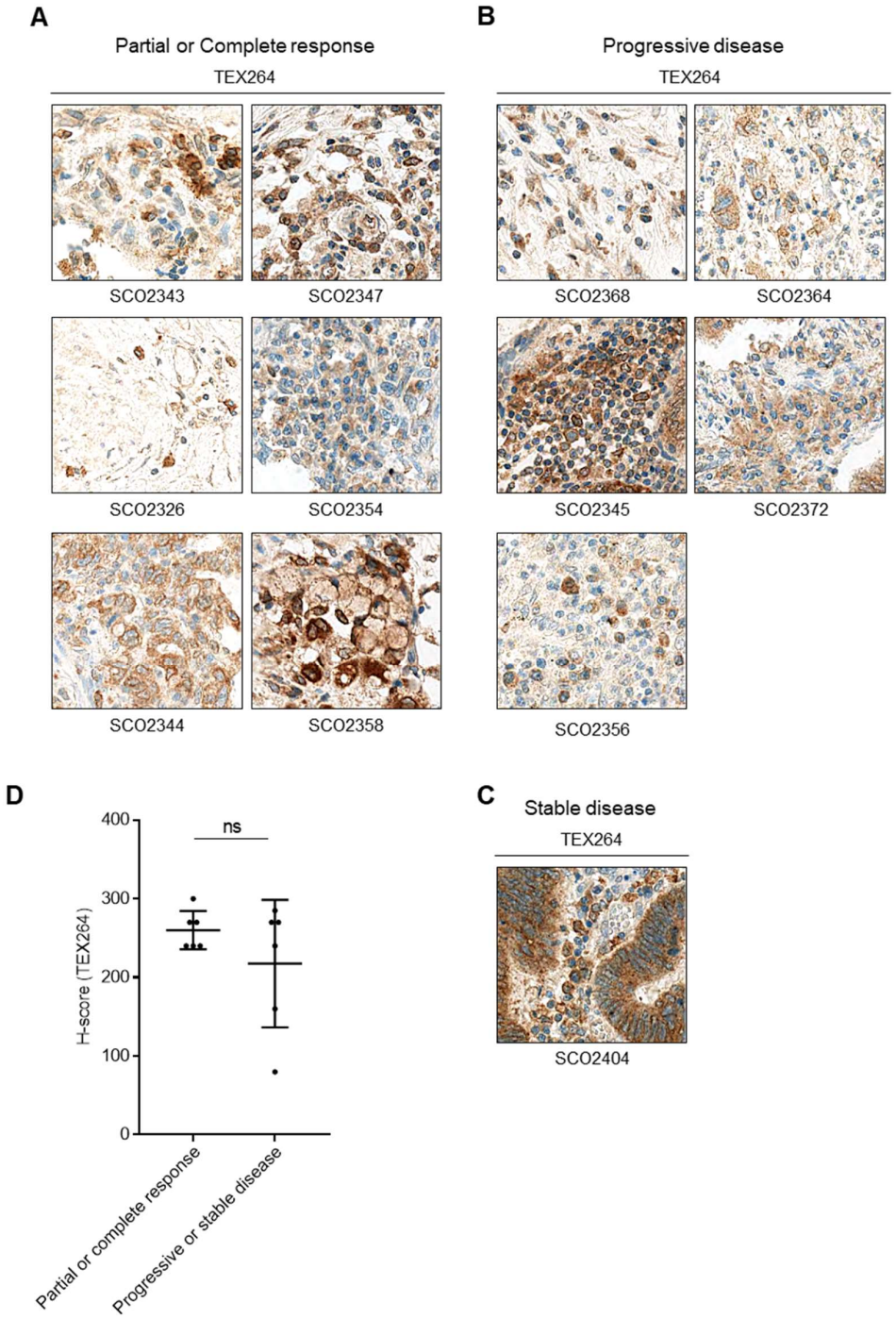



**Figure 38. TEX264 expression does not correlate with resistance to irinotecan in mCRC patients.**

Histological analysis of TEX264 expression in primary mCRC tumours before FOLFIRI. TEX264 expression in patients who exhibited a partial or complete response to therapy (A), patients with progressive disease (B), and in a patient with stable disease after therapy (C). (D) H-scores for TEX264, calculated by multiplying the percentage of positively stained tumour cells by the intensity of TEX264 staining (ns, not significant; Student's t-test). IHC samples were scored by Robert Goldin.

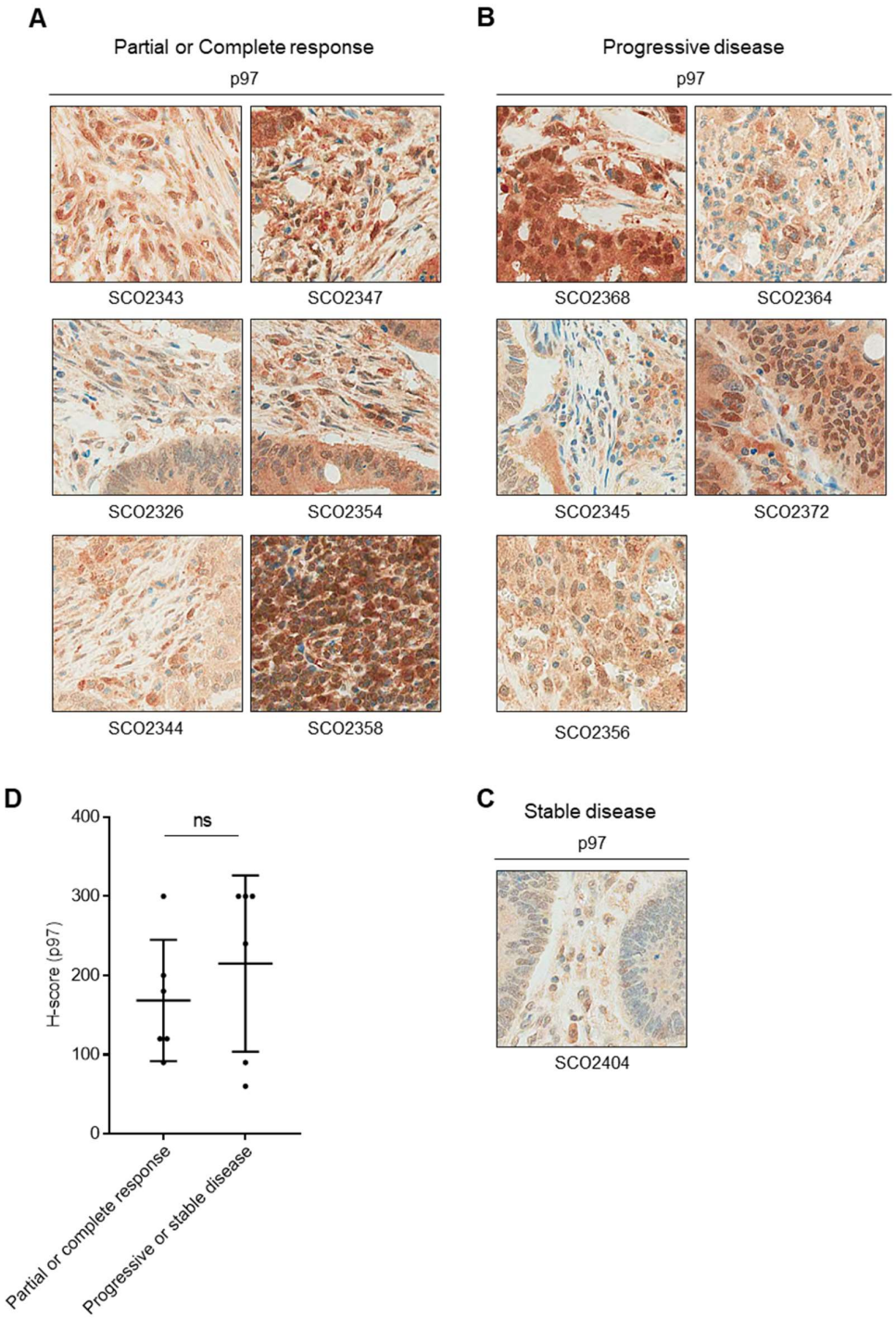



**Figure 39. p97 expression does not correlate with resistance to irinotecan in mCRC patients.**

Histological analysis of p97 expression in primary mCRC tumours before FOLFIRI. p97 expression in patients who exhibited a partial or complete response to therapy (A), patients with progressive disease (B), and in a patient with stable disease after therapy (C). (D) H-scores for p97, calculated by multiplying the percentage of positively stained tumour cells by the intensity of p97 staining (ns, not significant; Student's t-test). IHC samples were scored by Robert Goldin.

# 4 Discussion

The work presented in this thesis has contributed to our understanding of the roles of SPRTN, p97, and the previously uncharacterised protein TEX264 in maintaining genome stability. We have identified SPRTN as an essential, DNA replication-coupled DPC repair factor. Furthermore, we have established that TEX264, p97 and SPRTN cooperate to repair TOP1ccs. p97 and SPRTN are recruited to TOP1 by TEX264. TEX264 recognises both unmodified and SUMO1-modified TOP1. TEX264 is localised to the nuclear periphery and counteracts TOP1ccs at DNA replication forks. We propose that p97 remodels TOP1ccs to expose protein regions that are amenable to cleavage by SPRTN. This is supported by our observations that p97's ATPase activity and SPRTN's SHP box are necessary for TOP1cc repair. TEX264 and p97 act in the same TOP1cc repair pathway as TDP1, suggesting that the remnant DNA-bound peptide resulting from TOP1cc cleavage by SPRTN can be excised by TDP1. The importance of TEX264 for genome stability is demonstrated by the accumulation of DNA damage in TEX264-deficient cells.

Neither TEX264 nor p97 bind TOP1 in response to high dose CPT treatment. Low and high doses of CPT elicit dramatically different cellular responses. High doses of CPT induce DSB formation and replication fork collapse (i.e. the dissociation of replication factors from chromatin). Our data indicate that TEX264, p97, and SPRTN travel with the replisome and encounter TOP1ccs that impede fork progression. Thus, replication fork collapse would diminish the likelihood of these factors encountering TOP1ccs. Under such conditions, the proteasome or DNA endonucleases may play a greater role in TOP1cc repair, as previously reported

(Desai et al., 1997; Lin et al., 2008; Zhang et al., 2011). Alternatively, or in addition to this, the reported role of TDP1 in DSB repair may account for why TDP1-deficient cells, but not TEX264-deficient cells, are sensitive to micromolar doses of CPT (Kawale et al., 2018).

TEX264 was identified in a screen for proteins that interact with p97 upon IR. The interaction between p97 and TEX264 was only detected in the nucleus and was stimulated by irradiation. It is unclear whether the latter observation is linked to TOP1cc repair. IR can induce DPCs, however, CPT treatment does not notably stimulate the interaction between p97 and TEX264. It is possible that TEX264 may recruit p97 to other substrates, but it is unlikely to regulate DSB repair since TEX264-depleted cells are not sensitive to micromolar doses of CPT. It will be interesting to address which other DNA-damaging agents TEX264-deficient cells are sensitive to since this could provide clues about the other types of substrates and/or DNA lesions that TEX264 may recognise.

## 4.1 TEX264 at the INM

TEX264 has recently been implicated in a process known as ER-phagy (An et al., 2019; Chino et al., 2019). This is a recently described process in which components of the ER are degraded by autophagy (Khaminets et al., 2015). When cells are starved of amino acids, TEX264 is degraded by autophagy and TEX264 deficiency results in a failure to degrade certain components of the ER. This process what shown to be depend on a LIR motif in TEX264 which binds the autophagosomal protein LC3B (An et al., 2019). Overall, TEX264 was proposed to be anchored to the ER membrane via its N-terminal LRR and face into the

cytoplasm where it acts as a receptor for LC3B-containing phagophores to enable ER proteins to be shuttled into autophagosomes (An et al., 2019; Chino et al., 2019).

It will be interesting to test if TEX264's role in TOP1cc repair is linked to autophagy. The mass spectrometry data from An et al., 2019 hints that it may be. For example, TEX264 bound more strongly to TOP1 upon starvation and is found in proximity with p97 within autophagosomes. Autophagy has previously been linked to SUMO- and Cdc48-dependent repair of TOP1ccs. For example, Wss1 interacts with Doa1 and relocalises to vacuoles in response to replication stress (Balakirev et al., 2015). It is unclear why DPCs would be degraded by autophagy, however, evidence suggests that autophagy degrades proteins which are poor substrates for the proteasome (Lu et al., 2017). The proteasome is efficient at degrading soluble proteins but not protein aggregates. It is conceivable that DPCs are more akin to protein aggregates. For example, endogenous formaldehyde can induce protein-protein crosslinking in addition to DNA-protein crosslinking. Thus, DPCs are unlikely to occur in isolation. Autophagy may represent a way to export and recycle bulky protein-protein aggregates, DPCs, and/or protein fragments generated by proteolytic cleavage from the nucleus.

The fact that TEX264$^{\Delta LRR}$ relocalises to the nucleoplasm suggests it may be anchored at the INM. Formal proof of this will require an analysis of TEX264's localisation by electron microscopy. Nonetheless, our confocal microscopy analysis, biochemical fractionations, and iPOND experiments firmly demonstrate that a subpopulation of TEX264 is nuclear/chromatin-bound. Importantly, TEX264

is required for TOP1cc repair even in unchallenged cells. TEX264 would not be the first p97 cofactor to be membrane-bound. For example, UBXD8 is embedded in the ER membrane as well (Mueller et al., 2008).

TOP1 has recently been shown to predominantly act on heterochromatic genomic regions that are associated with the INM (Manzo et al., 2018). These regions are thought to be more prone to topological stress that needs to be resolved by TOP1. Therefore, it is conceivable that TOP1ccs are highly abundant at the nuclear periphery. An alternative hypothesis is that TOP1ccs are recruited to the INM. Indeed, other types of DNA lesions and SUMOylated proteins are recruited to the nuclear envelope. For example, SUMOylated histone H2Az is recruited to nuclear pore complex proteins upon DSB formation (Kalocsay et al., 2009). The re-localisation of damaged DNA to the nuclear periphery is thought to suppress excessive genomic rearrangements and provide an anchoring point for DNA repair. This concept is supported by the recent finding that the nucleoskeleton protein, IFFO1, interacts with lamin A/C and is recruited to DNA damage by XRCC4. Its interactions with both lamin A/C and XRCC4 are crucial for immobilising broken DNA ends and suppressing chromosomal rearrangements (Li et al., 2019). A similar mechanism may be at play in TOP1cc repair.

Another intriguing possibility is that the TEX264-p97 complex also acts downstream of the proteasome. Proteasomal proteolysis predominantly occurs at the nuclear envelope and rough ER (Enenkel et al., 1999). Cdc48 also acts at the nuclear envelope and in the ER (Latterich et al., 1995). It will be interesting to assess whether TOP1ccs do indeed relocalise to the nuclear envelope and to understand the mechanisms that underlie this. It is also conceivable, if unlikely,

that TEX264 dissociates from the INM to perform its function. These questions could be addressed by live-cell confocal microscopy analysis of endogenous TEX264 and TOP1, fused at their genetic loci with fluorescent protein tags.

## 4.2 The role of SUMO

The role of SUMO in p97-dependent TOP1cc repair is consistent with observations in yeast. While Cdc48 and Ufd1 can both bind directly to SUMO in yeast, this has never been demonstrated for their human homologs (Bergink et al., 2013; Nie et al., 2012). Moreover, to date, only one p97 cofactor – Ataxin-3 - has been shown to bind SUMO1 (Pfeiffer et al., 2017). Thus, TEX264 is quite unique amongst p97 cofactors in this respect. We do not however rule out a role for ubiquitin and Ufd1 in TOP1cc repair in humans. As discussed, the hierarchical binding of p97-Ufd1-Npl4 to additional cofactors can provide extra substrate-specificity control (Hänzelmann et al., 2011).

SUMOylation can influence protein function in a variety of ways. For example, selective substrate SUMOylation can facilitate protein-protein interactions, wherein a SUMOylated protein is recognised by SIMs in their binding partners. For example, a ‘split’ SIM in TDP2 is thought to enable it bind TOP2ccs more efficiently (Schellenberg et al., 2017). Alternatively, entire groups of proteins can be SUMOylated as is the case for HR proteins (Psakhye and Jentsch, 2012). In these circumstances, SUMO is proposed to help stabilise protein-protein interactions to facilitate the formation of larger protein assemblies. In the work presented here, SUMO1 appears to act as a signal for the recruitment of TEX264 to TOP1. It will be of interest to identify the E3 SUMO ligase(s) responsible for SUMOylating TOP1ccs, since these may contribute to clinical resistance to TOP1 poisons. It will also be interesting to assess whether SUMO1 has any impact on

SPRTN's proteolytic activity in the same way that ubiquitin does. SPRTN's cleavage activity is enhanced by the presence of free ubiquitin but not free SUMO2, however, SUMO1 has yet to be tested (Lopez-Mosqueda et al., 2016). Indeed, SPRTN possesses putative SIMs (Carmell et al., 2016).

*In vitro* and structural studies could improve our understanding of the role of SUMO1 in TOP1cc repair. For example, does SUMO1 enhance the strength of the TOP1cc-TEX264 interaction? Does it facilitate the ability of TOP1 binding residues on TEX264 to come into contact with TOP1? In particular, it will be interesting to crystallise the structure of TEX264 with its unique alpha helix (residues 193 - 210). It seems likely that this helix folds back onto the TOP1 structure and comes into proximity with SIM2 and TEX264's TOP1 binding region (Paul Elliott, personal communication).

### 4.2.1 ACRC: a SUMO-targeted DPC repair protease?

Recent evidence suggests that SUMOylation is also important for general DPC repair. Formaldehyde, but not other DNA-damaging agents, induces SUMOylation of chromatin-bound proteins and DPCs (Borgermann et al., 2019). How this promotes DPC repair is not well understood. However, SUMOylated DPCs persisted when SPRTN was depleted, suggesting SPRTN plays a role in clearing SUMOylated DPCs (Borgermann et al., 2019).

A second DPC repair protease may also exist in higher eukaryotes, known as ACRC (also known as GCNA). ACRC was recently identified as a SprT domain-containing protein (Borgermann et al., 2019; Carmell et al., 2016). ACRC is a member of a family of proteases that is more closely related to SPRTN than Wss1 (Fig. 40) (Fielden et al., 2018). While ACRC appears to be mostly expressed in

germ cells, exogenous ACRC can promote DPC repair in mammalian cells in a manner that depends on its putative protease active site (Borgermann et al., 2019). It will be important to address if ACRC is indeed a protease *in vitro* and, if so, what factors are required to promote its cleavage activity. ACRC binds SUMO2/3 via a cluster of SIMs and these are important for its role in DPC repair. It also appears to act in a DNA replication-independent manner, unlike SPRTN, indicating that it may act outside of S phase (Borgermann et al., 2019).

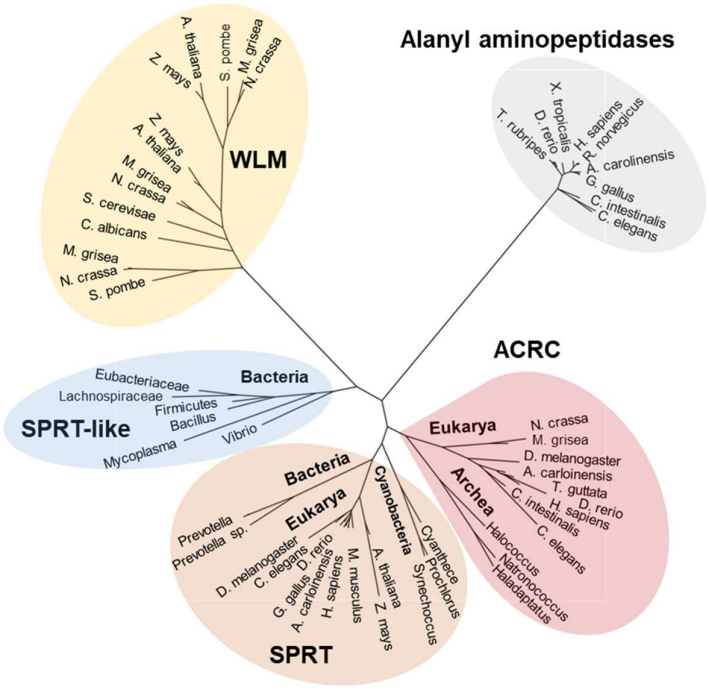



**Figure 40. Phylogenetic relationship of SPRT and WLM proteases.**

ACRC is closely related to the SPRT family of proteases and more distantly related to the WLM family of proteases. The alanyl aminopeptidase family was used as an outgroup. Protein sequences were aligned using MAFFT and the phylogenetic tree was constructed using maximum likelihood analysis (Fielden et al., 2018).

## 4.3 Novel insights into SPRTN function

Since 2016, at least 9 papers have been published on the role of SPRTN in DPC repair (Halder et al., 2019; Larsen et al., 2019; Li et al., 2019; Lopez-Mosqueda et al., 2016; Maskey et al., 2017; Mórocz et al., 2016; Sparks et al., 2019; Stingele et al., 2016; Vaz et al., 2016). Together these papers have provided new insights into SPRTN's function, regulation, and structure.

A crucial role for SPRTN in TOP1cc repair was supported by follow-up studies in *SPRTN*-hypomorphic mice (Maskey et al., 2017). MEFs derived from these mice were hypersensitive to camptothecin but not etoposide or PARP inhibition. TOP1cc foci accumulated from an early age in the livers of these mice, but not in other tissues. This may be explained by the fact that the TOP1 protein is much more abundant in the liver than in other tissues and therefore that TOP1ccs are more common (Maskey et al., 2017). However, it is also possible that other proteases/factors facilitate TOP1cc repair in other tissues. SPRTN-hypomorphic mice developed spontaneous tumours, particularly in the liver. As TOP1ccs can be highly mutagenic, this study raises the possibility that genome stability associated with defective TOP1cc repair may contribute to tumourigenesis (Balestrieri et al., 2001; Hashimoto et al., 1995). These findings may also explain our observation that TEX264 deficiency reduces SPRTN recruitment to DNA replication forks, without affecting total DPC repair: as is the case in mice, TOP1ccs may be among the most common substrates of SPRTN during DNA replication.

Much attention has been placed on trying to understand how SPRTN's function is regulated. Given that SPRTN cleavage activity is not protein sequence-specific,

its activity needs to be tightly regulated to prevent cleavage of DNA-bound proteins that are not DPCs near the replisome (Vaz et al., 2016). One way its activity may be controlled is through its DNA-binding activity. Some groups concluded that SPRTN's protease activity is equally activated by both ss- and ds-DNA (Lopez-Mosqueda et al., 2016; Mórocz et al., 2016; Vaz et al., 2016). However, others found that SPRTN is preferentially activated by ssDNA (Stingele et al., 2016). These seemingly contradictory conclusions appear to be explained by the different lengths of DNA templates used in the respective *in vitro* substrate cleavage assays. For example, 42 nucleotide-long ss- or ds-DNA fragments activate SPRTN equally well, whereas shorter single-strand phage DNA activates SPRTN more efficiently than double-stranded phage DNA (Li et al., 2019). The importance of ssDNA *in vivo* is supported by the observation that Rad18 is epistatic with SPRTN in DPC repair (Mórocz et al., 2016). This suggests that the collision between a replisome and a DPC could lead to helicase polymerase uncoupling, ssDNA formation, and Rad18 activation.

Initially this seemed unlikely to happen on a leading strand since a DPC would be expected to block helicase progression and thereby suppress helicase-polymerase uncoupling (Duxin et al., 2014). However, another study found that the CMG helicase can bypass a leading strand DPC with the aid of another DNA helicase, RTEL1 (Sparks et al., 2019). In Xenopus egg extract, RTEL1 promotes CMG bypass of a plasmid-borne model DPC. RTEL1 binds the lagging strand, opposite the CMG helicase. When a CMG helicase encounters a DPC, RTEL1 is thought to unwind the DNA ahead of the DPC. This step was found to be necessary for DPC bypass by the CMG helicase and precedes DPC proteolysis by SPRTN. This mechanism explains how ssDNA could be generated behind a

DPC, whereupon it could stimulate SPRTN's protease activity. It also provides an explanation for how the CMG helicase is protected from degradation by SPRTN. However, it is still unclear exactly how the CMG bypasses the DPC. A possible model is that the CMG helicase ring opens to enable it to slide past the DPC. An important point to note is that the model DPC used in these *in vitro* studies is 45 kDa. Thus, this model may explain how the CMG helicase bypasses smaller DPCs such as histones. However, it probably does not account for how the CMG helicase bypasses much larger DPCs such as TOP1ccs, which the CMG helicase would be unlikely to slide passed.

A second study performed in Xenopus extract identified a role for the proteasome in DPC repair (Larsen et al., 2019). The proteasome-dependent pathway requires DPC ubiquitination and the presence of ssDNA near the DPC. Interestingly, SPRTN-dependent DPC repair required SPRTN's UBZ but not DPC ubiquitination, suggesting SPRTN may be recruited to another ubiquitinated protein in the vicinity of the DPC. A likely candidate would be PCNA given the purported involvement of Rad18 in DPC repair (Mórocz et al., 2016). How these two proteases operate *in vivo* is an important line of future enquiry. The involvement of either protease over the other could be influenced by differences in DPC size or chemistry.

Data presented in this thesis suggest that an additional level of SPRTN regulation exists: one which is mediated by another p97 cofactor. The involvement of TEX264 and p97 may reflect the need to remodel and recognise TOP1ccs to facilitate their cleavage by SPRTN. This is likely due to the unique obstacle that a TOP1cc represents. The TOP1 protein is far larger than other known substrates

of SPRTN, such as histones, and lacks unstructured protein regions. It will be interesting to assess whether this mode of SPRTN regulation applies to other bulky substrates of SPRTN, such as TOP2ccs.

The activity of SPRTN is also regulated by its own PTMs. A subpopulation of SPRTN in cells is monoubiquitinated and this is reduced in response to formaldehyde treatment. A model was proposed in which the SPRTN protein folds in such a way that its UBZ binds to a ubiquitin attached to SPRTN (Mosbech et al., 2012). This conformation is thought to retain SPRTN in an inactive state, possibly by shielding its DNA-binding domains. Accordingly, SPRTN's deubiquitination coincides with its recruitment to chromatin (Stingele et al., 2016). More recently, CHK1 was found to phosphorylate SPRTN on its C-terminus to promote SPRTN recruitment to chromatin. SPRTN variants that cannot be phosphorylated by CHK1 fail to rescue defective DNA replication fork progression in SPRTN-depleted cells (Halder et al., 2019). Finally, an additional layer of SPRTN regulation occurs through self-cleavage. SPRTN molecules can cleave one another, which is more than likely a mechanism to restrain SPRTN’s protease activity on DNA. Autocleavage and substrate cleavage occur with similar kinetics and cleaved products of SPRTN are shorn of their DNA-binding residues, rendering them inactive (Vaz et al., 2016).

A replication-independent function for the Drosophila SPRTN ortholog MH in sperm before the first zygotic division has been proposed (Delabaere et al., 2014). Sperm DNA is densely packaged which can lead to topological stress. This led the authors to speculate that sperm may accumulate TOP-ccs that MH/SPRTN is needed to resolve. A replication-independent role of SPRTN is also suggested by

its importance for cell viability in post-mitotic *C. elegans* (Stingele et al., 2016). These findings underscore the importance of DPC repair outside of S phase and more work is needed to characterise the pathways involved in this process. As TEX264 has SPRTN-independent roles in TOP1cc repair and is expressed throughout the cell cycle, it is likely to play a role in TOP1cc repair outside of the S phase. In principle, TEX264 and p97 may be able to recognise and remodel TOP1ccs to enable their excision by TDP1. This would presumably prevent the proteolytic degradation of TOP1 protein and allow it to be recycled. The involvement of SPRTN during DNA replication may reflect an increased TOP1cc burden during DNA replication and/or the need to rapidly process TOP1ccs to prevent the cytotoxic consequences of persistent DNA replication fork stalling and collapse.

## 4.4 SPRTN as a driver of irinotecan resistance

Many chemotherapeutics in clinical use exert their anti-tumour activity by inducing DPCs. This raises the possibility that targeting DPC repair factors could further sensitise tumours to DPC-inducing agents. Analysing the expression of DPC repair factors may also predict the efficacy of DPC-inducing chemotherapeutics and improve patient stratification. Our finding that high SPRTN expression correlates with resistance to the TOP1 poison, irinotecan, in patients with mCRC supports this notion.

Previous studies have suggested that TOP1 and TDP1 may have utility as biomarkers in CRC patients treated with irinotecan. Specifically, in some clinical trials, patients with moderate or high TOP1 expression showed the greatest

responses to therapy (Braun et al., 2008). However, subsequent studies have found no correlation between TOP1 expression and response to therapy (Koopman et al., 2009). TDP1 expression is also reported to be significantly elevated in colorectal cancers compared with normal tissue (Snipstad et al., 2010; Yu et al., 2005). Our finding that the expression of neither of these two genes correlates with clinical response to irinotecan-based chemotherapy highlights the need to further untangle the molecular basis of irinotecan resistance.

Patient tissue and mRNA samples were obtained prior to the commencement of treatment. Differential SPRTN expression may therefore confer a selective advantage to tumours, even in the absence of irinotecan. For example, high SPRTN expression may promote DNA replication progression in tumours with a high DPC burden. This may explain why SPRTN, but not TEX264 or p97, was differentially expressed between patients prior to FOLFIRI treatment. Moreover, TEX264 and p97 have roles in other pathways whereas SPRTN's function is almost exclusively coupled to DNA replication. The fact that both TEX264 and p97 are more abundant than SPRTN suggests that SPRTN expression may be a 'limiting factor' in the ability of cells to repair TOP1ccs (Wang et al., 2012).

A phase I clinical trial assessing the effects of the p97 ATPase inhibitor, CB-5083, on lymphoid hematologic malignancies and advanced solid tumours was recently terminated due to the inhibitors cytotoxic side effects (https://clinicaltrials.gov/ct2/show/NCT02223598). Given that p97 has roles in numerous cellular pathways, including DSB repair, it is perhaps not surprising that inhibition of its function would have undesirable side effects (Meerang et al., 2011;

Torrecilla et al., 2017). Owing to its unique protease active site, low abundance, and S/G2 phase-specific expression, SPRTN may represent a more viable target of therapeutic intervention. For SPRTN to be used as an effective biomarker in future, it will be important to precisely define what constitutes ‘high’ and ‘low’ SPRTN expression. Specifically, a standardised IHC staining protocol would need to be introduced and a comprehensive analysis of SPRTN transcript abundance in a larger patient cohort should be performed.

# 5 Future work

DPC repair has emerged as an important and active area of research. DPCs contribute to human diseases, including neurodegeneration, aging, and cancer, but can also be exploited in chemotherapy.

Our understanding of the role of TEX264 in TOP1cc repair would be aided by a structure of TEX264 bound to SUMO1-TOP1 and p97. This could provide insight into the role of SUMO1, how TOP1ccs are processed by p97, and how this is coordinated with TOP1 cleavage by SPRTN. It will also be interesting to test if variants of TEX264 that cannot localise to the INM or bind LC3B are proficient for TOP1cc repair or not. Moreover, screens could be performed to assess whether TEX264 recruits p97 to other substrates. This could be done by comparing the p97 interactome in wild-type versus *TEX264*-knockout cells.

There are two recent manuscripts (including a pre-print manuscript on biorxiv.org) reporting a role for the yeast aspartic protease, Ddi1, in DPC repair (Serbyn et al., 2019; Svoboda et al., 2019). While there is no evidence that Ddi1 can

proteolytically cleave substrates, it does promote cellular resistance to DPC-inducing agents, including CPT, in a manner that depends on its putative protease active site. Ddi1 is known to interact with Cdc48 and, in mammalian cells, has a role in removing non-covalent DNA-bound protein obstacles to facilitate DNA replication fork restart (Kama et al., 2018; Kottemann et al., 2018). It will therefore be important to address if human Ddi1 is a protease, whether it is involved in DPC repair, and if its function is linked to p97.

Future work should focus on the relative contribution of different pathways and PTMs to DPC repair. This could help prioritise the development of inhibitors of DPC repair factors which could synergise with chemotherapy drugs that are already in clinical use. The TEX264-p97-SPRTN complex would represent a viable target since these proteins are important for cellular resistance to chemotherapy-relevant doses of CPT and SPRTN expression correlates with response to irinotecan-based chemotherapy. The identification and optimisation of inhibitors of SPRTN's protease activity could therefore prove a very effective and selective means of treating cancers.

# References


Ahnesorg, P., Smith, P., and Jackson, S.P. (2006). XLF interacts with the XRCC4-DNA Ligase IV complex to promote DNA nonhomologous end-joining. Cell *124*, 301–313.

Alagoz, M., Chiang, S.C., Sharma, A., and El-Khamisy, S.F. (2013). ATM Deficiency Results in Accumulation of DNA-Topoisomerase I Covalent Intermediates in Neural Cells. PLoS One *8*.

Amunugama, R., Willcox, S., Wu, R.A., Abdullah, U.B., El-Sagheer, A.H., Brown, T., McHugh, P.J., Griffith, J.D., and Walter, J.C. (2018). Replication Fork Reversal during DNA Interstrand Crosslink Repair Requires CMG Unloading. Cell Rep. *23*, 3419–3428.

An, H., Ordureau, A., Paulo, J.A., Shoemaker, C.J., Denic, V., and Harper, J.W. (2019). TEX264 Is an Endoplasmic Reticulum-Resident ATG8-Interacting Protein Critical for ER Remodeling during Nutrient Stress. Mol. Cell *74*, 891–908.e10.

Anderson, D.J., Le Moigne, R., Djakovic, S., Kumar, B., Rice, J., Wong, S., Wang, J., Yao, B., Valle, E., Kiss von Soly, S., et al. (2015). Targeting the AAA ATPase p97 as an Approach to Treat Cancer through Disruption of Protein Homeostasis. Cancer Cell *28*, 653–665.

Aparicio, T., Baer, R., Gottesman, M., and Gautier, J. (2016). MRN, CtIP, and BRCA1 mediate repair of topoisomerase II-DNA adducts. J. Cell Biol. *212*, 399–408.

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

Balakirev, M.Y., Mullally, J.E., Favier, A., Assard, N., Sulpice, E., Lindsey, D.F., Rulina, A. V., Gidrol, X., and Wilkinson, K.D. (2015). Wss1 metalloprotease partners with Cdc48/Doa1 in processing genotoxic SUMO conjugates. Elife *4*.

Balestrieri, E., Zanier, R., and Degrassi, F. (2001). Molecular characterisation of camptothecin-induced mutations at the hprt locus in Chinese hamster cells. Mutat. Res. - Fundam. Mol. Mech. Mutagen. *476*, 63–69.

Bass, T.E., Luzwick, J.W., Kavanaugh, G., Carroll, C., Dungrawala, H., Glick, G.G., Feldkamp, M.D., Putney, R., Chazin, W.J., and Cortez, D. (2016). ETAA1 acts at stalled replication forks to maintain genome integrity. Nat. Cell Biol. *18*, 1185–1195.

Baxter, J., and Diffley, J.F.X. (2008). Topoisomerase II Inactivation Prevents the Completion of DNA Replication in Budding Yeast. Mol. Cell *30*, 790–802.

Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011). The quantitative proteome of a human cell line. Mol. Syst. Biol. *7*.

Behan, F.M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C.M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., et al. (2019). Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. Nature *568*, 511–516.

Bennett, C.B., Lewis, A.L., Baldwin, K.K., and Resnick, M.A. (1993). Lethality induced by a single site-specific double-strand break in a dispensable yeast plasmid. Proc. Natl. Acad. Sci. U. S. A. *90*, 5613–5617.

Bergink, S., Ammon, T., Kern, M., Schermelleh, L., Leonhardt, H., and Jentsch, S.

(2013). Role of Cdc48/p97 as a SUMO-targeted segregase curbing Rad51-Rad52 interaction. Nat. Cell Biol. *15*, 526–532.

Beskow, A., Grimberg, K.B., Bott, L.C., Salomons, F.A., Dantuma, N.P., and Young, P. (2009). A Conserved Unfoldase Activity for the p97 AAA-ATPase in Proteasomal Degradation. J. Mol. Biol. *394*, 732–746.

Beucher, A., Birraux, J., Tchouandong, L., Barton, O., Shibata, A., Conrad, S., Goodarzi, A.A., Krempler, A., Jeggo, P.A., and Löbrich, M. (2009). ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. EMBO J. *28*, 3413–3427.

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

Blythe, E.E., Olson, K.C., Chau, V., and Deshaies, R.J. (2017). Ubiquitin- and ATP-dependent unfoldase activity of P97/VCP•NPLOC4•UFD1L is enhanced by a mutation that causes multisystem proteinopathy. Proc. Natl. Acad. Sci. *114*, E4380–E4388.

Bodnar, N., and Rapoport, T. (2017a). Toward an understanding of the Cdc48/p97 ATPase. F1000Research *6*, 1318.

Bodnar, N.O., and Rapoport, T.A. (2017b). Molecular Mechanism of Substrate Processing by the Cdc48 ATPase Complex. Cell *169*, 722–735.e9.

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

Borgermann, N., Ackermann, L., Schwertman, P., Hendriks, I.A., Thijssen, K., Liu, J.C., Lans, H., Nielsen, M.L., and Mailand, N. (2019). SUMO ylation promotes protective responses to DNA -protein crosslinks . EMBO J. *38*.

Bouwman, P., Aly, A., Escandell, J.M., Pieterse, M., Bartkova, J., Van Der Gulden, H., Hiddingh, S., Thanasoula, M., Kulkarni, A., Yang, Q., et al. (2010). 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. Nat. Struct. Mol. Biol. *17*, 688–695.

Bowman, G.D., O'Donnell, M., and Kuriyan, J. (2004). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. Nature *429*, 724–730.

Braun, M.S., Richman, S.D., Quirke, P., Daly, C., Adlard, J.W., Elliott, F., Barrett, J.H., Selby, P., Meade, A.M., Stephens, R.J., et al. (2008). Predictive biomarkers of chemotherapy efficacy in colorectal cancer: Results from the UK MRC FOCUS trial. J. Clin. Oncol. *26*, 2690–2698.

Brown, J.S., O'Carrigan, B., Jackson, S.P., and Yap, T.A. (2017). Targeting DNA repair in cancer: Beyond PARP inhibitors. Cancer Discov. *7*, 20–37.

Bruner, S.D., Norman, D.P.G., and Verdine, G.L. (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. Nature *403*, 859–866.

Buchberger, A., Schindelin, H., and Hänzelmann, P. (2015). Control of p97 function by cofactor binding. FEBS Lett. *589*, 2578–2589.

Budzowska, M., Zhang, J., Budzowska, M., Drohat, A.C., and Walter, J.C. (2016). Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. Cell *167*, 498–511.e14.

Bunting, S.F., Callén, E., Wong, N., Chen, H.T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., et al. (2010). 53BP1 inhibits

homologous recombination in brca1-deficient cells by blocking resection of DNA breaks. Cell *141*, 243–254.

Burgers, P.M.J. (2009). Polymerase dynamics at the eukaryotic DNA replication fork. J. Biol. Chem. *284*, 4041–4045.

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

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

Cappelli, E., Taylor, R., Cevasco, M., Abbondandolo, A., Caldecott, K., and Frosina, G. (1997). Involvement of XRCC1 and DNA ligase III gene products in DNA base excision repair. J. Biol. Chem. *272*, 23970–23975.

Carmell, M.A., Dokshin, G.A., Skaletsky, H., Hu, Y.C., van Wolfswinkel, J.C., Igarashi, K.J., Bellott, D.W., Nefedov, M., Reddien, P.W., Enders, G.C., et al. (2016). A widely employed germ cell marker is an ancient disordered protein with reproductive functions in diverse eukaryotes. Elife *5*.

Centore, R.C., Yazinski, S.A., Tse, A., and Zou, L. (2012). Spartan/C1orf124, a Reader of PCNA Ubiquitylation and a Regulator of UV-Induced DNA Damage Response. Mol. Cell *46*, 625–635.

Chapman, J.R., Sossick, A.J., Boulton, S.J., and Jackson, S.P. (2012). BRCA1-associated exclusion of 53BP1 from DNA: Damage sites underlies temporal control of DNA repair. J. Cell Sci. *125*, 3529–3534.

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

Chino, H., Hatta, T., Natsume, T., and Mizushima, N. (2019). Intrinsically Disordered Protein TEX264 Mediates ER-phagy. Mol. Cell *74*, 909–921.e6.

Chválová, K., Brabec, V., and Kašpárková, J. (2007). Mechanism of the formation of DNA-protein cross-links by antitumor cisplatin. Nucleic Acids Res. *35*, 1812–1821.

Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A.R., Laghmani, E.H., Joenje, H., McDonald, N., de Winter, J.P., et al. (2007). Identification of FAAP24, a Fanconi Anemia Core Complex Protein that Interacts with FANCM. Mol. Cell *25*, 331–343.

Clauson, C., Schärer, O.D., and Niedernhofer, L. (2013). Advances in understanding the complex mechanisms of DNA inter strand cross-link repair. Cold Spring Harb. Perspect. Med. *3*.

Cleaver, J.E. (2005). Cancer in xeroderma pigmentosum and related disorders of DNA repair. Nat. Rev. Cancer *5*, 564–573.

Compe, E., and Egly, J.M. (2012). TFIIH: When transcription met DNA repair. Nat. Rev. Mol. Cell Biol. *13*, 343–354.

Cortez, D., Guntuku, S., Qin, J., and Elledge, S.J. (2001). ATR and ATRIP: Partners in checkpoint signaling. Science (80-. ). *294*, 1713–1716.

Cotta-Ramusino, C., McDonald, E.R., Hurov, K., Sowa, M.E., Harper, J.W., and Elledge, S.J. (2011). A DNA damage response screen identifies RHINO, a 9-1-1 and

topBP1 interacting protein required for ATR signaling. Science (80-. ). *332*, 1313–1317.

Crossan, G., Langevin, F.P., Arends, M.J., and Patel, K.J. (2012). The fanconi pathway counteracts the toxic effects of naturally produced aldehydes. J. Pathol. *226*, S5.

Dalal, S., Rosser, M.F.N., Cyr, D.M., and Hanson, P.I. (2004). Distinct Roles for the AAA ATPases NSF and p97 in the Secretory Pathway. Mol. Biol. Cell *15*, 637–648.

Davis, E.J., Lachaud, C., Appleton, P., MacArtney, T.J., Näthke, I., and Rouse, J. (2012). DVC1 (C1orf124) recruits the p97 protein segregase to sites of DNA damage. Nat. Struct. Mol. Biol. *19*, 1093–1100.

Delabaere, L., Orsi, G.A., Sapey-Triomphe, L., Horard, B., Couble, P., and Loppin, B. (2014). The spartan ortholog maternal haploid is required for paternal chromosome integrity in the drosophila zygote. Curr. Biol. *24*, 2281–2287.

Delacroix, S., Wagner, J.M., Kobayashi, M., Yamamoto, K.I., and Karnitz, L.M. (2007). The Rad9-Hus1-Rad1 (9-1-1) clamp activates checkpoint signaling via TopBP1. Genes Dev. *21*, 1472–1477.

Demott, M.S., Beyret, E., Wong, D., Bales, B.C., Hwang, J.T., Greenberg, M.M., and Demple, B. (2002). Covalent trapping of human DNA polymerase β by the oxidative DNA lesion 2-deoxyribonolactone. J. Biol. Chem. *277*, 7637–7640.

Desai, S.D., Liu, L.F., Vazquez-Abad, D., and D'Arpa, P. (1997). Ubiquitin-dependent destruction of topoisomerase I is stimulated by the antitumor drug camptothecin. J. Biol. Chem. *272*, 24159–24164.

Dev, H., Chiang, T.W.W., Lescale, C., de Krijger, I., Martin, A.G., Pilger, D., Coates, J., Sczaniecka-Clift, M., Wei, W., Ostermaier, M., et al. (2018). Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. Nat. Cell Biol. *20*, 954–965.

Dewar, J.M., Budzowska, M., and Walter, J.C. (2015). The mechanism of DNA replication termination in vertebrates. Nature *525*, 345–350.

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

Dianov, G.L., and Hübscher, U. (2013). Mammalian base excision repair: The forgotten archangel. Nucleic Acids Res. *41*, 3483–3490.

Ding, X., Chaudhuri, A.R., Callen, E., Pang, Y., Biswas, K., Klarmann, K.D., Martin, B.K., Burkett, S., Cleveland, L., Stauffer, S., et al. (2016). Synthetic viability by BRCA2 and PARP1/ARTD1 deficiencies. Nat. Commun. *7*.

Doil, C., Mailand, N., Bekker-Jensen, S., Menard, P., Larsen, D.H., Pepperkok, R., Ellenberg, J., Panier, S., Durocher, D., Bartek, J., et al. (2009). RNF168 Binds and Amplifies Ubiquitin Conjugates on Damaged Chromosomes to Allow Accumulation of Repair Proteins. Cell *136*, 435–446.

Dong, H., Nebert, D.W., Bruford, E.A., Thompson, D.C., Joenje, H., and Vasiliou, V. (2015). Update of the human and mouse fanconi anemia genes. Hum. Genomics *9*.

Duxin, J.P., and Walter, J.C. (2015). What is the DNA repair defect underlying Fanconi anemia? Curr. Opin. Cell Biol. *37*, 49–60.

Duxin, J.P., Dewar, J.M., Yardimci, H., and Walter, J.C. (2014). Repair of a DNA-protein crosslink by replication-coupled proteolysis. Cell *159*, 349–357.

Dye, B.T., and Schulman, B.A. (2007). Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins. Annu. Rev. Biophys. Biomol. Struct. *36*, 131–150.

Enenkel, C., Lehmann, A., and Kloetzel, P.M. (1999). GFP-labelling of 26S proteasomes in living yeast: Insight into proteasomal functions at the nuclear envelope/rough ER. Mol. Biol. Rep. *26*, 131–135.

Escribano-Díaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J.T.F., Tkáč, J., Cook, M.A., Rosebrock, A.P., Munro, M., Canny, M.D., et al. (2013). A Cell Cycle-Dependent Regulatory Circuit Composed of 53BP1-RIF1 and BRCA1-CtIP Controls DNA Repair Pathway Choice. Mol. Cell *49*, 872–883.

Fachinetti, D., Bermejo, R., Cocito, A., Minardi, S., Katou, Y., Kanoh, Y., Shirahige, K., Azvolinsky, A., Zakian, V.A., and Foiani, M. (2010). Replication Termination at Eukaryotic Chromosomes Is Mediated by Top2 and Occurs at Genomic Loci Containing Pausing Elements. Mol. Cell *39*, 595–605.

Fielden, J., Ruggiano, A., Popović, M., and Ramadan, K. (2018). DNA protein crosslink proteolysis repair: From yeast to premature ageing and cancer in humans. DNA Repair (Amst). *71*, 198–204.

Fortini, P., and Dogliotti, E. (2007). Base damage and single-strand break repair: Mechanisms and functional significance of short- and long-patch repair subpathways. DNA Repair (Amst). *6*, 398–409.

Fousteri, M., Vermeulen, W., van Zeeland, A.A., and Mullenders, L.H.F. (2006). Cockayne Syndrome A and B Proteins Differentially Regulate Recruitment of Chromatin Remodeling and Repair Factors to Stalled RNA Polymerase II In Vivo. Mol. Cell *23*, 471–482.

Franz, A., Ackermann, L., and Hoppe, T. (2016a). Ring of change: CDC48/p97 drives protein dynamics at chromatin. Front. Genet. *7*.

Franz, A., Pirson, P.A., Pilger, D., Halder, S., Achuthankutty, D., Kashkar, H., Ramadan, K., and Hoppe, T. (2016b). Chromatin-associated degradation is defined by UBXN-3/FAF1 to safeguard DNA replication fork progression. Nat. Commun. *7*.

Friedberg, E.C., Wagner, R., and Radman, M. (2002). Specialized DNA polymerases, cellular survival, and the genesis of mutations. Science (80-. ). *296*, 1627–1630.

Gaillard, H., García-Muse, T., and Aguilera, A. (2015). Replication stress and cancer. Nat. Rev. Cancer *15*, 276–280.

Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D., and Labib, K. (2006). GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat. Cell Biol. *8*, 358–366.

Gao, R., Schellenberg, M.J., Huang, S.Y.N., Abdelmalak, M., Marchand, C., Nitiss, K.C., Nitiss, J.L., Williams, R.S., and Pommier, Y. (2014). Proteolytic degradation of topoisomerase II (Top2) enables the processing of Top2·DNA and Top2·RNA covalent complexes by tyrosyl-DNA-phosphodiesterase 2 (TDP2). J. Biol. Chem. *289*, 17960–17969.

Garaycoechea, J.I., Crossan, G.P., Langevin, F., Daly, M., Arends, M.J., and Patel, K.J. (2012). Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. Nature.

Ghezraoui, H., Oliveira, C., Becker, J.R., Bilham, K., Moralli, D., Anzilotti, C., Fischer,

R., Deobagkar-Lele, M., Sanchiz-Calvo, M., Fueyo-Marcos, E., et al. (2018). 53BP1 cooperation with the REV7–shieldin complex underpins DNA structure-specific NHEJ. Nature *560*, 122–127.

Ghosal, G., Leung, J.W.C., Nair, B.C., Fong, K.W., and Chen, J. (2012). Proliferating Cell Nuclear Antigen (PCNA)-binding protein C1orf124 is a regulator of translesion synthesis. J. Biol. Chem. *287*, 34225–34233.

Gibbs-Seymour, I., Oka, Y., Rajendra, E., Weinert, B.T., Passmore, L.A., Patel, K.J., Olsen, J. V., Choudhary, C., Bekker-Jensen, S., and Mailand, N. (2015). Ubiquitin-SUMO circuitry controls activated fanconi anemia ID complex dosage in response to DNA damage. Mol. Cell *57*, 150–164.

Gilbert, D.C., Chalmers, A.J., and El-Khamisy, S.F. (2012). Topoisomerase i inhibition in colorectal cancer: Biomarkers and therapeutic targets. Br. J. Cancer *106*, 18–24.

Gómez-Herreros, F., Schuurs-Hoeijmakers, J.H.M., McCormack, M., Greally, M.T., Rulten, S., Romero-Granados, R., Counihan, T.J., Chaila, E., Conroy, J., Ennis, S., et al. (2014). TDP2 protects transcription from abortive topoisomerase activity and is required for normal neural function. Nat. Genet. *46*, 516–521.

Goodarzi, A.A., Yu, Y., Riballo, E., Douglas, P., Walker, S.A., Ye, R., Härer, C., Marchetti, C., Morrice, N., Jeggo, P.A., et al. (2006). DNA-PK autophosphorylation facilitates Artemis endonuclease activity. EMBO J. *25*, 3880–3889.

Goodarzi, A.A., Noon, A.T., and Jeggo, P.A. (2009). The impact of heterochromatin on DSB repair. In Biochemical Society Transactions, pp. 569–576.

Gottlieb, T.M., and Jackson, S.P. (1993). The DNA-dependent protein kinase: Requirement for DNA ends and association with Ku antigen. Cell *72*, 131–142.

de Graaf, B., Clore, A., and McCullough, A.K. (2009). Cellular pathways for DNA repair and damage tolerance of formaldehyde-induced DNA-protein crosslinks. DNA Repair (Amst). *8*, 1207–1214.

Gupta, R., Somyajit, K., Narita, T., Maskey, E., Stanlie, A., Kremer, M., Typas, D., Lammers, M., Mailand, N., Nussenzweig, A., et al. (2018). DNA Repair Network Analysis Reveals Shieldin as a Key Regulator of NHEJ and PARP Inhibitor Sensitivity. Cell *173*, 972–988.e23.

Haahr, P., Hoffmann, S., Tollenaere, M.A.X., Ho, T., Toledo, L.I., Mann, M., Bekker-Jensen, S., Räschle, M., and Mailand, N. (2016). Activation of the ATR kinase by the RPA-binding protein ETAA1. Nat. Cell Biol. *18*, 1196–1207.

Halder, S., Torrecilla, I., Burkhalter, M.D., Popović, M., Fielden, J., Vaz, B., Oehler, J., Pilger, D., Lessel, D., Wiseman, K., et al. (2019). SPRTN protease and checkpoint kinase 1 cross-activation loop safeguards DNA replication. Nat. Commun. *10*.

Hande, K.R. (1998). Etoposide: Four decades of development of a topoisomerase II inhibitor. Eur. J. Cancer *34*, 1514–1521.

Hänzelmann, P., and Schindelin, H. (2016). Characterization of an Additional Binding Surface on the p97 N-Terminal Domain Involved in Bipartite Cofactor Interactions. Structure *24*, 140–147.

Hänzelmann, P., Buchberger, A., and Schindelin, H. (2011). Hierarchical binding of cofactors to the AAA ATPase p97. Structure *19*, 833–843.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-Resolution

CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell *163*, 1515–1526.

Hashimoto, H., Chatterjee, S., and Berger, N.A. (1995). Mutagenic Activity of Topoisomerase I Inhibitors. Clin. Cancer Res. *1*, 369–376.

Heck, H.D.A., and Casanova, M. (2004). The implausibility of leukemia induction by formaldehyde: A critical review of the biological evidence on distant-site toxicity. Regul. Toxicol. Pharmacol.

Hedglin, M., and Benkovic, S.J. (2015). Regulation of Rad6/Rad18 Activity During DNA Damage Tolerance. Annu. Rev. Biophys. *44*, 207–228.

Heideker, J., Prudden, J., Perry, J.J.P., Tainer, J.A., and Boddy, M.N. (2011). SUMO-targeted ubiquitin ligase, Rad60, and Nse2 SUMO ligase suppress spontaneous Top1-mediated DNA damage and genome instability. PLoS Genet. *7*.

Hishida, T., Kubota, Y., Carr, A.M., and Iwasaki, H. (2009). RAD6-RAD18-RAD5-pathway-dependent tolerance to chronic low-dose ultraviolet light. Nature *457*, 612–615.

Hoa, N.N., Shimizu, T., Zhou, Z.W., Wang, Z.Q., Deshpande, R.A., Paull, T.T., Akter, S., Tsuda, M., Furuta, R., Tsusui, K., et al. (2016). Mre11 Is Essential for the Removal of Lethal Topoisomerase 2 Covalent Cleavage Complexes. Mol. Cell *64*, 580–592.

Hoege, C., Pfander, B., Moldovan, G.L., Pyrowolakis, G., and Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature *419*, 135–141.

Horie, K., Tomida, A., Sugimoto, Y., Yasugi, T., Yoshikawa, H., Taketani, Y., and Tsuruo, T. (2002). SUMO-1 conjugation to intact DNA topoisomerase I amplifies cleavable complex formation induced by camptothecin. Oncogene *21*, 7913–7922.

Hsiang, Y.H., Hertzberg, R., Hecht, S., and Liu, L.F. (1985). Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. J. Biol. Chem. *260*, 14873–14878.

Ide, H., Nakano, T., Salem, A.M.H., and Shoulkamy, M.I. (2018). DNA–protein cross-links: Formidable challenges to maintaining genome integrity. DNA Repair (Amst). *71*, 190–197.

Inano, S., Sato, K., Katsuki, Y., Kobayashi, W., Tanaka, H., Nakajima, K., Nakada, S., Miyoshi, H., Knies, K., Takaori-Kondo, A., et al. (2017). RFWD3-Mediated Ubiquitination Promotes Timely Removal of Both RPA and RAD51 from DNA Damage Sites to Facilitate Homologous Recombination. Mol. Cell *66*, 622–634.e8.

Interthal, H., and Champoux, J.J. (2011). Effects of DNA and protein size on substrate cleavage by human tyrosyl-DNA phosphodiesterase 1. Biochem. J. *436*, 559–566.

Interthal, H., Chen, H.J., Kehl-Fie, T.E., Zotzmann, J., Leppard, J.B., and Champoux, J.J. (2005a). SCAN1 mutant Tdp1 accumulates the enzyme-DNA intermediate and causes camptothecin hypersensitivity. EMBO J. *24*, 2224–2233.

Interthal, H., Chen, H.J., and Champoux, J.J. (2005b). Human Tdp1 cleaves a broad spectrum of substrates, including phosphoamide linkages. J. Biol. Chem. *280*, 36518–36528.

Jazayeri, A., Falck, J., Lukas, C., Bartek, J., Smith, G.C.M., Lukas, J., and Jackson, S.P. (2006). ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. Nat. Cell Biol. *8*, 37–45.

Jeggo, P.A., and Löbrich, M. (2005). Artemis links ATM to double strand break rejoining. Cell Cycle *4*, 359–362.

Juhasz, S., Balogh, D., Hajdu, I., Burkovics, P., Villamil, M.A., Zhuang, Z., and Haracska, L. (2012). Characterization of human Spartan/C1orf124, an ubiquitin-PCNA interacting regulator of DNA damage tolerance. Nucleic Acids Res. *40*, 10795–10808.

Kalocsay, M., Hiller, N.J., and Jentsch, S. (2009). Chromosome-wide Rad51 Spreading and SUMO-H2A.Z-Dependent Chromosome Fixation in Response to a Persistent DNA Double-Strand Break. Mol. Cell *33*, 335–343.

Kama, R., Gabriely, G., Kanneganti, V., Gerst, J.E., and Gilmore, R. (2018). Cdc48 and ubiquilins confer selective anterograde protein sorting and entry into the multivesicular body in yeast. Mol. Biol. Cell *29*, 948–963.

Karaca, E., Tozluoğlu, M., Nussinov, R., and Haliloğlu, T. (2011). Alternative allosteric mechanisms can regulate the substrate and E2 in SUMO conjugation. J. Mol. Biol. *406*, 620–630.

Katyal, S., Lee, Y., Nitiss, K.C., Downing, S.M., Li, Y., Shimada, M., Zhao, J., Russell, H.R., Petrini, J.H.J., Nitiss, J.L., et al. (2014). Aberrant topoisomerase-1 DNA lesions are pathogenic in neurodegenerative genome instability syndromes. Nat. Neurosci. *17*, 813–821.

Kawale, A.S., Akopiants, K., Valerie, K., Ruis, B., Hendrickson, E.A., Huang, S. yin N., Pommier, Y., and Povirk, L.F. (2018). TDP1 suppresses mis-joining of radiomimetic DNA double-strand breaks and cooperates with Artemis to promote optimal nonhomologous end joining. Nucleic Acids Res. *46*, 8926–8939.

Kawamura, K., Bahar, R., Seimiya, M., Chiyo, M., Wada, A., Okada, S., Hatano, M., Tokuhisa, T., Kimura, H., Watanabe, S., et al. (2004). DNA polymerase θ is preferentially expressed in lymphoid tissues and upregulated in human cancers. Int. J. Cancer *109*, 9–16.

Khaminets, A., Heinrich, T., Mari, M., Grumati, P., Huebner, A.K., Akutsu, M., Liebmann, L., Stolz, A., Nietzsche, S., Koch, N., et al. (2015). Regulation of endoplasmic reticulum turnover by selective autophagy. Nature *522*, 354–358.

Kiianitsa, K., and Maizels, N. (2013). A rapid and sensitive assay for DNA-protein covalent complexes in living cells. Nucleic Acids Res. *41*.

Kim, M.S., Machida, Y., Vashisht, A.A., Wohlschlegel, J.A., Pang, Y.P., and Machida, Y.J. (2013). Regulation of error-prone translesion synthesis by Spartan/C1orf124. Nucleic Acids Res. *41*, 1661–1668.

Koegl, M., Hoppe, T., Schlenker, S., Ulrich, H.D., Mayer, T.U., and Jentsch, S. (1999). A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. Cell *96*, 635–644.

Kolas, N.K., Chapman, J.R., Nakada, S., Ylanko, J., Chahwan, R., Sweeney, F.D., Panier, S., Mendez, M., Wildenhain, J., Thomson, T.M., et al. (2007). Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. Science (80-. ). *318*, 1637–1640.

Koopman, M., Venderbosch, S., van Tinteren, H., Ligtenberg, M.J., Nagtegaal, I., Van Krieken, J.H., and Punt, C.J. (2009). Predictive and prognostic markers for the outcome of chemotherapy in advanced colorectal cancer, a retrospective analysis of the phase III randomised CAIRO study. Eur. J. Cancer *45*, 1999–2006.

Koster, D.A., Palle, K., Bot, E.S.M., Bjornsti, M.A., and Dekker, N.H. (2007). Antitumour

drugs impede DNA uncoiling by topoisomerase I. Nature *448*, 213–217.

Kottemann, M.C., Conti, B.A., Lach, F.P., and Smogorzewska, A. (2018). Removal of RTF2 from Stalled Replisomes Promotes Maintenance of Genome Integrity. Mol. Cell *69*, 24–35.e5.

Kulathu, Y., and Komander, D. (2012). Atypical ubiquitylation-the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. Nat. Rev. Mol. Cell Biol. *13*, 508–523.

Kumar, A., Mazzanti, M., Mistrik, M., Kosar, M., Beznoussenko, G. V., Mironov, A.A., Garrè, M., Parazzoli, D., Shivashankar, G. V., Scita, G., et al. (2014). ATR mediates a checkpoint at the nuclear envelope in response to mechanical stress. Cell *158*, 633–646.

Kumari, A., Lim, Y.X., Newell, A.H., Olson, S.B., and McCullough, A.K. (2012). Formaldehyde-induced genome instability is suppressed by an XPF-dependent pathway. DNA Repair (Amst). *11*, 236–246.

Langevin, F., Crossan, G.P., Rosado, I. V., Arends, M.J., and Patel, K.J. (2011). Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. Nature *475*, 53–59.

Larsen, N.B., Gao, A.O., Sparks, J.L., Gallina, I., Wu, R.A., Mann, M., Räschle, M., Walter, J.C., and Duxin, J.P. (2019). Replication-Coupled DNA-Protein Crosslink Repair by SPRTN and the Proteasome in Xenopus Egg Extracts. Mol. Cell *73*, 574–588.e7.

Latterich, M., Fröhlich, K.U., and Schekman, R. (1995). Membrane fusion and the cell cycle: Cdc48p participates in the fusion of ER membranes. Cell *82*, 885–893.

Ledesma, F.C., El Khamisy, S.F., Zuma, M.C., Osborn, K., and Caldecott, K.W. (2009). A human 5′-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. Nature *461*, 674–678.

Lee, J.H., and Paull, T.T. (2004). Direct Activation of the ATM Protein Kinase by the Mre11/Rad50/Nbs1 Complex. Science (80-. ). *304*, 93–96.

Lee, Y.C., Zhou, Q., Chen, J., and Yuan, J. (2016). RPA-Binding Protein ETAA1 Is an ATR Activator Involved in DNA Replication Stress Response. Curr. Biol. *26*, 3257–3268.

Lessel, D., Vaz, B., Halder, S., Lockhart, P.J., Marinovic-Terzic, I., Lopez-Mosqueda, J., Philipp, M., Sim, J.C.H., Smith, K.R., Oehler, J., et al. (2014). Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features. Nat. Genet. *46*, 1239–1244.

Li, F., Raczynska, J.E., Chen, Z., and Yu, H. (2019). Structural Insight into DNA-Dependent Activation of Human Metalloprotease Spartan. Cell Rep. *26*, 3336–3346.e4.

Li, M., Pokharel, S., Wang, J.T., Xu, X., and Liu, Y. (2015). RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. Nat. Commun. *6*.

Liang, C.C., Zhan, B., Yoshikawa, Y., Haas, W., Gygi, S.P., and Cohn, M.A. (2015). UHRF1 Is a sensor for DNA interstrand crosslinks and recruits FANCD2 to initiate the Fanconi Anemia pathway. Cell Rep. *10*, 1947–1956.

Lin, C.P., Ban, Y., Lyu, Y.L., Desai, S.D., and Liu, L.F. (2008). A ubiquitin-proteasome pathway for the repair of topoisomerase I-DNA covalent complexes. J. Biol. Chem. *283*, 21074–21083.

Lindahl, T. (1979). DNA Glycosylases, Endonucleases for Apurinic/Apyrimidinic Sites,

and Base Excision-Repair. Prog. Nucleic Acid Res. Mol. Biol. *22*, 135–192.

Lopez-Mosqueda, J., Maddi, K., Prgomet, S., Kalayil, S., Marinovic-Terzic, I., Terzic, J., and Dikic, I. (2016). SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. Elife *5*.

Lu, K., Den Brave, F., and Jentsch, S. (2017). Receptor oligomerization guides pathway choice between proteasomal and autophagic degradation. Nat. Cell Biol. *19*, 732–739.

Machida, Y., Kim, M.S., and Machida, Y.J. (2012). Spartan/C1orf124 is important to prevent UV-induced mutagenesis. Cell Cycle *11*, 3395–3402.

Mailand, N., Falck, J., Lukas, C., Syljuåsen, R.G., Welcker, M., Bartek, J., and Lukas, J. (2000). Rapid destruction of human Cdc25A in response to DNA damage. Science (80-. ). *288*, 1425–1429.

Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C., and Lukas, J. (2007). RNF8 Ubiquitylates Histones at DNA Double-Strand Breaks and Promotes Assembly of Repair Proteins. Cell *131*, 887–900.

Maloisel, L., Fabre, F., and Gangloff, S. (2008). DNA Polymerase  Is Preferentially Recruited during Homologous Recombination To Promote Heteroduplex DNA Extension. Mol. Cell. Biol. *28*, 1373–1382.

Manzo, S.G., Hartono, S.R., Sanz, L.A., Marinello, J., De Biasi, S., Cossarizza, A., Capranico, G., and Chedin, F. (2018). DNA Topoisomerase I differentially modulates R-loops across the human genome. Genome Biol. *19*.

Mao, Y., Desai, S.D., and Liu, L.F. (2000b). SUMO-1 conjugation to human DNA topoisomerase II isozymes. J. Biol. Chem. *275*, 26066–26073.

Mao, Y., Sun, M., Desai, S.D., and Liu, L.F. (2000a). SUMO-1 conjugation to topoisomerase I: A possible repair response to topoisomerase-mediated DNA damage. Proc. Natl. Acad. Sci. U. S. A. *97*, 4046–4051.

Mao, Y., Desai, S.D., Ting, C.Y., Hwang, J., and Liu, L.F. (2001). 26 S Proteasome-mediated Degradation of Topoisomerase II Cleavable Complexes. J. Biol. Chem. *276*, 40652–40658.

Maric, M., Maculins, T., De Piccoli, G., and Labib, K. (2014). Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. Science (80-. ). *346*.

Maric, M., Mukherjee, P., Tatham, M.H., Hay, R., and Labib, K. (2017). Ufd1-Npl4 Recruit Cdc48 for Disassembly of Ubiquitylated CMG Helicase at the End of Chromosome Replication. Cell Rep. *18*, 3033–3042.

Maskey, R.S., Kim, M.S., Baker, D.J., Childs, B., Malureanu, L.A., Jeganathan, K.B., Machida, Y., Van Deursen, J.M., and Machida, Y.J. (2014). Spartan deficiency causes genomic instability and progeroid phenotypes. Nat. Commun. *5*.

Maskey, R.S., Flatten, K.S., Sieben, C.J., Peterson, K.L., Baker, D.J., Nam, H.J., Kim, M.S., Smyrk, T.C., Kojima, Y., Machida, Y., et al. (2017). Spartan deficiency causes accumulation of Topoisomerase 1 cleavage complexes and tumorigenesis. Nucleic Acids Res. *45*, 4564–4576.

Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999). The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. Nature *399*, 700–704.

Maya-Mendoza, A., Moudry, P., Merchut-Maya, J.M., Lee, M., Strauss, R., and Bartek,

J. (2018). High speed of fork progression induces DNA replication stress and genomic instability. Nature *559*, 279–284.

McCulloch, S.D., and Kunkel, T.A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. Cell Res. *18*, 148–161.

McCulloch, S.D., Kokoska, R.J., Chilkova, O., Welch, C.M., Johansson, E., Burgers, P.M.J., and Kunkel, T.A. (2004). Enzymatic switching for efficient and accurate translesion DNA replication. Nucleic Acids Res. *32*, 4665–4675.

McVey, M., Khodaverdian, V.Y., Meyer, D., Cerqueira, P.G., and Heyer, W.-D. (2016). Eukaryotic DNA Polymerases in Homologous Recombination. Annu. Rev. Genet. *50*, 393–421.

Meerang, M., Ritz, D., Paliwal, S., Garajova, Z., Bosshard, M., Mailand, N., Janscak, P., Hübscher, U., Meyer, H., and Ramadan, K. (2011). The ubiquitin-selective segregase VCP/p97 orchestrates the response to DNA double-strand breaks. Nat. Cell Biol. *13*, 1376–1382.

Melis, J.P.M., Van Steeg, H., and Luijten, M. (2013). Oxidative DNA damage and nucleotide excision repair. Antioxidants Redox Signal. *18*, 2409–2419.

Meselson, M., and Stahl, F.W. (1958). The replication of DNA in Escherichia coli. Proc. Natl. Acad. Sci. *44*, 671–682.

Meyer, H., Bug, M., and Bremer, S. (2012). Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. Nat. Cell Biol. *14*, 117–123.

Michelena, J., Lezaja, A., Teloni, F., Schmid, T., Imhof, R., and Altmeyer, M. (2018). Analysis of PARP inhibitor toxicity by multidimensional fluorescence microscopy reveals mechanisms of sensitivity and resistance. Nat. Commun. *9*.

Michelena, J., Gatti, M., Teloni, F., Imhof, R., and Altmeyer, M. (2019). Basal CHK1 activity safeguards its stability to maintain intrinsic S-phase checkpoint functions. J. Cell Biol. jcb.201902085.

Ming, X., Groehler, A., Michaelson-Richie, E.D., Villalta, P.W., Campbell, C., and Tretyakova, N.Y. (2017). Mass Spectrometry Based Proteomics Study of Cisplatin-Induced DNA-Protein Cross-Linking in Human Fibrosarcoma (HT1080) Cells. Chem. Res. Toxicol. *30*, 980–995.

Moldovan, G.-L., and D'Andrea, A.D. (2009). How the Fanconi Anemia Pathway Guards the Genome. Annu. Rev. Genet. *43*, 223–249.

Moreno, S.P., Bailey, R., Campion, N., Herron, S., and Gambus, A. (2014). Polyubiquitylation drives replisome disassembly at the termination of DNA replication. Science (80-. ). *346*, 477–481.

Mórocz, M., Zsigmond, E., Tóth, R., Zs Enyedi, M., Pintér, L., and Haracska, L. (2016). DNA-dependent protease activity of human Spartan facilitates replication of DNA-protein crosslink-containing DNA. Nucleic Acids Res. *45*, 3172–3188.

Mosbech, A., Gibbs-Seymour, I., Kagias, K., Thorslund, T., Beli, P., Povlsen, L., Nielsen, S.V., Smedegaard, S., Sedgwick, G., Lukas, C., et al. (2012). DVC1 (C1orf124) is a DNA damage-targeting p97 adaptor that promotes ubiquitin-dependent responses to replication blocks. Nat. Struct. Mol. Biol. *19*, 1084–1092.

Motnenko, A., Liang, C.C., Yang, D., Lopez-Martinez, D., Yoshikawa, Y., Zhan, B., Ward, K.E., Tian, J., Haas, W., Spingardi, P., et al. (2018). Identification of UHRF2 as a novel DNA interstrand crosslink sensor protein. PLoS Genet. *14*.

Mueller, B., Klemm, E.J., Spooner, E., Claessen, J.H., and Ploegh, H.L. (2008). SEL1L nucleates a protein complex required for dislocation of misfolded glycoproteins. Proc. Natl. Acad. Sci. U. S. A. *105*, 12325–12330.

Mullen, J.R., Chen, C.-F., and Brill, S.J. (2010). Wss1 Is a SUMO-Dependent Isopeptidase That Interacts Genetically with the Slx5-Slx8 SUMO-Targeted Ubiquitin Ligase. Mol. Cell. Biol. *30*, 3737–3748.

Murai, J., Huang, S.Y.N., Das, B.B., Renaud, A., Zhang, Y., Doroshow, J.H., Ji, J., Takeda, S., and Pommier, Y. (2012). Trapping of PARP1 and PARP2 by clinical PARP inhibitors. Cancer Res. *72*, 5588–5599.

Murai, J., Zhang, Y., Morris, J., Ji, J., Takeda, S., Doroshow, J.H., and Pommier, Y. (2014). Rationale for poly(ADP-ribose) polymerase (PARP) inhibitors in combination therapy with camptothecins or temozolomide based on PARP trapping versus catalytic inhibition. J. Pharmacol. Exp. Ther. *349*, 408–416.

Muramatsu, S., Hirai, K., Tak, Y.S., Kamimura, Y., and Araki, H. (2010). CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol ε, and GINS in budding yeast. Genes Dev. *24*, 602–612.

Nakanishi, A., Oshida, T., Matsushita, T., Imajoh-Ohmi, S., and Ohnuki, T. (1998). Identification of DNA gyrase inhibitor (GyrI) in Escherichia coli. J. Biol. Chem. *273*, 1933–1938.

Nakano, T., Morishita, S., Katafuchi, A., Matsubara, M., Horikawa, Y., Terato, H., Salem, A.M.H., Izumi, S., Pack, S.P., Makino, K., et al. (2007). Nucleotide Excision Repair and Homologous Recombination Systems Commit Differentially to the Repair of DNA-Protein Crosslinks. Mol. Cell *28*, 147–158.

Nie, M., Aslanian, A., Prudden, J., Heideker, J., Vashisht, A.A., Wohlschlegel, J.A., Yates, J.R., and Boddy, M.N. (2012). Dual recruitment of Cdc48 (p97)-Ufd1-Npl4 ubiquitin-selective segregase by small ubiquitin-like modifier protein (SUMO) and ubiquitin in SUMO-targeted ubiquitin ligase-mediated genome stability functions. J. Biol. Chem. *287*, 29610–29619.

Niedernhofer, L.J., Daniels, J.S., Rouzer, C.A., Greene, R.E., and Marnett, L.J. (2003). Malondialdehyde, a product of lipid peroxidation, is mutagenic in human cells. J. Biol. Chem. *278*, 31426–31433.

Nitiss, J.L. (2002). DNA topoisomerases in cancer chemotherapy: using enzymes to generate selective DNA damage. Curr. Opin. Investig. Drugs *3*, 1512–1516.

Noordermeer, S.M., Adam, S., Setiaputra, D., Barazas, M., Pettitt, S.J., Ling, A.K., Olivieri, M., Álvarez-Quilón, A., Moatti, N., Zimmermann, M., et al. (2018). The shieldin complex mediates 53BP1-dependent DNA repair. Nature *560*, 117–121.

Nouspikel, T. (2009). DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. Cell. Mol. Life Sci. *66*, 994–1009.

Ochi, T., Blackford, A.N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., Travers, J., Wu, Q., Draviam, V.M., Robinson, C. V., et al. (2015). PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. Science (80-. ). *347*, 185–188.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., and Sugino, A. (1968). Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. Proc. Natl. Acad. Sci. U. S. A. *59*, 598–605.

Oleinick, N.L., Chiu, S.M., Ramakrishnan, N., and Xue, L.Y. (1987). The formation,

identification, and significance of DNA-protein cross-links in mammalian cells. Br. J. Cancer. Suppl.

Ossareh-Nazari, B., Bonizec, M., Cohen, M., Dokudovskaya, S., Delalande, F., Schaeffer, C., Van Dorsselaer, A., and Dargemont, C. (2010). Cdc48 and Ufd3, new partners of the ubiquitin protease Ubp3, are required for ribophagy. EMBO Rep. *11*, 548–554.

Papadopoulos, C., Kirchner, P., Bug, M., Grum, D., Koerver, L., Schulze, N., Poehler, R., Dressler, A., Fengler, S., Arhzaouy, K., et al. (2017). VCP /p97 cooperates with YOD 1, UBXD 1 and PLAA to drive clearance of ruptured lysosomes by autophagy . EMBO J. *36*, 135–150.

Patel, A.G., Flatten, K.S., Peterson, K.L., Beito, T.G., Schneider, P.A., Perkins, A.L., Harki, D.A., and Kaufmann, S.H. (2016). Immunodetection of human topoisomerase I-DNA covalent complexes. Nucleic Acids Res. *44*, 2816–2826.

Patel, K., Dickson, J., Din, S., Macleod, K., Jodrell, D., and Ramsahoye, B. (2010). Targeting of 5-aza-2'-deoxycytidine residues by chromatin-associated DNMT1 induces proteasomal degradation of the free enzyme. Nucleic Acids Res. *38*, 4313–4324.

Pellegrini, L. (2012). The Pol α -Primase Complex. Subcell. Biochem. *62*, 157–169.

Pfeiffer, P. (2000). Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. Mutagenesis *15*, 289–302.

Pfeiffer, A., Luijsterburg, M.S., Acs, K., Wiegant, W.W., Helfricht, A., Herzog, L.K., Minoia, M., Böttcher, C., Salomons, F.A., Attikum, H., et al. (2017). Ataxin-3 consolidates the MDC 1-dependent DNA double-strand break response by counteracting the SUMO -targeted ubiquitin ligase RNF 4 . EMBO J. *36*, 1066–1083.

Pilzecker, B., Buoninfante, O.A., and Jacobs, H. (2019). DNA damage tolerance in stem cells, ageing, mutagenesis, disease and cancer therapy. Nucleic Acids Res.

Pinder, J.B., Attwood, K.M., and Dellaire, G. (2013). Reading, writing, and repair: The role of ubiquitin and the ubiquitin-like proteins in DNA damage signaling and repair. Front. Genet. *4*.

Pommier, Y. (2013). Drugging topoisomerases: Lessons and Challenges. ACS Chem. Biol. *8*, 82–95.

Pommier, Y., Barcelo, J.M., Rao, V.A., Sordet, O., Jobson, A.G., Thibaut, L., Miao, Z.H., Seiler, J.A., Zhang, H., Marchand, C., et al. (2006). Repair of Topoisomerase I-Mediated DNA Damage. Prog. Nucleic Acid Res. Mol. Biol. *81*, 179–229.

Pommier, Y., Leo, E., Zhang, H., and Marchand, C. (2010). DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. Chem. Biol. *17*, 421–433.

Prakash, S., Tian, L., Ratliff, K.S., Lehotzky, R.E., and Matouschek, A. (2004). An unstructured initiation site is required for efficient proteasome-mediated degradation. Nat. Struct. Mol. Biol. *11*, 830–837.

Prasad, R., Horton, J.K., Chastain, P.D., Gassman, N.R., Freudenthal, B.D., Hou, E.W., and Wilson, S.H. (2014). Suicidal cross-linking of PARP-1 to AP site intermediates in cells undergoing base excision repair. Nucleic Acids Res. *42*, 6337–6351.

Prasad, R., Horton, J.K., Dai, D.P., and Wilson, S.H. (2019). Repair pathway for PARP-1 DNA-protein crosslinks. DNA Repair (Amst). *73*, 71–77.

Prioleau, M.N., and MacAlpine, D.M. (2016). DNA replication origins—Where do we

begin? Genes Dev. *30*, 1683–1697.

Psakhye, I., and Jentsch, S. (2012). Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. Cell *151*, 807–820.

Puumalainen, M.R., Lessel, D., Rüthemann, P., Kaczmarek, N., Bachmann, K., Ramadan, K., and Naegeli, H. (2014). Chromatin retention of DNA damage sensors DDB2 and XPC through loss of p97 segregase causes genotoxicity. Nat. Commun. *5*.

Rabinovich, E., Kerem, A., Frohlich, K.-U., Diamant, N., and Bar-Nun, S. (2002). AAA-ATPase p97/Cdc48p, a Cytosolic Chaperone Required for Endoplasmic Reticulum-Associated Protein Degradation. Mol. Cell. Biol. *22*, 626–634.

Raman, M., Havens, C.G., Walter, J.C., and Harper, J.W. (2011). A Genome-wide Screen Identifies p97 as an Essential Regulator of DNA Damage-Dependent CDT1 Destruction. Mol. Cell *44*, 72–84.

Reardon, J.T., and Sancar, A. (2006). Repair of DNA-polypeptide crosslinks by human excision nuclease. Proc. Natl. Acad. Sci. U. S. A. *103*, 4056–4061.

Reardon, J.T., Cheng, Y., and Sancar, A. (2006). Repair of DNA-protein cross-links in mammalian cells. Cell Cycle *5*, 1366–1370.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F.X. (2009). Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing. Cell *139*, 719–730.

Ren, J., Pashkova, N., Winistorfer, S., and Piper, R.C. (2008). DOA1/UFD3 plays a role in sorting ubiquitinated membrane proteins into multivesicular bodies. J. Biol. Chem. *283*, 21599–21611.

Rivera, B., Polak, P., and Foulkes, W.D. (2018). Monogenic Diseases of DNA Repair. N. Engl. J. Med. *378*, 491.

Robson, M., Im, S.A., Senkus, E., Xu, B., Domchek, S.M., Masuda, N., Delaloge, S., Li, W., Tung, N., Armstrong, A., et al. (2017). Olaparib for metastatic breast cancer in patients with a germline BRCA mutation. N. Engl. J. Med. *377*, 523–533.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J. Biol. Chem. *273*, 5858–5868.

Romanowski, M.J., Gibney, S.A., and Burley, S.K. (2002). Crystal structure of the Escherichia coli SbmC protein that protects cells from the DNA replication inhibitor microcin B17. Proteins Struct. Funct. Genet. *47*, 403–407.

Ruijs, M.W.G., van Andel, R.N.J., Oshima, J., Madan, K., Nieuwint, A.W.M., and Aalfs, C.M. (2003). Atypical progeroid syndrome: An unknown helicase gene defect? Am. J. Med. Genet. *116A*, 295–299.

Saldivar, J.C., Cortez, D., and Cimprich, K.A. (2017). The essential kinase ATR: Ensuring faithful duplication of a challenging genome. Nat. Rev. Mol. Cell Biol. *18*, 622–636.

Saldivar, J.C., Hamperl, S., Bocek, M.J., Chung, M., Bass, T.E., Cisneros-Soberanis, F., Samejima, K., Xie, L., Paulson, J.R., Earnshaw, W.C., et al. (2018). An intrinsic S/G2 checkpoint enforced by ATR. Science (80-. ). *361*, 806–810.

Sanchez, Y., Wong, C., Thoma, R.S., Richman, R., Wu, Z., Piwnica-Worms, H., and Elledge, S.J. (1997). Conservation of the Chk1 checkpoint pathway in mammals: Linkage of DNA damage to Cdk regulation through Cdc25. Science (80-. ). *277*, 1497–

1501.

Santamaria, D. (2000). Bi-directional replication and random termination. Nucleic Acids Res. *28*, 2099–2107.

Schellenberg, M.J., Lieberman, J.A., Herrero-Ruiz, A., Butler, L.R., Williams, J.G., Muñoz-Cabello, A.M., Mueller, G.A., London, R.E., Cortés-Ledesma, F., and Williams, R.S. (2017). ZATT (ZNF451)–mediated resolution of topoisomerase 2 DNA-protein cross-links. Science (80-. ). *357*, 1412–1416.

Sczepanski, J.T., Wong, R.S., McKnight, J.N., Bowman, G.D., and Greenberg, M.M. (2010). Rapid DNA-protein cross-linking and strand scission by an abasic site in a nucleosome core particle. Proc. Natl. Acad. Sci. U. S. A. *107*, 22475–22480.

Sengupta, S., and Nagaraja, V. (2008). YacG from Escherichia coli is a specific endogenous inhibitor of DNA gyrase. Nucleic Acids Res. *36*, 4310–4316.

Serbyn, N., Noireterre, A., Bagdiul, I., Plank, M., Michel, A.H., Loewith, R., Kornmann, B., and Stutz, F. (2019). The Aspartic Protease Ddi1 Contributes to DNA-Protein Crosslink Repair in Yeast. BioRxiv 575860.

Seymour, M.T., Maughan, T.S., Ledermann, J.A., Topham, C., James, R., Gwyther, S.J., Smith, D.B., Shepherd, S., Maraveyas, A., Ferry, D.R., et al. (2007). Different strategies of sequential and combination chemotherapy for patients with poor prognosis advanced colorectal cancer (MRC FOCUS): a randomised controlled trial. Lancet *370*, 143–152.

Shibata, A., Conrad, S., Birraux, J., Geuting, V., Barton, O., Ismail, A., Kakarougkas, A., Meek, K., Taucher-Scholz, G., Löbrich, M., et al. (2011). Factors determining DNA double-strand break repair pathway choice in G2 phase. EMBO J. *30*, 1079–1092.

Shibata, A., Moiani, D., Arvai, A.S., Perry, J., Harding, S.M., Genois, M.M., Maity, R., van Rossum-Fikkert, S., Kertokalio, A., Romoli, F., et al. (2014). DNA Double-Strand Break Repair Pathway Choice Is Directed by Distinct MRE11 Nuclease Activities. Mol. Cell *53*, 7–18.

Singh, N.P., McCoy, M.T., Tice, R.R., and Schneider, E.L. (1988). A simple technique for quantitation of low levels of DNA damage in individual cells. Exp. Cell Res. *175*, 184–191.

Sirbu, B.M., Couch, F.B., Feigerle, J.T., Bhaskara, S., Hiebert, S.W., and Cortez, D. (2011). Analysis of protein dynamics at active, stalled, and collapsed replication forks. Genes Dev. *25*, 1320–1327.

Slichenmyer, W.J., Rowinsky, E.K., Donehower, R.C., and Kaufmann, S.H. (1993). The current status of camptothecin analogues as antitumor agents. J. Natl. Cancer Inst. *85*, 271–291.

Snipstad, K., Fenton, C.G., Kjæve, J., Cui, G., Anderssen, E., and Paulssen, R.H. (2010). New specific molecular targets for radio-chemotherapy of rectal cancer. Mol. Oncol. *4*, 52–64.

Somyajit, K., Gupta, R., Sedlackova, H., Neelsen, K.J., Ochs, F., Rask, M.B., Choudhary, C., and Lukas, J. (2017). Redox-sensitive alteration of replisome architecture safeguards genome integrity. Science (80-. ). *358*, 797–802.

Sørensen, C.S., Syljuåsen, R.G., Falck, J., Schroeder, T., Rönnstrand, L., Khanna, K.K., Zhou, B.B., Bartek, J., and Lukas, J. (2003). Chk1 regulates the S phase checkpoint by coupling the physiological turnover and ionizing radiation-induced accelerated proteolysis of Cdc25A. Cancer Cell *3*, 247–258.

Sparks, J.L., Chistol, G., Gao, A.O., Räschle, M., Larsen, N.B., Mann, M., Duxin, J.P., and Walter, J.C. (2019). The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. Cell *176*, 167–181.e21.

Stapf, C., Cartwright, E., Bycroft, M., Hofmann, K., and Buchberger, A. (2011). The general definition of the p97/valosin-containing protein (VCP)-interacting motif (VIM) delineates a new family of p97 cofactors. J. Biol. Chem. *286*, 38670–38678.

Stingele, J., Schwarz, M.S., Bloemeke, N., Wolf, P.G., and Jentsch, S. (2014). A DNA-dependent protease involved in DNA-protein crosslink repair. Cell *158*, 327–338.

Stingele, J., Habermann, B., and Jentsch, S. (2015). DNA-protein crosslink repair: Proteases as DNA repair enzymes. Trends Biochem. Sci. *40*, 67–71.

Stingele, J., Bellelli, R., Alte, F., Hewitt, G., Sarek, G., Maslen, S.L., Tsutakawa, S.E., Borg, A., Kjær, S., Tainer, J.A., et al. (2016). Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. Mol. Cell *64*, 688–703.

Strumberg, D., Pilon, A.A., Smith, M., Hickey, R., Malkas, L., and Pommier, Y. (2000). Conversion of Topoisomerase I Cleavage Complexes on the Leading Strand of Ribosomal DNA into 5'-Phosphorylated DNA Double-Strand Breaks by Replication Runoff. Mol. Cell. Biol. *20*, 3977–3987.

Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell *123*, 1213–1226.

Su, D., and Hochstrasser, M. (2010). A WLM Protein with SUMO-Directed Protease Activity. Mol. Cell. Biol. *30*, 3734–3736.

Sugasawa, K., Ng, J.M.Y., Masutani, C., Iwai, S., Van Der Spek, P.J., Eker, A.P.M., Hanaoka, F., Bootsma, D., and Hoeijmakers, J.H.J. (1998). Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. Mol. Cell *2*, 223–232.

Sun, Y., and Nitiss, J. (2017). A SUMO-ubiquitin mediated proteasome pathway in repair of DNA damage induced by topoisomerase II inhibitors. FASEB J. *31*, lb57-lb57.

Svoboda, M., Konvalinka, J., Trempe, J.F., and Grantz Saskova, K. (2019). The yeast proteases Ddi1 and Wss1 are both involved in the DNA replication stress response. DNA Repair (Amst). *80*, 45–51.

Takahara, P.M., Rosenzweig, A.C., Frederick, C.A., and Lippard, S.J. (1995). Crystal structure of double-stranded DNA containing the major adduct of the anticancer drug cisplatin. Nature *377*, 649–652.

Takashima, H., Boerkoel, C.F., John, J., Saifi, G.M., Salih, M.A.M., Armstrong, D., Mao, Y., Quiocho, F.A., Roa, B.B., Nakagawa, M., et al. (2002). Mutation of TDP1, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy. Nat. Genet. *32*, 267–272.

Takimoto, C.H., and Arbuck, S.G. (1997). Clinical status and optimal use of topotecan. Oncology (Williston Park).

Thangavel, S., Berti, M., Levikova, M., Pinto, C., Gomathinayagam, S., Vujanovic, M., Zellweger, R., Moore, H., Lee, E.H., Hendrickson, E.A., et al. (2015). DNA2 drives processing and restart of reversed replication forks in human cells. J. Cell Biol. *208*, 545–562.

Thomas, A., Murai, J., and Pommier, Y. (2018). The evolving landscape of predictive biomarkers of response to PARP inhibitors. J. Clin. Invest. *128*, 1727–1730.

Tong, Z., Han, C., Luo, W., Wang, X., Li, H., Luo, H., Zhou, J., Qi, J., and He, R. (2013). Accumulated hippocampal formaldehyde induces age-dependent memory decline. Age (Omaha). *35*, 583–596.

Torrecilla, I., Oehler, J., and Ramadan, K. (2017). The role of ubiquitin-dependent segregase p97 (VCP or Cdc48) in chromatin dynamics after DNA double strand breaks. Philos. Trans. R. Soc. B Biol. Sci. *372*.

Tretyakova, N.Y., Groehler, A., and Ji, S. (2015). DNA-Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes. Acc. Chem. Res. *48*, 1631–1644.

Tubbs, A., and Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. Cell *168*, 644–656.

Tubbs, A., Sridharan, S., van Wietmarschen, N., Maman, Y., Callen, E., Stanlie, A., Wu, W., Wu, X., Day, A., Wong, N., et al. (2018). Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse. Cell *174*, 1127–1142.e19.

Vanoli, F., Fumasoni, M., Szakal, B., Maloisel, L., and Branzei, D. (2010). Replication and recombination factors contributing to recombination-dependent bypass of DNA lesions by template switch. PLoS Genet. *6*.

Vaz, B., Halder, S., and Ramadan, K. (2013). Role of p97/VCP (Cdc48) in genome stability. Front. Genet. *4*.

Vaz, B., Popovic, M., Newman, J.A., Fielden, J., Aitkenhead, H., Halder, S., Singh, A.N., Vendrell, I., Fischer, R., Torrecilla, I., et al. (2016). Metalloprotease SPRTN/DVC1 Orchestrates Replication-Coupled DNA-Protein Crosslink Repair. Mol. Cell *64*, 704–719.

Vaz, B., Popovic, M., and Ramadan, K. (2017). DNA–Protein Crosslink Proteolysis Repair. Trends Biochem. Sci. *42*, 483–495.

Wall, M.E. (1998). Camptothecin and taxol: Discovery to clinic. Med. Res. Rev. *18*, 299–314.

Wang, J.L., Duboc, C., Wu, Q., Ochi, T., Liang, S., Tsutakawa, S.E., Lees-Miller, S.P., Nadal, M., Tainer, J.A., Blundell, T.L., et al. (2018). Dissection of DNA double-strand-break repair using novel single-molecule forceps. Nat. Struct. Mol. Biol. *25*, 482–487.

Wang, M., Weiss, M., Simonovic, M., Haertinger, G., Schrimpf, S.P., Hengartner, M.O., and Von Mering, C. (2012). PaxDb, a database of protein abundance averages across all three domains of life. Mol. Cell. Proteomics *11*, 492–500.

Wei, Y., Diao, L.X., Lu, S., Wang, H.T., Suo, F., Dong, M.Q., and Du, L.L. (2017). SUMO-Targeted DNA Translocase Rrp2 Protects the Genome from Top2-Induced DNA Damage. Mol. Cell *66*, 581–596.e6.

Wen Li, Xiuzhen Bai, Jun Li, Yichao Zhao, Jingyan Liu, Huayu Zhao, Lan Liu, Miao Ding, Qingsong Wang, Fang-Yuan Shi, Mei Hou, Jianguo Ji, Ge Gao, Rong Guo, Yujie Sun, Y.L.& D.X. (2019). The nucleoskeleton protein IFFO1 immobilizes broken DNA and suppresses chromosome translocation during tumorigenesis. Nat. Cell Biol.

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., et al. (2019). TRAIP is a master regulator of DNA interstrand crosslink repair. Nature *567*, 267–272.

Xu, G., Ross Chapman, J., Brandsma, I., Yuan, J., Mistrik, M., Bouwman, P., Bartkova,

J., Gogola, E., Warmerdam, D., Barazas, M., et al. (2015). REV7 counteracts DNA double-strand break resection and affects PARP inhibition. Nature *521*, 541–544.

Xue, L., Blythe, E.E., Freiberger, E.C., Mamrosh, J.L., Hebert, A.S., Reitsma, J.M., Hess, S., Coon, J.J., and Deshaies, R.J. (2016). Valosin-containing protein (VCP)-adaptor interactions are exceptionally dynamic and subject to differential modulation by a VCP inhibitor. Mol. Cell. Proteomics *15*, 2970–2986.

Yamanaka, K., Chatterjee, N., Hemann, M.T., and Walker, G.C. (2017). Inhibition of mutagenic translesion synthesis: A possible strategy for improving chemotherapy? PLoS Genet. *13*.

Yang, W., and Gao, Y. (2018). Translesion and Repair DNA Polymerases: Diverse Structure and Mechanism. Annu. Rev. Biochem. *87*, 239–261.

Yang, S.W., Burgin, A.B., Huizenga, B.N., Robertson, C.A., Yao, K.C., and Nash, H.A. (1996). A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. Proc. Natl. Acad. Sci. *93*, 11534–11539.

Ye, Y., Meyer, H.H., and Rapoport, T.A. (2003). Function of the p97-Ufd1-Npl4 complex in retrotranslocation from the ER to the cytosol: Dual recognition of nonubiquitinated polypeptide segments and polyubiquitin chains. J. Cell Biol. *162*, 71–84.

Ye, Y., Tang, W.K., Zhang, T., and Xia, D. (2017). A mighty “protein extractor” of the cell: Structure and function of the p97/CDC48 ATPase. Front. Mol. Biosci. *4*.

Yoon, J.H., McArthur, M.J., Park, J., Basu, D., Wakamiya, M., Prakash, L., and Prakash, S. (2019). Error-Prone Replication through UV Lesions by DNA Polymerase θ Protects against Skin Cancers. Cell *176*, 1295–1309.e15.

Yu, J., Shannon, W.D., Watson, M.A., and McLeod, H.L. (2005). Gene expression profiling of the irinotecan pathway in colorectal cancer. Clin. Cancer Res. *11*, 2053–2062.

Zagnoli-Vieira, G., Bruni, F., Thompson, K., He, L., Walker, S., Brouwer, A.P.M.D., Taylor, R., Niyazov, D., and Caldecott, K.W. (2018). Confirming TDP2 mutation in spinocerebellar ataxia autosomal recessive 23 (SCAR23). Neurol. Genet. *4*.

Zegerman, P., and Diffley, J.F.X. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature *445*, 281–285.

Zellweger, R., Dalcher, D., Mutreja, K., Berti, M., Schmid, J.A., Herrador, R., Vindigni, A., and Lopes, M. (2015). Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. J. Cell Biol. *208*, 563–579.

Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. Nat. Cell Biol. *16*, 2–9.

Zhang, J., and Walter, J.C. (2014). Mechanism and regulation of incisions during DNA interstrand cross-link repair. DNA Repair (Amst). *19*, 135–142.

Zhang, F., Fan, Q., Ren, K., and Andreassen, P.R. (2009). PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. Mol. Cancer Res. *7*, 1110–1118.

Zhang, H.F., Tomida, A., Koshimizu, R., Ogiso, Y., Lei, S., and Tsuruo, T. (2004). Cullin 3 Promotes Proteasomal Degradation of the Topoisomerase I-DNA Covalent Complex. Cancer Res. *64*, 1114–1121.

Zhang, J., Dewar, J.M., Budzowska, M., Motnenko, A., Cohn, M.A., and Walter, J.C. (2015). DNA interstrand cross-link repair requires replication-fork convergence. Nat.

Struct. Mol. Biol. *22*, 242–247.

Zhang, Y.W., Regairaz, M., Seiler, J.A., Agama, K.K., Doroshow, J.H., and Pommier, Y. (2011). Poly(ADP-ribose) polymerase and XPF-ERCC1 participate in distinct pathways for the repair of topoisomerase I-induced DNA damage in mammalian cells. Nucleic Acids Res. *39*, 3607–3620.

Zhao, W., Vaithiyalingam, S., San Filippo, J., Maranon, D.G., Jimenez-Sainz, J., Fontenay, G. V., Kwon, Y., Leung, S.G., Lu, L., Jensen, R.B., et al. (2015). Promotion of BRCA2-Dependent Homologous Recombination by DSS1 via RPA Targeting and DNA Mimicry. Mol. Cell *59*, 176–187.

Zhao, W., Steinfeld, J.B., Liang, F., Chen, X., Maranon, D.G., Jian Ma, C., Kwon, Y., Rao, T., Wang, W., Sheng, C., et al. (2017). BRCA1-BARD1 promotes RAD51-mediated homologous DNA pairing. Nature *550*.

Zhu, Z., Chung, W.H., Shim, E.Y., Lee, S.E., and Ira, G. (2008). Sgs1 Helicase and Two Nucleases Dna2 and Exo1 Resect DNA Double-Strand Break Ends. Cell *134*, 981–994.

Zou, L., and Elledge, S.J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science (80-. ). *300*, 1542–1548.

Zou, Y., Liu, Y., Wu, X., and Shell, S.M. (2006). Functions of human replication protein A (RPA): From DNA replication to DNA damage and stress responses. J. Cell. Physiol. *208*, 267–273.