Structure-Function Analysis of SLX4 and Its Scaffolded Nucleases

Malitha Ratnaweera

University College & Department of Oncology, University of Oxford

Supervisor: Professor Peter McHugh

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2024

Acknowledgements

Foremost, I extend my deepest gratitude to my supervisor, Peter, for his unwavering support throughout this project. Your guidance during periods of uncertainty, especially during the pandemic and beyond, your vast and ever-evolving expertise in the field, and your encouragement to explore various avenues have been invaluable. It has been a profound privilege to work under your mentorship.

I would like to express my sincere appreciation to those who assisted in the core facilities at the WIMM. I am thankful to the members of the WIMM flow cytometry (Craig, Kevin, and Sally), mass spectrometry (Marjorie), and oncology facilities (Linda, Caroline, Eleanor, Joe, Andy, and Saji) teams for their critical contributions to my work. I am also indebted to indispensable discussions and resources offered by Dr. Philip Hublitz and Prof. Yin Dong.

The members of Peter's lab, both past and present, have provided an immensely supportive environment. Much of my work is thanks to Denis, and I appreciate his work initiating the SMX project, and to Abi, Jenni, Sanja and Will for their continued dedication to its advancement. The young(er) researchers, including Michael, Niamh, Ellie, and Iulia, have played a pivotal role in nurturing my passion for lab work and teaching. I hope I have been able to repay, in some measure, the energy and commitment you contributed to our shared work. I have cherished many happy moments with members of adjoining labs over the years - Matthew, Ann-Marie, Grid, Spencer, Tomas, Nat and Gulnar - thank you for shortening many long-feeling days. I would especially like to acknowledge Julius, whose unshakable confidence and passion for science were both a source of endless amusement and inspiration.

I am also particularly grateful for the exceptional technical expertise provided by Will, both in Oxford and in the welcoming environment of the Nussenzweig lab in Bethesda. My sincere thanks go to Andre for guiding me through the exciting work in his lab. Two remarkable scientist-mentors have had a profound influence on almost all the work in this thesis: Lonnie, thank you for your unrelenting attention to all details lab-related and for answering the never-ending bombardment of questions I brought to you every week/day/hour; and Kaima, whose near-daily discussions consistently placed the following week's work in better stead than I could ever have planned. I am also deeply appreciative of the enduring support I received from Hannah and Rebecca, two extraordinarily compassionate and talented role models who timed their tenures in the lab so considerately that I was never without their guiding light and kindness.

I have been blessed with two friends from school, Sarah and Grace, who have shaped the person I am in more ways than I can express. Having you both nearby as I navigated this chapter has brought immense comfort in hard times. Your dependability and encouragement over a decade of friendship have been unfailing, and I am extremely grateful. Choosing to study biochemistry felt like a natural path from the very start of my undergraduate studies, thanks in large part to the wonderful solidarity and enthusiasm that David brought to our shared experience. Thank you for being a constant source of both academic and personal encouragement. I am also very grateful to Anna, Kieran, Elizabeth, and Zoulikha for their steady friendship over the years and to the entire Oxenford family for their welcoming and excellent company both on and off the pitch.

Graduate student life, especially during a year as uncertain as 2020, could only have been made bearable through meeting two especially kind and loving friends, Ed and Rachel. I couldn't have asked for two more caring, wholehearted people to help guide me through such a daunting transition. Sharing those early, often uncertain, moments with you both brought not only a sense of belonging but real joy, and I'll always be grateful for the laughter and devoted friendship you offered so freely. I feel completely in debt to Orion for bringing a welcome sense of balance and humour to the often intense demands of research life. Thank you also for sharing your impeccable taste in music and TV, along with generous doses of both academic and worldly wisdom. To Lucy, your patience and endless care have been my rock throughout this journey. Thank you for standing by me through many highs and even more lows and for always knowing how to lift me when I doubted myself.

I'm grateful to all my seeyas and acchis, from Mellor to Borella to Sydney. To my cousins Megan and Jacob, I was delighted to reconnect after so many years, and I deeply cherish the time and energy you dedicated to including me in your home and lives. I am also grateful to Uncle Mike and Auntie Manuja for taking me in and making my brief time in DC so memorable, filled with love, Gia's company, and pistachio ice cream.

To my family, words cannot express the depth of my gratitude. To my brother, thank you for being by my side your whole life and for always reminding me of my worth. To my parents, thank you for instilling in me a fraction of the values that make you such wonderful people, and for always believing in me. Your constant encouragement has been the foundation of anything I have achieved.

Abstract

The ever-expanding utility of cryoelectron microscopy (cryo-EM) has enabled resolving architectures of protein complexes that were once considered beyond reach[1–3]. However, DNA damage-associated protein complexes pose unique challenges due to their pronounced flexibility and intrinsic disorder, which render them poorly defined even by advanced *in silico* machine-learning-based structural predictions[4]. Such complexes denote high-priority but structurally elusive targets, necessitating careful purification and experimental structure determination.

One such complex is the SLX4-scaffolded nuclease complex (SMX) comprising the DNA nucleases XPF-ERCC1, MUS81-EME1 and SLX1. This complex is profoundly implicated in cancer[5], ageing[6], and Fanconi anaemia (FA)[7] - diseases intricately linked to DNA damage and repair mechanisms. SLX4 serves as a platform for recruiting these nucleases and coordinates the resolution of diverse DNA lesions, particularly interstrand crosslinks (ICLs), which are among the most cytotoxic forms of DNA damage[8]. The structural characterisation of the SMX complex remains a critical and ambitious objective of this project. Furthermore, endogenous causative agents generating ICLs in cells remain an area of active investigation.

This work contributes to the growing evidence implicating formaldehyde as a key endogenous source of ICLs, supporting its intrinsic connection to aldehyde toxicity and the Fanconi anaemia pathway[9]. This study further proposes that the accumulation of formaldehyde during epigenetic reprogramming may represent a significant source of ICLs, potentially contributing to the differentiation defects observed in the blood lineages of patients with FA[10]. In parallel, CRISPR-Cas9-targeted disruption of SLX1 and XPF was undertaken to determine biological roles and genetic interactions underlying their loss. Moreover, significant efforts were made toward the homogeneous purification of SMX, employing multi-pronged expression systems to address obstacles in achieving stability and functionality. Given challenges associated with purifying the entire complex, functional studies on sub-complexes and investigations into novel interactions of SLX4 provide valuable insights into the post-translational modifications regulating SLX4's association with MUS81-EME1 and other DNA repair factors. The ultimate goal of these efforts is to obtain high-quality SMX suitable for structural characterisation by cryo-EM, advancing understanding of its molecular architecture and biological roles.

Contents

List of Figures

List of Tables

1 | Introduction

Contents

1.1 | Genome stability

"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."- while appended casually like a footnote to Watson and Crick's proposed structure of DNA[11], the notion that DNA

duplication carried the genetic information between generations also necessitated its integrity. Any DNA modifications encountered in a cell's lifespan, rendering the code unreadable, present a risk to its progeny if inherited. Therefore, protecting DNA's fidelity in living cells is an essential process continually antagonised through exposure to damaging sources, whether exogenous or endogenous. Robust, redundant and efficient DNA repair pathways have evolved in all domains of life in this relentless battle against genomic assaults. These pathways are collectively known as the DNA damage response (DDR) and include hundreds of proteins that sense, signal, and repair specific DNA lesions.

1.2 | Damage to DNA

A simplistic view of DNA repair might suggest that a predetermined pathway is perfectly adapted to correct any single DNA lesion. Often and for many insults, this can be true, yet emerging evidence highlights the significant crosstalk, redundancy, and contextuality that repairing DNA actually requires. Despite this, the frequency and estimated toxicity of different lesions vary (Table 1.1), and these numbers can be reflected in the complexity of repair pathways dedicated to repairing each lesion (Figure 1.1). It appears that the greater the ramifications of a particular form of DNA damage to a cell, the greater the need to dedicate more genes to its repair. For example, relatively common depurination events (estimated to occur 12,000 times per cell per day[12]) are repaired in a manner that requires three main genes within the base excision repair (BER) pathway (though redundancy exists with other genes). Meanwhile, the repair of up to only 10 crosslinks per cell per day[13] accounts for a pathway that requires at least 23 genes of the Fanconi anaemia pathway[14]. Perhaps what is better reflected in pathway complexity is the difficulty encountered in resolving the chemistry that defines a DNA injury. An ICL requires more chemical steps to return the original underlying DNA sequence, while a depurinated base (at a minimum) requires only base excision, single-nucleotide replacement and religation (Section 1.3.1).

Damage to one of the two DNA strands endangers cells through the possible introduction of point mutations or small insertions/deletions (indels), which can alter critical protein sequence or expression. It can be argued that the damage to DNA isn't the fatal issue – it's how that damage is encountered by DNA-engaging machinery. Nevertheless, given that DNA is mostly non-coding, why is a cell unable to ignore the lesion and allow minor

modifications to pass? Based on a frequency of 70,000 lesions per day[12], transcription alone would encounter 350-700 lesions per cell in some way. The remaining 99% of lesions would obstruct replication and signal towards DNA repair machinery. However, tolerance is still an intrinsic safeguard against DNA damage (Figure 1.1). Equally, these lesions can impair replicative and transcriptional polymerases, hampering regular cell functions and preventing these cells from propagating.

All cells encounter a basal level of endogenous DNA damage due to normal cellular processes that impart damage as a by-product of their typical function (Figure 1.1). However, cells are usually adept at overcoming these assaults when in steady state. DNA is, nevertheless, also hit with environment-dependent exogenous sources of damage. This exogenous damage can be further divided into typical physiological damage, *i.e.* a source that stresses the average cell of a healthy individual, such as UV damage, and unnatural sources of damage, *i.e.* what afflicts unique cells in a particular context, such as chemotherapeutic agents.

Damage	**Examples**	**Estimated frequency (per cell per day)**
Damaged base	8-oxo-G	2,800 (ref[12])
Single-strand break		55,000 (ref[12])
Mismatch	DNA replication errors	30,000-30 million (assuming $10^{-4} - 10^{-7}$ proof-reading polymerase error rates (ref[15]))
Intrastrand crosslinks	UV-induced photodimers	unknown, in the order of 100,000 from sunlight (ref[16])
Double-strand break		25 (ref[12])
Interstrand crosslink		unknown, perhaps up to 10 (ref[13])
DNA-protein crosslinks		15,000-135,000 (ref[17])

Table 1.1 | Estimate DNA damage frequencies.

1.2.1 | Exogenous lesions

DNA lesions can be generated from electromagnetic radiation sources, which are capable of energising particular atoms within DNA or activating nearby molecules that can then react with DNA. Ingested or inhaled sources of DNA-damaging agents also form a notable portion of the damage repertoire encountered by an individual. Such sources vary, arising

from components of seemingly innocuous ingested food chemicals or by, for example, the administration of drugs either clinically or recreationally. Proximity to various microbial agents can also endanger cells through genotoxicity.

The average cell is hit with a continual bombardment of high-energy particles capable of damaging DNA. Low-level exposure to ultraviolet radiation (UV), in the form of UV-B (290-320 nm)[18], reaches individuals despite the ozone layer limiting their transmission to the Earth's surface[19]. Serendipitously for molecular biologists quantifying nucleic acid concentrations, DNA (and RNA) absorb maximally at 260 nm[20], and through absorption can result in the promotion of states favouring formation of *e.g.* cyclobutane pyrimidine dimers and 6-4 pyrimidine photoproducts, both well-described intrastrand crosslinks[21,22]. With somewhat lower incidence, some interstrand crosslinks can also form from UV radiation[23]. Ionising radiation (IR) also delivers significant energy to DNA, but crucially imparts enough energy to detach electrons from DNA, and therefore ionise[24]. Most humans sustain non-trivial annual IR exposure through cosmic rays[25,26] and a steady stream of α-particles radiating from decaying soil radon[27,28]. Radiated photons (x-rays, gamma rays) exert their ionising effects indirectly through the release of β-particles, capable of ionising DNA[24]. Ionisation, occurring directly through the absorption of such β-particles or *via* oxidative radicals through indirect radiolysis, can induce multiple DNA lesions, including double-strand breaks (DSBs), SSBs, and base lesions such as 8-oxo-G or depurination/depyrimidation events[29]. Clustered damage, defined as more than one lesion per helical turn, is challenging to revert and is often left unrepaired and accounts for the unique effects of acute radiation exposure[30,31]. In a strictly dose-dependent manner, exposure to either radiation form increases cancer risk[32].

Carcinogenic chemical elements appear ever increasingly in our daily lives, and as such, an appreciable number of exogenous toxins in the environment are ingested or inhaled, harming DNA. Ingested toxins, such as mycotoxins, further add to the DNA damage burden. Aflatoxin B1, a potent mycotoxin produced by *Aspergillus* species growing on poorly stored crops like peanuts and maize, induces guanine adducts in DNA[33]. These lesions lead to characteristic G:C to T:A transversions, which are hallmarks of hepatocellular carcinoma[34]. Tobacco smoke, a significant source of polycyclic aromatic hydrocarbons (PAHs), delivers highly mutagenic compounds that form DNA adducts,

disrupt repair processes, and increase mutation rates[35]. Many chemotherapeutic agents were initially developed as wartime poisons and later repurposed at far lower doses to exploit the vulnerabilities of rapidly dividing cancer cells[36]. Alkylating agents, like cyclophosphamide, and platinum-based drugs, such as cisplatin, create DNA crosslinks that cancer cells often cannot repair due to dysfunctional DDR[37]. Similarly, topoisomerase inhibitors like doxorubicin induce strand breaks that can overwhelm the repair machinery of cancer cells[38]. While effective, these agents usually have significant side effects, including mutagenesis and secondary malignancies, necessitating ongoing research to refine their administration[39]. The contribution of physical extremes, such as hypoxia (low oxygen levels) and low temperatures, is less understood but increasingly recognised to contribute to DNA damage. Hypoxia, for example, can impair DNA repair by limiting ATP-dependent processes and increasing oxidative stress[40]. Cold environments may indirectly contribute to DNA damage by altering cellular repair capacity and increasing the production of reactive oxygen species during reoxygenation events following cold-induced ischaemia[41].

1.2.2 | Endogenous DNA damage

One of the notable threats to the genome is the presence of reactive oxygen species (ROS); toxic byproducts of a cell's indispensable need for catabolism. Superoxide anions, hydrogen peroxide and hydroxyl radicals ($O_2^{-\bullet}$, H_2O_2, $^\bullet OH$) can escape from the electron transport chain during oxidative phosphorylation[42] but are counteracted by cellular reductants like the thiol-containing glutathione[43]. Indiscriminate oxidation is the great danger of ROS. For example, hydroxylation (by $^\bullet OH$) of the C^8 position within the purine ring of guanine forms 8-oxoguanine (8-oxo-G)[44], which base pairs erroneously with adenine using a Hoogsteen edge when in the 'syn' conformation (otherwise 'anti' when base-pairing with cytosine). This base-pairing results in a G>T transversion substitution in the following division[45].

An understated source of endogenous DNA lesions, albeit a promising avenue for future research, is the microbiome[46]. Emerging evidence suggests that microbiota-derived genotoxins contribute to unique mutational signatures, predominantly in colorectal cancers[47,48]. Consisting of two electrophilic cyclopropane warheads, colibactin, widely produced by the Enterobacteriaceae family (which includes commensal *Escherichia coli*),

can single-strand alkylate or even crosslink DNA strands, akin to the nucleophilic attack of N^3 amines on adenine/guanine by carcinogenic nitrogen mustards (HN1-3)[49–51]. Assays of DNA damage on plasmids allowed researchers to identify damage patterns resulting from released compounds of 18 phylogenetically diverse isolates from inflammatory bowel disease patients[52]. Here, DNase-like damage fragments and alkylations were abundant, while some ICLs were also observed[52]. Interestingly, a strategy used by certain microorganisms to eliminate competitors in their microenvironment is the production of genotoxic agents, such as cordycepin, which effectively function as antibiotics[53]. This represents an interesting evolutionary strategy where DNA is deliberately targeted, highlighting the importance of genome stability and drawing parallels with the use of these agents as anti-cancer drugs.

Imperfect replication represents another source of endogenous DNA damage. Standard replicative polymerases misincorporate bases every 10^5 (0.1 Mbp)[15]. Difficult-to-replicate regions, such as repetitive segments at centromeres and telomeres, or common fragile sites (CFSs) can slow replication forks. CFSs are especially susceptible to breakage in mitosis, a phenomenon which is exploited to study the response to replication stress through the administration of hydroxyurea or aphidicolin[54,55].

Epigenetic remodelling, particularly during processes such as cellular differentiation or transcriptional reactivation, is increasingly recognised as another potential source of DNA damage. A key contributor to this is the demethylation of proteins such as histones, catalysed by enzymes like LSD1 (*KDM1A*) and members of the Jumonji C (JmjC) domain-containing demethylase family[56,57]. These enzymes remove methyl groups from lysine residues on histones *via* oxidative reactions, producing formaldehyde as a by-product[56]. In normal physiological settings, this formaldehyde is rapidly detoxified by aldehyde-processing enzymes such as ADH5[58]. However, during periods of extensive epigenetic reprogramming, such as stem cell differentiation, the cumulative release of formaldehyde might exceed detoxification capacity[59], potentially resulting in the formation of genotoxic lesions like DNA-protein crosslinks (DPCs) or ICLs.

1.2.3 | Replication stress

A source of DNA fragility comes from imperfect cellular responses to replication stress. In its simplest form, replication stress is defined by an uncoupling of the replicative DNA helicase and polymerase[60,61] and multiple scenarios are capable of forcing this uncoupling. DNA-trapped proteins (Section 1.3.6.1), such as PARP1[62] or TOP1[63,64], or DNA secondary structures (*e.g.* G-quadruplexes, formed by the Hoogsteen-type base pairing of stretches of guanines[65]) can present obstacles to replication. Similarly, impeding transcriptional machinery can cause a replication fork to stall[66,67]. Hydroxyurea (HU)[68,69] and aphidicolin (APH, a natural metabolite of the fungus *Cephalosporium aphidicola*[70]) are routinely used to elicit artificial replication stress. HU inhibits ribonucleotide reductase, decreasing dNTP levels to a replication-limiting level[71], while APH directly blocks the activity of B-family DNA polymerases[72]. This replication stress is marked by replication fork-stalling events that fall short of the activity of DNA helicases, manifesting as extended ssDNA tracts that quickly become coated by replication protein A (RPA)[73]. An ATR-CHK1 axis (Section 1.3.4.4) coordinates a replication stress response (RSR) for the stabilisation and restart of a replication fork[74–76]. Where RSR fails, forks collapse and an overburden of ssDNA tracts exhaust the cell's RPA pool, creating deprotected stretches of ssDNA that are vulnerable to nucleolytic digestion[77,78]. This culminates in replication catastrophe, but it can be mitigated through a delicate balancing act between fork remodelling and degradation.

Replication fork stabilisation and restart The accurate restart or stabilisation of stalled replication forks is crucial to maintain genome integrity. The key step in restarting replication forks is their reversal, which forms a transient four-way junction structure[79,80]. This regressed fork allows parental strands to reanneal, unveiling the newly synthesised strand for hybridisation and forming an overall four-way junction reminiscent of a Holliday junction, referred to as a 'chicken-foot' structure[81]. The annealed daughter strands of a reversed fork mimic a one-ended DSB, and their protection is paramount to prevent erroneous nuclease activity[82–85]. The delay offered by a stabilised fork structure allows time for the original replication-blocking lesion to be removed before either restart or a template-switching-based bypass mechanism[86]. Some persistent stalled forks can be

processed into DSBs by MUS81-EME1 to facilitate homologous recombination (HR)-mediated fork-restart[87,88] (Section 1.4.5.1). Combined single-molecule DNA fibre and electron microscopy approaches[82,85,89] have elucidated how the translocase activities of SMARCAL1 and ZRANB3 can enact fork remodelling, the former sensing RPA-coated ssDNA tracts[90] and the latter relying on polyubiquitylation of PCNA by HLTF[89]. RAD54[91,92] and FANCM[93,94] can catalyse fork reversal *in vitro* and may be redundant with other SWI/SNF family translocases. Many HR and DDR factors contribute secondarily towards fork protection. RAD51 promotes reversal and coats ssDNA strands to prevent nucleolytic attack of reverted arms[95–98]. In particular, the MRE11 nuclease is implicated in generating nicks that create substrates for further cleavage by MUS81, although this activity is censured through BRCA2-stabilised, RAD51 nucleofilament formation[99]. Several separation-of-function mutants distinguish HR-competent, but fork protection-deficient variants of BRCA2 (such as S3291A[96,100,101]) and even BRCA1[102], highlighting how these functions have been modularised through evolution. Several of the FA pathway genes beyond BRCA1 (FANCS), BRCA2 (FANCD1), and RAD51 (FANCR) also moonlight in the protection of stressed replication forks. Unlike cells deficient for other members of the FA pathway, the depletion of FANCD2 is required for resistance to HU. However, this requirement is independent of its monoubiquitylation by the core complex, despite the observation that it is monoubiquitylated upon HU treatment[103]. S-phase-specific interactions have also been described between FANCD2 and BRCA1/RAD51[104], and FANCD2 helps conserve fork stability during BRCA1/2-deficiency[105,106]. FANCD2 interacts with MCM proteins and restrains DNA synthesis in HU/APH-based replication stress[107]. FANCD2, within a complex including FANCI, forms a strong nucleoprotein lock around dsDNA[1,108,109], and this might be recapitulated in cells binding to reversed replication forks, preventing their MRE11- or DNA2-dependent degradation.

Transcription-replication conflicts (TRCs) An unfortunate consequence of cellular complexity for dividing cells is that transcription and replication often come into direct conflict. Origins of replication, by design, frequently overlap with promoters initiating transcription, as regions with higher AT content reduce the energy burden required for double-stranded DNA melting[110]. This dual functionality introduces the potential for collisions between the two processes, either head-on, where the replication machinery

meets the transcription complex from the opposite direction, or co-directionally, where replication machinery catches up to an actively transcribing RNA polymerase II (RNA pol II/RNAPII)[111]. While it has been hypothesised that transcription might aid replication by pre-melting DNA, evidence suggests this relationship is more disruptive than synergistic, particularly under conditions of genomic stress[112].

A major disruption to replication is the formation of R-loops, which occur when the nascent RNA hybridises with the template DNA strand, displacing the non-transcribed strand as ssDNA[113]. R-loops are particularly hazardous because the exposed ssDNA is highly susceptible to mutagenic insults, including oxidative damage and spontaneous base deamination[114,115]. The displaced strand can also fold into non-canonical secondary structures such as G-quadruplexes (G4s), which can stall replication and transcription machinery, exacerbating genomic instability and requiring RTEL1 for their unwinding[116]. The Fanconi anaemia-related FANCJ helicase is also capable of unwinding G4 structures[117–119], which is required for replication fork progression, while other FA proteins have been implicated in preventing R-loops and genomic instability caused by TRC[120,121]. R-loops have been implicated in promoting replication fork collapse and triggering DNA damage signalling pathways, including those regulated by the Fanconi anaemia (FA) pathway[122]. Another significant challenge arising from concurrent transcription and replication is torsional stress. As replication and transcription progress, they generate supercoiling in the DNA ahead of the fork or transcription complex[123]. Positive supercoiling can impede both processes, while negative supercoiling behind the replication fork can promote R-loop formation[124,125]. Topoisomerases alleviate torsional stress by introducing transient breaks in DNA, but their failure or dysregulation can lead to fork stalling, DSBs, and chromosomal rearrangements[126]. The exposure of the non-transcribed strand during R-loop formation presents further challenges. Extruded ssDNA is particularly vulnerable to nucleases and other genotoxic agents, leading to increased mutation rates or fork collapse[127]. In resolving these conflicts effectively, cells maintain genomic integrity while supporting the high demands of replication and transcription, particularly in proliferative tissues and during development.

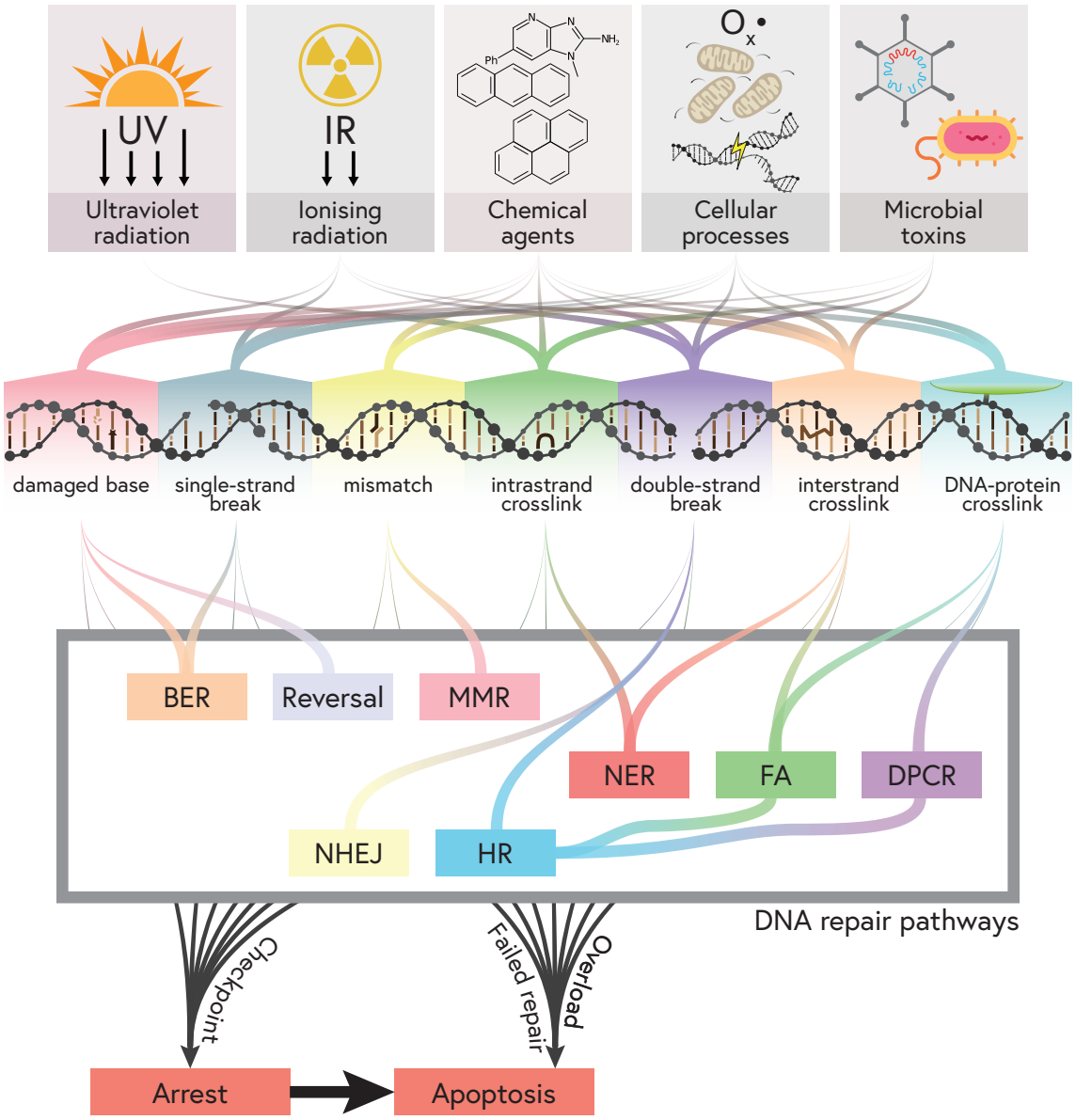


Figure 1.1 | Overview of DNA damage and repair pathways. DNA damage arises from various endogenous and exogenous sources, including ultraviolet (UV) radiation, ionising radiation (IR), chemical agents, cellular processes, and microbial toxins. These insults lead to diverse types of DNA lesions, such as damaged bases, single-strand breaks, mismatches, intrastrand crosslinks, double-strand breaks, ICLs, and DPCs. To maintain genomic integrity, cells employ specialised DNA repair pathways: base excision repair (BER) mainly for repair of damaged bases, mismatch repair (MMR), nucleotide excision repair (NER) for bulky adducts and intrastrand crosslinks, homologous recombination (HR) and non-homologous end joining (NHEJ) for double-strand breaks, the Fanconi anaemia (FA) pathway for ICLs, and DPC repair (DPCR). Failure to repair damage effectively can result in checkpoint activation, leading to cell cycle arrest or apoptosis, thereby preventing the propagation of damaged DNA.

1.3 | The DDR

Though individual repair pathways cannot be mapped 1:1 to a given DNA lesion, we can delineate pathways by the most common type of lesion they would usually repair. How lesions arise and the 'classical' pathway that enacts their resolution is summarised in Figure 1.1. Though not a strict mandate, the major events of most repair pathways can be thought of as a sequential checklist. The first stage often involves some 'sensing' mechanism, stressing the problem-at-hand to signal amplifiers, exemplified by the ATM and ATR kinases[128] (Section 1.3.4.4). Sensing, in turn, recruits effectors of the various pathways, where a certain number of 'pre-processing' steps are usually required; an example within NHEJ involves the action of nucleases Artemis and polymerases pol λ, pol μ, or TDT (*DNTT*) in processing non-ligatable DNA ends at a DSB such that they are compatible with the action of DNA ligases (Section 1.3.4.1)[129]. Pre-processing might also represent more considerable structural changes like the homology search and strand invasion performed by RAD51 in HR (Section 1.3.4) or the remodelling of ICL-converged replication forks to a chicken-foot structure by FANCM (Section 1.3.5). Many pathways, however, converge on some later 'synthesis' step; inevitably, several nucleotides are lost from the duplex during upstream repair, and polymerases are necessary to fill these gaps. A final 'ligation' step covalently links the backbone.

1.3.1 | Base excision repair

While generally considered a milder lesion, base modifications (such as those arising from DNA alkylation), if unresolved, alter the pattern of hydrogen bonding between bases of opposite strands, possibly directing the incorporation of an incorrect base during the next round of DNA replication[130,131].

Direct reversal is a quick fix for some of these lesions - DNA alkyltransferases can catalyse the removal of alkylation products. Of these, MGMT covalently displaces the alkylation on guanosine at the O^6 position, rendering itself unavailable for further catalysis, and thereby being referred to as a suicide enzyme[132]. This is a particularly costly process to the cell, since immediate turnover by the proteasomal machinery[133] is required, but it demonstrates a preference for eliminating these lesions quickly. Where a base modification

blocks replication, a monoubiquitylation switch in the replicative clamp PCNA occurring at K164 in human PCNA can direct polymerase switching from the high-fidelity pol δ or pol ε to more error-prone, but bypass-competent, Y-family translesion polymerases such as REV1[134], pol η[134–136], pol κ[136,137], pol ι[135]. This switch effectively delays the issue so that a compensatory pathway can resolve the modification at a less critical cell cycle stage[138].

One pathway, base excision repair (Figure 1.2), engages DNA glycosylases as sensors that scan the genome for alkylated bases and are triggered by minor distortions of the DNA helix[139]. Lesion verification by base-flipping (bending of DNA such that a base is flipped out of the helix) is a common mechanism for all DNA glycosylases[140,141]. Damaged bases are processed by cleavage of the N-glycosidic bond[141–143] tethering the base to the DNA backbone so that an abasic (AP) site forms. An AP endonuclease further processes these AP sites into a single-strand break with 3′ OH group and 5′ deoxyribose phosphate (dRP) termini[144,145]. When bifunctional glycosylases are employed, an additional AP lyase activity cleaves the strand to either give single-strand breaks of 3′ α,β-unsaturated aldehyde with a phosphate (PUA)/5′ phosphate (through the β-elimination reaction of OGG1 or NTH1) or 3′ phosphate/5′ phosphate (from the β and δ elimination by NEILs)[146–148]. In all cases, the product of APE1 or bifunctional glycosylases requires end cleaning to generate ligatable 3′ OH and 5′ phosphate: the repair polymerase, pol β, with its intrinsic dRP lyase activity, can purge 5′-dRP[149]; APE1 can remove 3′ blocking PUA ends[150–152]; and polynucleotide kinase (PNK) can remove a 3′ phosphate[153,154]. A base, however, is still missing, and at this point, BER diverges into a short-patch (or single-nucleotide, SP-BER) or long-patch (LP-BER) branch[155]. The polymerase, pol β, usually favouring SP-BER, fails to repair oxidised dRPs, instead being covalently trapped at these DNA sites[156]. Therefore, long-patch repair is adopted[157]. The arrival of canonical replication machinery can promote LP-BER[158,159]; PCNA coordinates the polymerases pol δ or pol ε to fill-in the gap and even extend beyond it, displacing a small 5′-flap which like that in Okazaki fragment maturation, can be cleaved by PCNA-recruited FEN1[160,161]. DNA ligase I (LIG1) ligates the resulting single-strand nick[162,163]. In SP-BER, single-nucleotide gap filling by pol β is followed by ligation, predominantly performed by LIG3, a handover that is likely coordinated by the LIG3-scaffold, XRCC1[164–166].

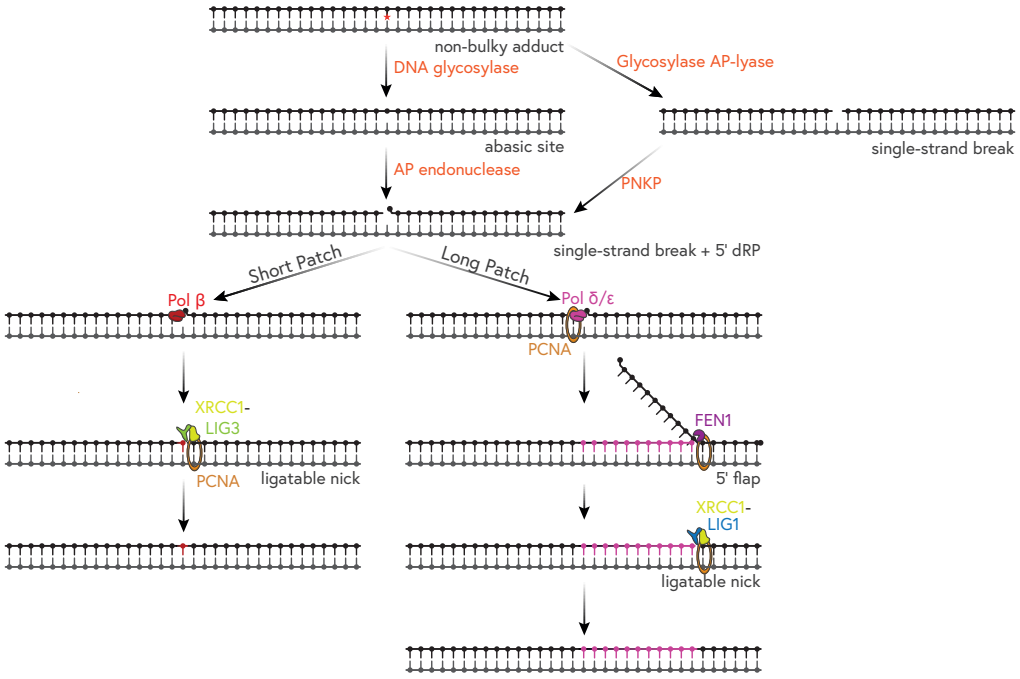


Figure 1.2 | Schematic representation of the base excision repair (BER) pathway. Non-bulky DNA lesions are recognised and excised by DNA glycosylases, which generate an abasic (AP) site. If a single-strand break is introduced, glycosylase AP-lyase or polynucleotide kinase phosphatase (PNKP) processes the lesion. AP endonuclease cleaves the AP site, generating a substrate for downstream repair. Two sub-pathways, short-patch BER and long-patch BER, then resolve the lesion. In short-patch BER, DNA polymerase β (pol β) inserts a single nucleotide, and XRCC1-LIG3 ligates the ensuing nick. Long-patch BER involves longer extension by DNA polymerase δ/ε (pol δ/ε), supported by PCNA, which displaces the damaged strand to create a 5' flap. FEN1 cleaves the flap, and LIG1 seals the nick to complete repair.

1.3.1.1 | Single-strand break repair

Single-strand break repair can be considered a component of the BER pathway since modified bases are channelled towards an SSB. SSBs also arise from sugar damage, *i.e.* where ROS can fracture oxidised deoxyribose[167]. The abortive activity of topoisomerase (TOP1), forced by collision with RNA pol II or proximity to other lesions, creates a TOP1-linked SSB (TOP1-SSB). These SSBs are immediately sensed by PARP1, activating it and stimulating the swift accumulation of repair factors, especially XRCC1[168,169]. Activation and recruitment of other repair factors depend on poly(ADP-ribose) (PAR) polymerisation (also termed PARylation) on PARP1 itself, as well as on other factors in the vicinity. Poly(ADP-ribose) glycohydrolase (PARG) restores PARP1 to its de-ribosylated state and is key for recycling PARP1 for ensuing rounds of SSB detection[170]. PARylation is

an NAD^+-dependent process[171], indirectly requiring ATP and therefore can be a taxing process for the cell, sometimes leading to necrosis[38].

1.3.1.2 | Ribonucleotide excision repair

The genome is afflicted with the accidental incorporation of ribonucleotides into DNA by replicative DNA polymerases, for example, yeast pol ε, at an alarming maximal rate of 1 every 1,250 bp[172,173]. This establishes embedded ribonucleotides as the most abundant form of DNA lesion. Replicative DNA polymerases correctly incorporate a deoxyribonucleotide $10^4/10^5$ fold over ribonucleotides, but a 36-190 fold higher concentration of ribonucleotides gives rise to this misincorporation rate[173]. The consequence of embedded ribonucleotides is that its $2'$ hydroxyl risks DNA backbone cleavage, endangering genome stability. A unique repair pathway, ribonucleotide excision repair (RER), has evolved to alleviate this 'ribo-damage'. RER is initiated through an incision by RNase H2 of ribonucleotide-embedded dsDNA, which releases ends - $3'$ hydroxyl on one end, and $5'$ phosphate ribonucleotide on the other[172,174,175] - that resemble intermediates processed by factors involved in Okazaki fragment maturation, which typically expel $5'$-RNA-terminated DNA flaps. With the $3'$ OH in place, pol δ, or less efficiently, pol ε, can perform DNA synthesis, leading to strand displacement and creating the flap resembling an Okazaki fragment structure[173]. Such a structure is an ideal substrate for removal by FEN1 or EXO1[176]. A DNA ligase, most commonly LIG1, then seals the remaining nick[174].

1.3.2 | Mismatch repair

The fundamental role of mismatch repair (MMR) is to correct misincorporations that have escaped polymerase proofreading. The failure of mismatch repair (MMR) captures a specific hallmark of cancer biology: genomic instability through the disruption of microsatellites (also known as microsatellite instability or MSI)[177,178]. Microsatellites are a catch-all term for the $\sim$3% of the eukaryotic genome consisting of repetitive sequences[179]. As a result of the sequence, there can be considerable diversity in the length of these repeats within a population and even between alleles within an individual. They may exist as mononucleotide repeats, such as poly(A) tracts[180]. Tandem repeats of dinucleotides also exist, such as the $(TA)_n$ repeats whose expansion confers significant stalling of

replication forks from a propensity of long TA stretches to form non-B DNA cruciform structures[181]. Tandem repeats can consist of larger nucleotide repeats; the expansion of repeats consisting of three nucleotides causes trinucleotide repeat disorders[182], exemplified by $(CGG)_n$ repeats in fragile X syndrome[183] or $(CAG)_n$ repeats in Huntington's[184,185]. Such repeats may have arisen and propagated throughout evolution due to the slippage of these sequences during replication. However, a few also comprise certain protein repeat domains, such as armadillo repeats[186]. The restoration of these common base mismatches is so crucial that the mismatch repair mechanisms are essential even in *E. coli*, where genes *mutS*, *mutL*, *mutH* and *uvrD* were implicated early in MMR research[187,188]. Through comparative genomics, the homologous system in eukaryotes was elucidated[189,190]. Human MutS and MutL heterodimers (MutLα: PMS2-MLH1, and MutSα - MSH2-MSH6; or MutLα-MutSβ(MSH2-MSH3) or MLH1/MLH3-MutSβ) translocate away from detected mismatches and generate a nick on the DNA[191] lesion-side oriented from PCNA, which helps guide the repair of the newly synthesised strand[192,193]. EXO1 then excises the nicked strand in a $5'$-$3'$ direction past the mismatch[194,195]. This action is likely redundant with the proof-reading exonuclease functions of pol ε and pol δ[196], but together, allows PCNA-pol δ to complete DNA re-synthesis and LIG1 to ligate the remaining nick[197,198].

1.3.3 | Nucleotide excision repair (NER)

NER is uniquely agnostic to a great variety of structurally distinct DNA lesions. One branch of NER, global genome NER (GG-NER), engages UV-DDB proteins to scan the genome for distortions to the DNA helix[199], while the other, transcription-coupled NER (TC-NER), is triggered by RNA pol II stalling at a lesion on the template strand[200] (Figure 1.3).

In GG-NER, XPC stabilised alongside UV excision repair protein RAD23 homolog B (RAD23B)[201] and centrin 2 (CETN2)[202,203], continually browses DNA for signs of DNA adducts through distortions to the regular helix, unveiling ssDNA zones that can be bound[103,204]. This is aided by a complex of DDB1 and DDB2 (the UV-DDB or UV radiation-DNA damage-binding protein complex), which acts as an additional damage-recognition factor by encouraging the binding of XPC[205,206], especially to minimally distorting cyclobutane-pyrimidine dimers (CPD)[207,208]. As the XPC complex binds damage, RAD23B dissociates, perhaps unveiling an XPC DNA-binding interface that

ensures the formation of the damage complex[209]. TFIIH is a transcription initiation and excision repair complex consisting of ten protein subunits[210], providing a binding surface to XPC[211]. Two TFIIH-resident DNA helicases with opposing polarities, XPB (*ERCC3*) and XPD (*ERCC2*)[212], expand a stretch of ssDNA to form an NER bubble, which allows for damage verification by XPD[213,214], aided by XPA[215]. The 5′-3′ unwinding activity of XPD is a vital step in NER, and its XPA-assisted activity scans for helicase-blocking lesions[216], in a mechanism defined by undamaged ssDNA passing freely through its internal channel but halting at damaged DNA[214,217,218]. Where XPD doesn't detect any damage, NER may be aborted[214], and XPA assists in damage verification by sensing ssDNA nucleotides with altered chemistry[215]. The CDK-activating subcomplex (CAK) of TFIIH is solely a transcription initiation factor, which is released by association with XPA and XPG upon its recruitment[219], essential for de-repressing CAK-dependent inhibition of XPD helicase activity[220].

1.3.3.1 | Transcription-coupled NER

Encountering inefficiently repaired lesions (such as CPDs, even despite UV-DDB activity) can stall replication, eventually leading to DSBs. Yet, damage tolerance mechanisms, such as template switching during HR (Section 1.3.4.2) and/or lesion bypass, alleviate these scenarios[138]. However, unlike in polymerase switching, RNA pol II cannot be switched to enact a bypass for helix-distorting lesions, and such a mechanism comes at the cost of a mistranscribed mRNA[222,223]. Hence, where RNA pol II has stalled, TC-NER machinery promotes the essential activity that repairs the intervening damage and reinstates RNA pol II transcriptional activity. This process is incomplete in individuals with Cockayne syndrome[224]. Cockayne syndrome factors CSA (*ERCC8*) and CSB (*ERCC6*) are recruited to a lesion-stalled RNA pol II through a constitutive engagement with elongating RNA pol II[225]. Lesion-stalling, in turn, amasses core NER factors, except for GG-NER-specific proteins, UV-DDB and XPC, but additionally including TC-NER-specific UVSSA (UV-stimulated scaffold protein A), USP7 (ubiquitin-specific-processing protease 7), XAB2 (XPA-binding protein 2; SYF1) and HMGN1 (high mobility group nucleosome-binding domain-containing protein 1)[226]. Despite typical repair of oxidative damage by BER (Section 1.3.1) and RNA pol II's ability to continue past oxidised bases[227], CSB has an

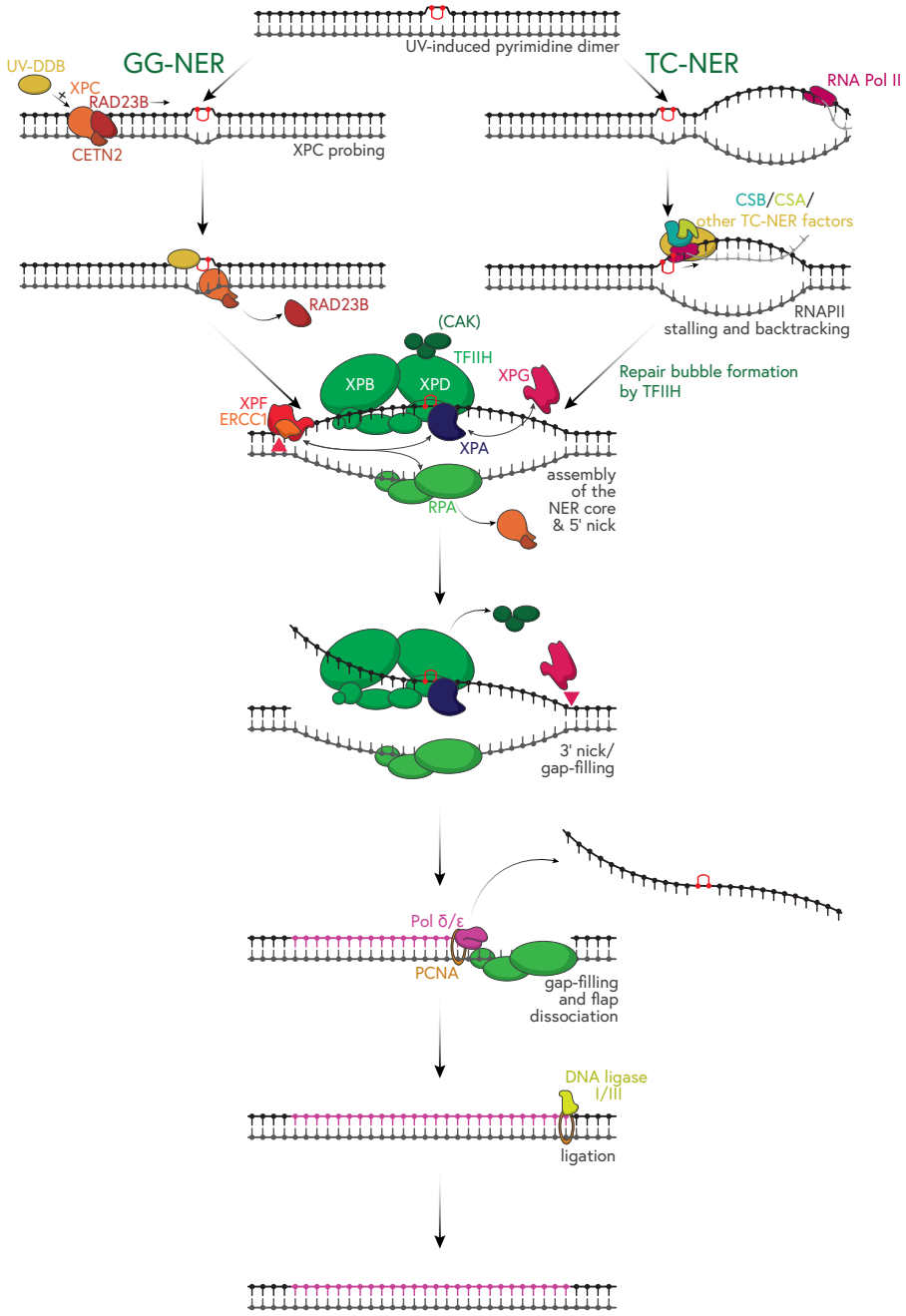


Figure 1.3 | Mechanisms of global genome (GG-NER) and transcription-coupled nucleotide excision repair (TC-NER). Nucleotide excision repair (NER) addresses bulky DNA lesions, such as UV-induced pyrimidine dimers, through two sub-pathways. GG-NER is initiated by lesion recognition through UV-DDB (DDB1 and DDB2), XPC, and RAD23B, which probe DNA for damage, eventually triggering the recruitment of additional repair factors after lesion verification by UV-DDB. For TC-NER, activated by RNA polymerase II (RNA pol II) stalling at transcription-blocking lesions. CSB, CSA, and other TC-NER factors facilitate RNA pol II backtracking and recruit the repair machinery. In both pathways, TFIIH, containing XPB and XPD, unwinds the DNA around the lesion, forming a repair bubble. XPA and RPA stabilise the bubble, while XPF-ERCC1 and XPG create 5' and 3' nicks around the lesion, respectively. The damaged strand is excised, leaving a gap, which is primed for gap filling by DNA polymerase δ/ε in coordination with PCNA, and ligation follows. Adapted from[221].

NER-unique feature in accumulating at sites of local oxidative damage, in a transcription-dependent manner, suggesting a subset of lesions are sensed by BER, fruitlessly recruit TC-NER only to continue being repaired by BER instead[228,229]. Perhaps a more global and diverse transcription-coupled repair system exists, coalescing parts of BER and NER[221].

RNA pol II occupancy of the transcribed strand envelopes ∼35 nucleotides[230], abrogating the NER core machinery's access to the lesion. Models enabling access to NER machinery included backtracking, dissociation or even degradation of RNA pol II. A precedent exists for RNA pol II backtracking since, during initiation, backtracking is used to license productive transcription[231]. While backtracking hasn't yet been directly supported, CSB, CSA and ELOF1 direct the ubiquitylation of RNA pol II's RPB1 subunit at K1268[232,233], channelling towards the degradation of a persistently stalled RNA pol II[234], which is considered a last resort. The more favourable scenario involves the dissociation of RNA pol II by TFIIH, possibly through its ATP-dependent remodelling functions. The active state of TFIIH is promoted by remodelling functions of STK19[235].

The ability to abort NER prevents the formation of undesirable and irreversible DNA alterations. It follows, however, that a commitment to completing the pathway must proceed from any point where strands are incised, pivotal in preventing the lasting presence of toxic intermediates. Within the bubble structure formed, the ds-ss discontinuity is a prime substrate for incision 5′ to the lesion by the structure-selective endonuclease XPF (alongside its obligate partner, ERCC1)[236–239]. A second incision 3′ to the lesion by XPG, recruited by TFIIH[240] or independently, removes a 22-30 nt stretch containing the damaged nucleotide[221]. Herein, PCNA-RFC, DNA pol δ and ε, or DNA pol κ, govern gap filling, and either LIG1 or XRCC1-LIG3 seals the remaining nick to complete NER[241,242]. Assembly of XPA, XPG and RPA at NER lesions is a prerequisite for the accurate coordination of incisions; XPA is regarded as a principal orchestrator of NER, binding almost all NER proteins, while RPA correctly orientates XPF and XPG[239]. Furthermore, it has been suggested that the XPF-ERCC1 5′ incision can instigate DNA synthesis, filling the intervening gap, even before XPG's cut on the 3′ side[236]. In this scenario, the extrusion of a 5′-flap might even create an improved substrate for XPG and also prevent excessive accumulation of ssDNA gaps, triggering disproportionate damage signalling[243].

It is important to note that NER deficiency results in complex disorders. Where GG-NER alone is perturbed, *i.e.* in XP-C (deficiency in XPC), individuals are singularly

predisposed to only slight photo-aversion, perhaps reflective of the greater diversity in lesion recognition of TC-NER[244]. Other XP patients with defects in proteins beyond the GG/TC branch point within NER suffer from any light exposure, with a concomitant substantial increase in the likelihood of skin cancer[244]. Some XP patients are also burdened with progressive neurological degeneration, usually where a more severe NER deficiency presides (*i.e.* XP-A)[245,246]. While the brain is not exposed to UV radiation, the progressive nature of the disease suggests an inability to repair ongoing endogenous damage, perhaps such as oxidative damage[247]. In post-mitotic neurons, such unrepaired damage can impair regular function and lead to cell death[248]. Neurological degeneration and physical abnormalities, alongside classical NER-deficient photosensitivity, distinguish Cockayne syndrome (CS)[249]. Here, CSA and, in particular, CSB deficiencies likely result in a significant transcription-coupled repair defect, which has a disproportionately negative impact on cells with greater transcriptional output, such as neurons[250]. Given neurons impose a greater metabolic demand, increased release of ROS can create bulky DNA lesions, evading repair in CS patients[251].

1.3.4 | DSB repair

A consequence of double-strand breaks (DSBS) is their analogy to the ends of chromosomes. A discontinuous chromosome is capable of gross genomic rearrangements that materially alter a cell's genotype and can lead to apoptosis. Therefore, robust and varied DSB repair pathways defend against the persistence of these breaks. Where a DSB is encountered, two solutions are employed: the first immediately re-ligates the two ends together, and the other copies around the break using a homologous template. The former process is termed non-homologous end-joining (NHEJ) and actively competes for DSBs with the latter pathway, termed homologous recombination (HR). Emerging work suggests redundancy with other sub-pathways, usually arising from a failure to repair by NHEJ or HR[252]. The fate of a DSB's resolution can be considered to fall within a decision tree, taking inputs of the cell cycle stage and the presence/absence of various factors[253] (Figure 1.4). For example, it appears that the two major DSB repair pathways are repressed during mitosis since any attempt to repair *via* classical NHEJ or HR in mitosis is potentially disastrous to the cell, with end-to-end fusions of chromosomes and

chromosome exchanges/lagging chromosomes being particularly detrimental consequences of doing so[254,255]. Additionally, the chromatin context and other regulatory factors govern the preference of HR for S and G2, where a sister chromatid is available[256]. Conversely, cNHEJ appears to function through most of the remaining cell cycle[257].

1.3.4.1 | Non-homologous end-joining

Classical NHEJ (cNHEJ), distinguished from alternative end-joining (Section 1.3.4.3), is the primary pathway of DSB repair, rejoining two DNA breaks agnostic of the sequence at these ends. Early insights into the mechanisms of this pathway in mammals arose from studying V(D)J recombination or class-switch recombination (CSR) during lymphocyte development[258,259]. cNHEJ is a process intrinsic to the repair of the programmed DSBs that occur during these immune repertoire diversifying processes[260,261].

Ku70-Ku80 (XRCC5-XRCC6), an abundant nuclear heterodimer binding immediately to each end of a two-ended DSB[262,263] is a crucial instigator of cNHEJ. The dimer, with high affinity for blunt DNA ends or those with short ssDNA overhangs, both shields the two ends and fulfils a landing-pad role, recruiting DNA-dependent protein kinase catalytic subunit (DNA-PKcs), DNA ligase IV (LIG4), and additional XRCC4, XRCC4-like factor (XLF) and paralogue of XRCC4 and XLF (PAXX)[264–266]. The supernumerary factors here appear as redundant scaffolding factors for LIG4 stabilisation, but XRCC4 is essential[267]. The transition from a long-range to a short-range synapse bridging two DSB ends has been proposed and later characterised structurally[262,268]. The long-range synaptic complex, with DNA ends more than 100 Å apart, is contorted through a combination of the Ku heterodimer, DNA-PKcs, LIG4, XRCC4 and XLF (Figure 1.4)[268]. Transitioning to the short-range synaptic complex requires XLF, XRCC4-LIG4 (though not its catalytic function) and the kinase activity of DNA-PKcs. However, this resembles a dynamic process where long-range to short-range transition cycles might precede cNHEJ processing[269]. A DNA-PKcs autophosphorylation trigger switches its role from end protection to processing[128]. Structures of this synaptic complex suggest that positioning two DNA-PKcs across the DSB gap allows for autophosphorylation *in trans*[270]. Artemis (SNM1C; gene *DCLRE1C*), a $5'$-$3'$ exonuclease bound and phosphorylated by DNA-PKcs, processes DNA ends, extricating them of non-ligatable DNA aberrations and together with

pol λ and pol μ, ensures DNA end compatibility[271]. The activity of these end-processing factors is restricted to the short-range synaptic complex, and Artemis is considered to lock DNA-PK in a kinase-inactive state[270]. Artemis can regenerate blunted ends where a DSB already contains an overhang, which can occur from non-blunt breaks or where resection has started. This serves to re-engage Ku70-Ku80, which may have fallen off at DSBs with lengthy ssDNA tails, thereby recommitting a DSB through repair by cNHEJ[253,272]. Analogously, fill-in can be achieved at resected DSB ends by a complex of CTC1, STN1 and TEN1 (CST)-pol α-primase (PRIM1, PRIM2)[273]. This complex is controlled by the NHEJ-promoting 53BP1 axis involving RIF1 and its associated shieldin complex (SHLD1-3 and REV7)[274]. To the detriment of the cell, however, trimming or filling in nucleotides to renew Ku binding promotes mutational repair, a significant consequence of cNHEJ that is fortunately advantageous to the generation of knockouts by CRISPR-Cas9[275]. DNA-PKcs phosphorylation, auto- and/or *via* the multifunctional DDR kinase ATM is critical for Ku70-Ku80 affinity and disassembly following ligation[128,276,277].

The MRE11-RAD50-NBS1 (MRN) complex has multifaceted roles in processing DSBs, mainly contributing to the progression of HR[278]. However, *Mre11* mutant mice, lacking entirely the MRN complex, showcase a severe CSR defect[279]. MRN could promote cNHEJ by abridging distant breaks[280] or, using its $3'$-$5'$ exonuclease activity, partially processing DNA termini with $3'$ blocking lesions[281]. However, it might instead reflect some limited resection by MRE11, promoting MMEJ (Section 1.3.4.3).

1.3.4.2 | Homologous recombination

The other major pathway to resolve DSBs is HR. HR, through faithful copying of a template, usually ensures no unwanted nucleotide deletions/insertions are introduced. Exhaustive links between HR genes and cancer predisposition showcase how HR deficiency can guide genomic instability, a hallmark of cancer[282,283]. By chance, it is the prevalence within certain cancers for HR deficiency (*i.e.* the 15% of ovarian cancers harbour *BRCA1/2* mutations) that has permitted the success of drugs that overload tumours with unrepairable DSBs[284]. HR culminates in the loading of RAD51 to generate an ssDNA-RAD51 oligomer filament, equipped with a homology searching function[285]. For suitably long ssDNA tracts to exist, HR either relies on resecting blunt DNA ends or that an ssDNA gap of sufficient

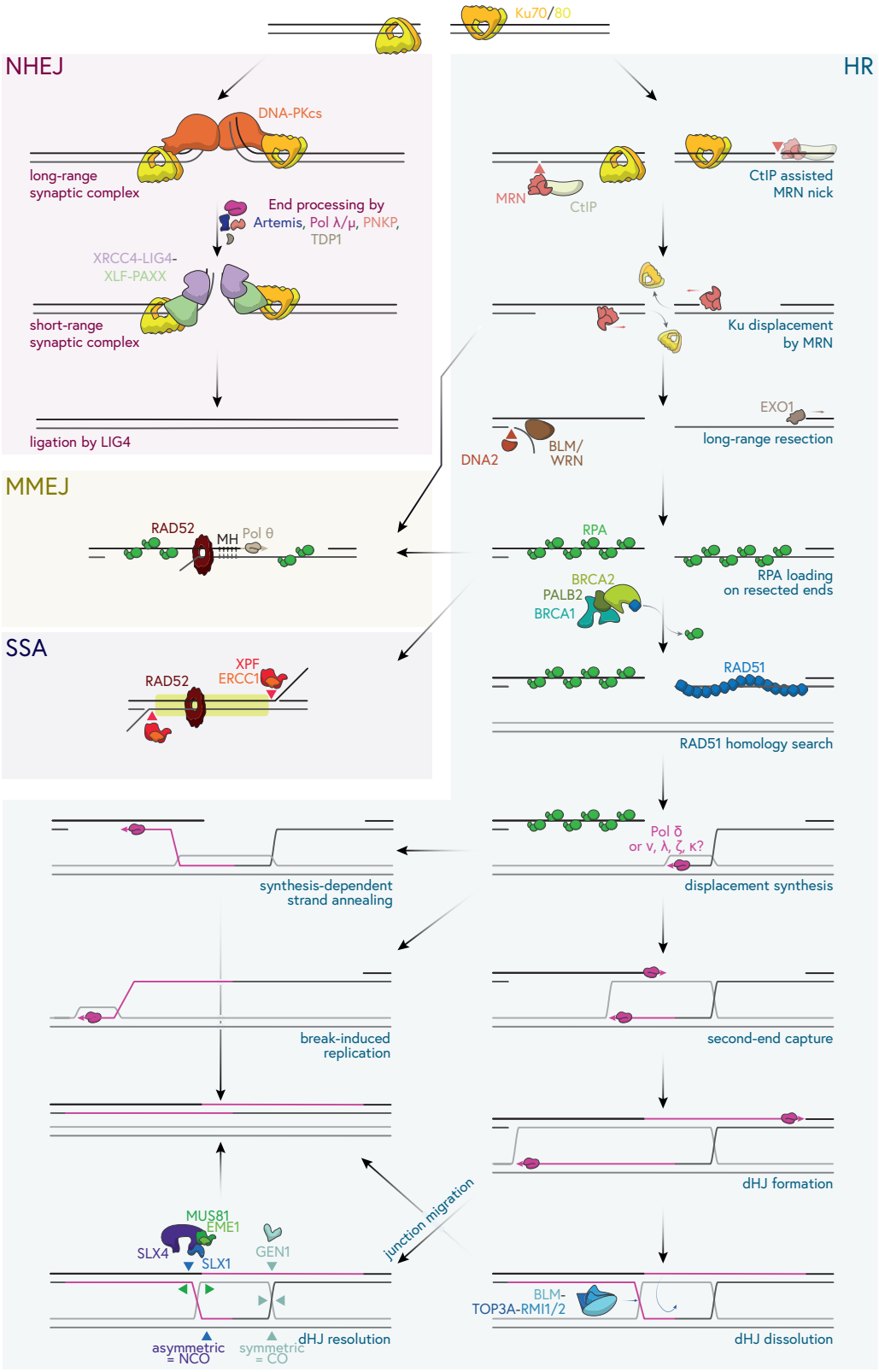

Ku70/80
NHEJ
DNA-PKcs
long-range synaptic complex
End processing by Artemis, Pol λ/μ, PNKP, TDP1
XRCC4-LIG4-XLF-PAXX
short-range synaptic complex
ligation by LIG4
HR
MRN
CtIP
CtIP assisted MRN nick
Ku displacement by MRN
EXO1
long-range resection
DNA2
BLM/WRN
RPA
RPA loading on resected ends
BRCA2
PALB2
BRCA1
RAD51
RAD51 homology search
Pol δ or ν, λ, ζ, κ?
displacement synthesis
second-end capture
dHJ formation
BLM-TOP3A-RMI1/2
dHJ dissolution
junction migration
MMEJ
RAD52
MH
Pol θ
SSA
RAD52
XPF ERCC1
synthesis-dependent strand annealing
break-induced replication
MUS81 EME1
SLX4
SLX1
GEN1
dHJ resolution
asymmetric = NCO
symmetric = CO

Figure 1.4 | Major double-strand break repair pathways. Double-strand breaks (DSBs) in DNA are repaired through distinct pathways based on end processing and sequence homology. During non-homologous end joining (NHEJ), DNA ends are recognised by the Ku70/80 heterodimer, followed immediately by DNA-PKcs, which phosphorylate and activate downstream repair factors. Of these, Artemis, pol λ/μ, PNKP and TDP1 prepare the DNA ends for ligation by the XRCC4-LIG4-XLF-PAXX complex. Ku70/80 binding can instead be followed by MRE11-NBS1-RAD50 (MRN) binding, which, alongside BRCA1-assisted CtIP, performs short-range resection. The switch to long-range resection involves either the 5′-3′ exonuclease activity of EXO1 or 3′-flap excision by DNA2 of WRN/BLM-unwound DNA. RPA immediately coats and stabilises single-stranded DNA, which is displaced by RAD51 in a manner dependent on BRCA1-BRCA2-PALB2 for filament assembly. RAD51 mediates homology search and strand invasion, which is followed by pol δ displacement synthesis that can either continue to the end of the chromosome in conservative break-induced replication (BIR) or lead to the capture of the second end, forming a double Holliday junction (dHJ). In some cases, synthesis leads to strand annealing to the second end, referred to as synthesis-dependent strand annealing (SDSA), which results in the non-crossover resolution of a DSB. For second-end capture, the resulting dHJ is dissolved by the branch migration of BLM and strand disentanglement by TOP3A of the BLM-TOP3A-RMI1-RMI2 (BTR) complex. Occasionally, dHJs are resolved through either symmetric nuclease cuts by GEN1 (which leads to crossovers) or asymmetric incisions by the SLX4-scaffolded MUS81-EME1 or SLX1. After short-range resection, uncovered short homologies can instead be annealed and stabilised by RAD52, with fill-in synthesis dependent on DNA polymerase θ, also facilitating alignment and ligation of broken ends. In single-strand annealing, which occurs as RAD51 becomes limiting, RAD52 aids the annealing of uncovered homologous repeats at the resected ends. A 3′-flap is extruded and cleaved by XPF-ERCC1, with ligation following to restore dsDNA, albeit with loss of repeat sequences.

length persists during replication[286,287]. The latter scenario partly describes HR's close connection to the rate of replication - HR-deficient cells have reduced replication fork speed[288]. In any case, DNA end resection is an obligatory licensing step for HR.

Eukaryotic end resection begins with the endonuclease activity of loaded MRN, which also serves a vital signalling role by activating another major DSB kinase, ATM[128,289,290]. Evidence points towards MRE11 nicking, up to 300 nucleotides upstream, the strand that leads to a 5′ terminus of a DSB, a particularly important process where the end is blocked, such as in the presence of TOP1-trapped cleavage complexes (Section 1.3.6.1)[291–293]. A 3′-5′ exonuclease function extends the initial nick towards the DNA break, effectively forging 'short-range' resection. CtBP-interacting protein (CtIP; RBBP8), facilitated by the BRCA-A complex[294], is stimulated by DNA end blocking through *e.g.* Ku70-Ku80, RPA or nucleosomes, and is required for efficient short-range resection by MRE11, which is thought to displace Ku70-Ku80[295–298]. A combination of activities mediates long-range resection to liberate a long 3′ ssDNA tail: 5′-3′ exonuclease activity of EXO1[299,300]; unwinding from the 5′ termini by Bloom syndrome helicase (BLM) or Werner helicase (WRN), freeing a 5′-flap structure to which DNA2 can excise through endonuclease activity[301–303].

A vast excess of an ssDNA binding complex, replication protein A (RPA; consisting of a trimer of RPA1-3) guarantees that any free ssDNA stretches are immediately coated and protected from annealing to other DNA and/or forming internal secondary structures[304]. In this protective role, however, RPA is a barrier to the loading of RAD51, and this displacement must be enforced by recombination 'mediators'[305,306]. Rad52, in yeast, is the key recombination mediator[307], with BRCA2 having an equivalent role in vertebrates, competing with RPA for binding to ssDNA[306]. BRCA2 constitutively binds the 26S proteasome complex subunit DSS1 (SEM1), RAD51 monomers, and BRCA1-BARD1 *via* PALB2 (partner and localiser of BRCA2)[308]. The necessity of BRCA1 to HR, and the viability of organisms in itself[309], is multifaceted, being involved in coupling initial CtIP stimulation[294] to the mediation of RAD51 loading alongside PALB2-BRCA2[310] and stimulation of the homology search function of RAD51[311]. Once formed, the stability of the nucleofilament of RAD51 and ssDNA is not settled, and multiple regulatory activities stabilising and disrupting this filament optimise HR efficiency while restraining homology search to appropriate genome regions. RAD51-ssDNA performs homology search by invading duplex DNA and facilitating base-pairing with complementary sequences[312]. Through a three-stranded helix intermediate, sufficient base-pairing stabilises the synapse and forms a displacement loop[313].

The appropriate RAD51-mediated synapse is a node with several distinct outcomes. The primary pathway, synthesis-dependent strand annealing (SDSA), elicits invasion of only one end of a two-ended DSB, while the second remains passive but resected[314,315]. Strand displacement to the passive resected second end allows displacement of the template strand[316]. Evidence is lacking on how the asymmetry between two DNA ends is established. In yeast, mutants lacking MRX or Sae2 (CtIP homolog) show asymmetric resection of IR-induced DSBs, suggesting MRX/MRN may control invading-non-invading strand interactions[317]. Without forming a Holliday junction, SDSA remains a non-crossover pathway, suppressing the extent of gene conversion. A secondary branch of D-loop processing is predominant in meiosis, where the resected second end captures the region displaced in the D-loop, forming an entangled DNA structure referred to as a double Holliday junction (dHJ)[318,319]. The resolution of a dHJ requires the action of structure-specific endonucleases. GEN1, or the dimeric MUS81-EME1 (likely within

a complex with SLX4) can separately make opposing incisions to disentangle the dHJ, but sometimes with crossover events where asymmetric nicks occur[320–324]. Dissolution, instead, avoids crossovers. Here, branch migration of the two HJs towards each other leads to a hemicatenated intermediate, which can be dissolved through topoisomerases. Housing a ssDNA-specific type IA topoisomerase, TOPIIIα, the dissolvase complex (BTR) also includes the branch-migrating RecQ helicase, BLM[325] and structural proteins RMI1-RMI2. The study of shorter model substrates to identify mechanisms has obscured the influence of nucleosomes in this process. Still, with the advent of 1.8 kb DNA morphed into dHJs, it has recently been demonstrated that chromatin inhibits GEN1- or SMX-dependent resolution while permitting BTR-mediated dissolution[318]. Failure to disentangle these joint intermediate molecules can subvert HR into a pathway of continual error-prone replicative recombination - faulty re-engagement of the second end or displacing the nascent strand promotes uncontrolled migration of the D-loop bubble through a chromosome - this process is named break-induced replication, a conservative DNA synthesis process entailing long-tract gene conversion that can lead to a gross loss of heterozygosity (LOH)[326,327].

1.3.4.3 | Alternative DSB repair pathways

Where RAD51 loading is faulty, or its strand invasion function does not result in productive HR, an opportunity arises for error-prone pathways such as MMEJ or single-strand annealing (SSA). Resection unveils chance sections of limited homology at the remaining 3′ overhangs.

Microhomology-mediated end-joining (MMEJ) MMEJ, or sometimes alternative end-joining (aEJ), is, despite its name, devoid of the key end-joining factors, DNA-PKcs, Ku, or LIG4, and is initiated through the generation of short single-strand overhangs by MRE11 and CtIP[328,329]. Microhomologies at the breakpoint, defined as 5-25 base pairs of complementarity though most commonly 4-6 bp, anneal with the help of the polymerase pol θ, encoded by the pol θ gene[330–335], which also displaces protective RPA from 3′ ssDNA overhangs[336]. pol θ, with reported helicase-like activity[331], may then also stabilise the synapsis between the ends[337]. After annealing of internal homologies, APEX2 and FEN1 cleave the extruded 3′-flaps[338,339] while pol θ extends these intermediates[340,341] and XRCC1-LIG3 completes end-joining[342]. While pol θ is a central cog in MMEJ, it is also its

erroneous synthesis that confers MMEJ mutational hallmarks[341]. Its recruitment to DSBs is mitosis-constricted; RHINO, an interacting partner of the 9-1-1 complex (RAD9A-RAD1-HUS1), only assembles in mitosis *via* polo-like kinase 1 (PLK1) phosphorylation[343].

Single-strand annealing (SSA) Homologous 3′ ends can anneal in a RAD51-independent manner when extensive resection reveals underlying repeat sequences, such as those found at tandem repeats[344,345]. This annealing activity results in the displacement of the DNA sequences located 3′ to the annealed repetitions (usually 29-400 bp in length[346]) into flap structures, leading to the loss of these intervening DNA sequences. Consequently, single-strand annealing (SSA) is inherently error-prone. This pathway also requires the disassembly of RPA from the 3′ tails to allow the two ends to anneal. In yeast, the displacement of RPA is mediated by the SSA-essential factor RAD52, however as the key RAD51-mediator, RAD52 might promote either HR or SSA. In mammals, the derivation of BRCA2 to fulfil RAD51 loading requirements frees RAD52 for a predominant role in SSA. However, RAD59[346] and FANCA, alongside FANCG, have been implicated in SSA[347]. Given the association of SSA with the loss of intervening repeat sequences, it is conceivable that this pathway evolved to counteract repeat expansions[348]. However, deleted DNA sequences can come at an appreciable cost. The observed correlation between high RAD52 expression and poorer outcomes in human cancers suggests that SSA may accelerate genome instability, thereby driving cancer progression[349].

The non-homologous 3′ ends flanking annealed stretches generate 3′-flaps that require removal. In yeast, this removal is catalysed by the $Rad1^{XPF}$-$Rad10^{ERCC1}$ complex[350], with assistance from the flap-stabilising MSH2-MSH3 complex[351]. In mammalian cells, RAD52 directly recruits the XPF-ERCC1 complex[352,353], whereas in yeast, Saw1 facilitates the connection between these components[354]. The gap-filling step likely involves the activities of polymerases δ, ζ, or θ, although the exact polymerase responsible remains undetermined[355–358]. Similarly, the specific ligase involved in sealing the repaired DNA ends has yet to be identified[358].

Drugging the alternative DSB repair pathways The synthetic lethality of pol θ loss in HR/NHEJ deficiency[359] suggests MMEJ as simply a backup pathway acting where preferred DSB repair is unavailable. However, recent evidence supports the claim that MMEJ is a first-line repair route for certain DSBs[360]. Alongside NHEJ, MMEJ promotes the random integration of foreign DNA into the genome and aids in the repair of CRISPR-Cas9-induced DSBs at certain loci[361,362]. Where RHINO was discovered as a crucial MMEJ factor, it was also noted that mitotic MMEJ is necessary for the resolution of DSBs originating from and escaping repair during S phase[343]. While considerably mutagenic, MMEJ and SSA alleviate possible structural rearrangements in DNA caused by DSBs and are necessary for the survival of cells lacking sufficient HR and NHEJ, providing a key synthetic lethal relationship explored in clinical trials of pol θ or RAD52 inhibitors[363,364]. The efficacy of pol θ inhibitors was showcased through its specific targeting of BRCA-deficient tumours, synergy with PARPi, and ability to eliminate certain PARPi resistant tumours[365].

1.3.4.4 | DSB signalling

At its most fundamental level, signalling around DNA double-strand breaks (DSBs) can be divided into three overlapping cascades, each mediated by one of the apical kinases of the phosphoinositide 3-kinase (PI3K)-related kinase (PIKK) family[128]. Activation of these kinases—ATM, ATR, and DNA-PKcs—coordinates an intricate phosphoproteome that regulates processes such as DNA repair, cell cycle control, DNA replication, transcription, RNA metabolism, senescence, and apoptosis[366]. At DSBs destined for classical non-homologous end joining (cNHEJ), Ku-bound DSBs recruit and activate DNA-PKcs[128,367]. Conversely, in contexts favouring homologous recombination (HR), the MRN complex recruits and activates ATM *via* an interaction with NBS1[289,290]. At sites with extensive single-stranded DNA (ssDNA), such as resected DSBs or stalled replication forks, ATR is recruited by its binding partner ATRIP to RPA-coated ssDNA[368].

ATM, once activated by the MRN complex, phosphorylates the histone variant H2AX at serine 139, producing γH2AX, a critical marker for DSB signalling and repair[369]. A positive feedback loop amplifies this signal: MDC1 binds phosphorylated H2AX *via* its BRCT domain[370] and is stabilised on chromatin by ATM-dependent phosphorylation at

threonine 4[371,372]. This, in turn, recruits MRN through interactions with NBS1, locally amplifying ATM signalling. γH2AX domains extend sufficiently to enable detection of DSBs by γH2AX foci visualised through microscopy[373] or chromatin immunoprecipitation (ChIP), which has revealed γH2AX tracks spanning more than 20 Mb[374]. ATM phosphorylates CHK2 at multiple sites, notably threonine 68[375], and CHK2 further phosphorylates p53 at serine 15, promoting the transcription of the CDK inhibitor p21. This inhibits the Cyclin A/E-CDK2 complex, preventing entry into S phase[376]. At ssDNA regions surrounding a DSB or stalled replication fork, RPA recruits ATR *via* ATRIP[368]. ATR activation requires cofactors such as ETAA1 or TopBP1, the latter of which involves the 9-1-1 clamp complex in its activation of ATR[377]. Through an ATR-CHK1-CDC25A axis, ATR regulates the DNA damage-dependent restriction of mitotic entry. ATR phosphorylates CHK1 at serines 317 and 345[378], and this activated CHK1 targets CDC25A for proteasomal degradation, thereby preventing the removal of inhibitory phosphate groups on WEE1 and PKMYT1[379]. This inhibition allows these kinases to phosphorylate and inactivate mitotic CDKs (CDK1 and CDK2), halting the cell cycle[380,381]. ATR also prevents replication fork collapse under conditions of replication stress by restraining replication origin firing, conserving nucleotide pools, and preventing the exhaustion of repair factors such as RPA[61]. Furthermore, ATR-dependent upregulation of RRM2 increases nucleotide availability[382].

Repair pathway choice at DSBs is tightly regulated by the cell cycle stage to maintain genomic stability. Key regulators 53BP1 and BRCA1 act antagonistically, with 53BP1 promoting NHEJ and suppressing HR, while BRCA1 promotes HR by countering 53BP1's end-protective activities. 53BP1 recruits PTIP and RIF1 following ATM-dependent phosphorylation[383,384], whereas BRCA1 suppresses these activities through multiple mechanisms, including CDK1-dependent phosphorylation of CtIP, which prevents RIF1 recruitment. Both factors localise to ionising radiation-induced foci, but S phase foci exhibit BRCA1 accumulation and 53BP1 exclusion[385]. This exclusion may involve BRCA1-dependent ubiquitylation of H2A, recruiting SMARCAD1 to antagonise 53BP1 or destabilising RIF1 binding. The interplay between 53BP1 and BRCA1 is further modulated by a methylation switch on histone H4. 53BP1 binds the H4K20me1 and H4K20me2 marks found on 'old' histones, which are more abundant in G1-phase chromatin, whereas BARD1 (a BRCA1 binding partner) recognises non-methylated H4 found in post-replicative

chromatin[386–388]. This 53BP1-BRCA1 dynamic switch allows repair pathway choice to align with cell cycle stage, ensuring appropriate genomic maintenance, explaining why the deletion of 53BP1 rescues the HR defect of Brca1 mutant cells[389] and PARPi sensitivity[390]. Besides BRCA1/2 reversion mutations[391], any strategy that dampens NHEJ (*e.g.* LIG3 loss[392] or loss of Shieldin[393–396]), restoring the balance between HR and NHEJ in BRCA1-deficient cells, results in resistance towards PARPi.

1.3.5 | ICL repair

ICLs arise from the covalent linking of the two strands of duplex DNA and represent a particularly hazardous form of DNA damage[8]. The danger ensues from the block to essential homeostatic processes, transcription and replication, since the strands cannot be separated for reading by polymerases. Crosslinking can arise endogenously from reactive aldehydes or lipid peroxidation byproducts[397] and exogenously from exposure to bifunctional alkylating agents such as mitomycin C, cisplatin, or psoralen[398]. The unique structure of ICLs poses significant challenges to the cell, as repair requires the coordinated action of multiple DNA repair pathways. During S phase, a replication-coupled repair mechanism dominates and is initiated by stalling of replication forks at the crosslink site[399]. This process involves the Fanconi anaemia (FA) pathway, which orchestrates lesion recognition, nucleolytic incisions on either side of the crosslink, and subsequent translesion synthesis (TLS) across the adduct[400]. HR is then employed to restore genomic integrity. The consequence of mis-repaired crosslinks can result in chromosomal instability, cell cycle arrest, or apoptosis, underlining the importance of efficient ICL repair for genomic maintenance. Aberrations in ICL repair mechanisms are associated with several human diseases, including Fanconi anaemia, and contribute to the cytotoxicity of crosslinking chemotherapeutics used in cancer treatment[401].

1.3.5.1 | ICL repair deficiency in the disease, Fanconi anaemia

In 1927, Guido Fanconi reported three brothers presenting with symptoms resembling a pernicious anaemia[402,403]. A generation of DNA discoveries passed before Fanconi himself hypothesised that these symptoms arose from an underlying chromosomal instability[404]. By then, such conditions were aptly referred to as Fanconi's anaemia (hereafter FA). Others

then highlighted a key consolidating characteristic of FA, sensitivity to chemicals capable of covalently crosslinking the two strands of DNA. Notably, mitomycin C (MMC)- or diepoxybutane (DEB) caused instability of FA patient cells and chromosome breakage events therein, later adopted as the gold-standard diagnostic test for suspected FA cases[405]. Therefore, FA was deemed a pathology of inefficient ICL repair, distinguished by bone marrow failure[10], skin irregularities[406] and decreased fertility[407], perhaps caused by the faster accumulation of replication errors in these rapidly-dividing tissues. Since the cloning of *FANCC* in 1992, genetic dissection of FA has culminated in the discovery of, now, 23 FA complementation groups from *FANCA* to *FANCW* and encompass various stages of classical ICL repair[400] (Table 1.2).

1.3.5.2 | Classical model of ICL repair

Studies using plasmids containing site-specific cisplatin-induced ICLs in *Xenopus laevis* egg extracts have been foundational in elucidating the mechanisms of replication-coupled ICL repair. Current consensus suggests that ICL repair is intrinsically linked to DNA replication, ensuring that the intermediate breaks, which are unavoidable during repair, are addressed in a way that maintains genomic fidelity (Figure 1.5). Due to their topology, the repair of ICLs necessitates the generation of DNA breaks as an integral part of the process. Repair is initiated when two replication forks converge at the ICL[399], although whether this mechanism applies universally, particularly in cells with greater inter-origin distances ($\sim$100 kb[408]) compared to the size of plasmids used in experimental systems (5-6 kb[399]), remains debated[400]. TRAIP is an E3 ubiquitin ligase central to the DDR, such that its loss is lethal and mutations lead to primordial dwarfism[409,410]. In the context of ICL repair, when replisomes converge, TRAIP dictates pathway choice by depositing ubiquitin chains ahead of the leading replisome, especially on the upstream CMG helicase[411]. The glycosylase, NEIL3, can cleave either the N-glycosyl bond of an abasic site or psoralen ICL, effectively unhooking the ICL and denoting a likely preferred pathway that avoids the need for the entire FA pathway, which can be recombinogenic in part due to the unavoidable generation of one- or two-ended DSBs[412]. TRAIP promotes repair by NEIL3, through deposition of short ubiquitin chains on CMG sufficient for NEIL3 recruitment[411]. The chemistry of the lesion might render the ICL refractory

FA group	Alias	Function
FANCA		FA core complex, part of "AG20" subcomplex
FANCB		FA core complex, part of "BL100" subcomplex
FANCC		FA core complex, part of "CEF" subcomplex
FANCD1	BRCA2	Homologous recombination, promotes RAD51 filament formation
FANCD2		Monoubiquitylated by FA core complex, part of "ID2" complex
FANCE		FA core complex, part of "CEF" subcomplex
FANCF		FA core complex, part of "CEF" subcomplex
FANCG	XRCC9	FA core complex, part of "AG20" subcomplex
FANCI		Monoubiquitylated by FA core complex, part of "ID2" complex
FANCJ	BRIP1	DNA helicase
FANCL		FA core complex, part of "BL100" subcomplex, E3 ubiquitin ligase
FANCM		Recruitment of FA core complex, DNA:RNA helicase activity
FANCN	PALB2	Homologous recombination, interacting with BRCA1 and BRCA2
FANCO	RAD51C	Homologous recombination
FANCP	SLX4	Nuclease scaffold for XPF-dependent ICL unhooking
FANCQ	ERCC4, XPF	Nuclease responsible for primary incision
FANCR	RAD51	Homologous recombination
FANCS	BRCA1	Homologous recombination, promotes RAD51 filament formation
FANCT	UBE2T	E2 ubiquitin-conjugating enzyme
FANCU	XRCC2	Homologous recombination
FANCV	REV7	Translesion synthesis
FANCW	RFWD3	E3 ubiquitin ligase for turnover of RAD51

Table 1.2 | Summary of FA complementation groups, gene aliases, and functions.

to NEIL3 glycosylase activity[413], and where it (tries and) fails, TRAIP continues to lengthen the ubiquitin chain on CMG, eventually leading to its unloading by the p97 ATPase[414]. The unloading of the CMG helicase creates sufficient space for the leading strand polymerases to approach within one nucleotide of the ICL[399,415] and for one fork to undergo reversal, which generates a chicken-foot structure that protects the stalled replication fork[416]. The enzyme responsible for fork reversal and whether fork reversal is a prerequisite for ICL repair remains contentious to date. However, FANCM, alongside its partner FAAP24 is recruited to DNA crosslinks and stabilised by histone-fold proteins MHF1 and MFH2[417–419] and demonstrates fork reversal activity *in vitro*[94]. Either through

establishing the reversed fork structure or as a direct protein platform, the FANCM complex facilitates the recruitment of secondary FA proteins[420]. These secondary proteins form the FA core complex, composed of subcomplexes AG20 (FANCA, FANCG, FAAP20), BL100 (FANCB, FANCL, FAAP100), and CEF (FANCC, FANCE, FANCF)[421]. The core complex is responsible for recruiting and orienting FANCT (UBE2T), an E2 ubiquitin ligase, which associates with the FANCL E3 ubiquitin ligase that catalyses the conjugation of ubiquitin molecules to the tertiary FA proteins FANCI and FANCD2, collectively referred to as the D2-I complex[422–427]. The monoubiquitylation of FANCD2 at K561 is essential for ICL repair[428], while the monoubiquitylation of FANCI at K523 plays a regulatory role protecting the deubiquitylation of FANCD2 by USP1-UAF1[429–432]. FANCI is phosphorylated by ATR[433]. The ubiquitylated ID2 complex accumulates in foci that colocalise with recombination machinery and was initially believed to directly recruit the effectors required for repair and damage bypass[104,434,435]. D2-I binds and diffuses on dsDNA as a locked sliding clamp when FANCD2 is monoubiquitylated[1–3,108,109,430,436,437] but stalls at ssDNA-dsDNA junctions[438], such as the structures present when replication forks stall at an ICL. ATR primes the D2-I complex for ubiquitylation by phosphorylating FANCI at S558, S561 and T567[433]. The ubiquitylated residues of FANCI and FANCD2 could provide simple landing platforms for the recruitment of XPF (FANCQ) *via* SLX4 (FANCP), which contains two ubiquitin-binding UBZ domains (Section 1.4.4). While this notion was assumed for a number of years, the burial of a necessary hydrophobic patch, which includes I44 of ubiquitin, prevents engagement by ubiquitin-binding motifs[1,109]. Instead, the sliding clamp of FANCD2-FANCI likely protects intermediate structures, like reversed forks, until SLX4-XPF is recruited by some other means[439]. The UHRF1 scaffold protein has been reported to sense ICLs and recruit XPF-ERCC1 and MUS81-EME1, independently of D2-I and SLX4[440].

However it may be recruited, XPF and its obligate partner ERCC1, introduces an incision 5′ to the ICL, similar to XPF-ERCC1's role in NER[435,441–443]. The classical X-shaped DNA structure present at ICL-converged forks is not a typical XPF substrate due to the lack of dsDNA sequence at the junction[435,444]. Therefore, fork reversal might be a critical step in permitting XPF incision. However, the nuclease responsible for the complementary 3′ incision remains unidentified. Of the candidates, SNM1A appears the

most likely, based on its ability to digest ICL-containing substrates *in vitro*, its epistatic relationship with XPF in response to mitomycin C (MMC), and evidence linking *DCLRE1A* (encoding SNM1A) to FA genes in CRISPR screens[445,446]. Potential redundancy with other nucleases, such as FAN1 or SNM1B, which share similar nuclease polarities, might explain the moderate sensitivity of SNM1A-depleted cells to crosslinking agents[447,448].

Once the ICL is unhooked in this replication-associated FA mechanism, the intermediary structures remaining undergo divergent forms of repair. On the 5′ side of the crosslink, leading strand polymerases halt one nucleotide short of the unhooked ICL, with a gap extending to the downstream Okazaki fragment generated by the lagging strand polymerase. The polymerase(s) responsible for bypassing the unhooked ICL and incorporating a nucleotide opposite an ICL has yet to be definitively identified. TLS polymerases, such as REV1 and REV7 (MAD2L2 but also FANCV or pol ζ when complexed with REV3), are implicated in this step, as a product of their established roles in FA and lesion bypass[449–451]. On the unhooked strand, where nuclease incisions have generated a two-ended DSB, homologous recombination (HR) is the primary repair mechanism during S phase. This reliance on HR likely explains why conventional HR factors, including RAD51 and its associated loading factors BRCA1, BRCA2, PALB2, and the RAD51 paralogue complex BCDX2 complex components RAD51C and XRCC2, are categorised as FA complementation groups (FANCR, FANCS, FANCD1, FANCN, and FANCO, respectively)[452,453]. Additionally, the E3 ubiquitin ligase RFWD3 (FANCW)[454] and the 3′-5′ DNA helicase BRIP1 (FANCJ)[455], both HR-associated factors, are implicated in FA.

The final repair step involves resolving the residual mono-adduct remaining on one strand of the replicated DNA[456]. In *Xenopus* egg extracts, most of these mono-adducts persist indefinitely[449]. However, full resolution likely involves a combination of NER and/or BER, with the choice of pathway depending on the specific crosslink[457].

1.3.5.3 | Alternatives to the converging forks model of ICL repair

DNA combing experiments in human cells suggested that single replication forks passed through 60% of ICLs, conflicting with the converging forks model of ICL repair in egg extracts[458]. This process, ‘traverse’ is dependent on the ATPase activity of the FANCM complex (with MHF1-MFH2, but not FAAP24), ATR, FANCD2[459], and the

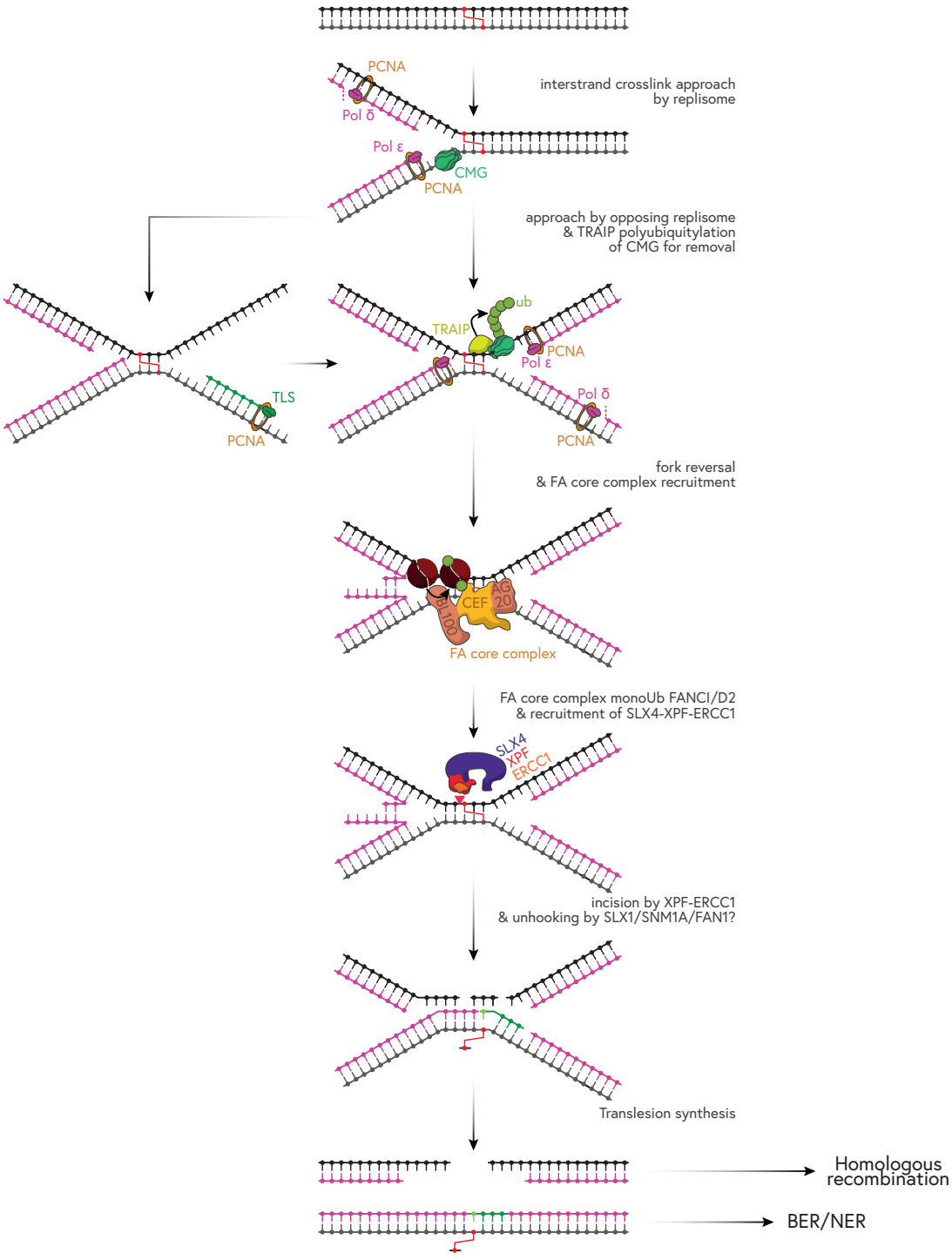

PCNA
Pol δ
interstrand crosslink approach
by replisome
Pol ε
CMG
PCNA
approach by opposing replisome
& TRAIP polyubiquitylation
of CMG for removal
ub
TRAIP
PCNA
Pol ε
TLS
PCNA
Pol δ
PCNA
fork reversal
& FA core complex recruitment
CEF
AC
20
100
FA core complex
FA core complex monoUb FANCI/D2
& recruitment of SLX4-XPF-ERCC1
SLX4
XPF
ERCC1
incision by XPF-ERCC1
& unhooking by SLX1/SNM1A/FAN1?
Translesion synthesis
Homologous
recombination
BER/NER

Figure 1.5 | The repair of interstrand crosslinks by the Fanconi anaemia pathway. Interstrand crosslinks (ICLs) block replication fork progression, triggering a coordinated repair process. When the replisome encounters an ICL, the CMG helicase stalls ahead of the leading strand PCNA-pol δ. Stalling recruits TRAIP, which polyubiquitylates and targets p97 to remove CMG from the replisome, allowing the approach of leading strand forks to the -1 position relative to the crosslink. Fork reversal may be carried out by FANCM-FAAP24-MHF1-MHF2, leading to the recruitment of the FA core complex, which monoubiquitylates each subunit of the FANCD2-FANCI heterodimer. In some manner, this promotes the recruitment of the SLX4-XPF-ERCC1 complex. This complex incises the DNA to the 5′ flank of the ICL, and unhooking ensues *via* some unknown mechanism but likely through the action of SNM1A or SLX1 or FAN1. Translesion synthesis bypasses the lesion, filling the gap left by the unhooking step and the remaining adduct may be resolved by base excision repair (BER) or nucleotide excision repair (NER). The remaining double-strand break is repaired by homologous recombination.

BLM helicase[460]. Traverse involves GINS displacement from the CMG helicase at ICLs, which is regulated by ATR phosphorylation of FANCM[459]. It is possible that the dissociation of GINS permits the CMG ring to open such that it can bypass an ICL block, even though GINS is not required for traverse of CMG through a DPC[461]. Even in traverse, SLX4 and XPF are required for unhooking[462], and the ensuing mechanisms for resolution are likely identical.

1.3.6 | DNA-protein crosslink (DPC) repair

While the above forms of damage showcase some form of grouped chemical identity, DNA damage, where proteins are covalently crosslinked to DNA (a DPC), comprises a diverse and heterogeneous set of DNA lesions. The DNA component, the chemical nature of the crosslink, and the exact type of protein tethered can vary greatly and distinguishing these variations is a key contributor to determining the precise mechanism adopted for DPC repair[463].

1.3.6.1 | Forms of DPCs

It is believed four major DNA-varying DPCs can form in cells. Endogenous DPCs can frequently occur on dsDNA from aldehydes persisting through catabolism. As an example, the Schiff base intermediate that materialises during a formaldehyde-induced crosslink can be attacked, instead of an opposing DNA base amine, by nucleophiles within basic amino acids such as arginine or lysine, or those of cysteine, histidine or tryptophan[464]. An important note on the ability of aldehydes to crosslink DNA is that this likely represents a minority of all aldehyde crosslinking events - instead, protein-protein

crosslinks, unencumbered by the need to disassemble Watson-Crick base pairing, are probably more readily formed[465]. This places an asterisk on attributing the cytotoxicity of aldehydes to a defective DNA repair phenotype. Chemotherapeutics can also generate artificial DPCs. Cisplatin and nitrogen mustards are suggested to induce DPCs, and cells deficient in the main DPC-removal pathway (*e.g.* that involving SPRTN) show sensitivity to cisplatin, implicating a fraction of this drug's chemotoxicity stems from the generation of DPCs[466,467]. The enzyme DNMT1 senses and corrects hypomethylated DNA regions, especially following DNA replication[468]. Incorporating a cytosine analogue, 5-aza-dC, traps DNMT1 behind the replication fork by preventing its β-elimination after DNMT1 methylates the daughter strand[469]. Radiation, such as UV and IR, also contribute to DPCs, especially in low-oxygen environments. Under hypoxic conditions that define certain solid tumours, IR-induced DNA radicals, usually channelled towards DNA decomposition under normoxic conditions, promote crosslink formation[470].

Despite identical chemistries underlying the crosslinks and an equal (if not higher) propensity for their formation, protein crosslinks on single-stranded DNA present a very unique substrate for repair. In some instances, DPC formation is actively preferred over erroneous processing, such as where a thiazoline link bonds HMCES (5-hydroxymethylcytosine (5hmC) binding, ES-cell specific) to abasic sites on ssDNA[471,472]. Where proteins enzymatically act on DNA, they may be covalently crosslinked as part of their catalytic cycle. Torsional stress that builds during transcription or replication is alleviated through the action of topoisomerases, which rely on a covalent link to DNA such that tangled strands (after an SSB in the case of TOP1)[473] or duplexes (after a DSB for TOP2)[474] can be weaved through one another to ensure their disentanglement. Vicinal DNA lesions or the action of intercalating agents like camptothecin or topotecan can displace the $5'$ hydroxyl that otherwise would release the TOP1-DNA phosphotyrosine link, thereby trapping the enzyme[474,475]. This trapped TOP1 is referred to as a TOP1 cleavage complex (TOP1cc) and can be deconstructed through tyrosyl-DNA phosphodiesterase 1 (TDP1), which is mutated in a rare autosomal neurodegenerative disorder, SCAN1 (spinocerebellar ataxia with axonal neuropathy) marked by TOP1cc accumulation in cells[476]. These TOP1ccs exemplify a class of DPCs that are bordered by an SSB. Enzymes with AP lyase activity (Section 1.3.1), like pol β can become crosslinked in their pursuit of AP site

repair[477]. Similarly, PARP1s can be crosslinked to abasic sites, which might partly explain the efficacy of PARP inhibitors in the clinic, as PARPi showcases a continuum of differing trapping abilities[478]. The final class of DPCs concerns where proteins are tethered adjacent to a double-strand break. TOP2 homodimers induce opposing breaks, forming a covalent bond with both 5′ ends while disentangling dsDNA before release[479]. The re-ligation of both these nicks is prevented by dsDNA intercalating agents like etoposide and doxorubicin, forming TOP2 cleavage complexes (TOP2cc), which can be relieved by TDP2[480,481]. The germ-cell-specific meiotic recombinase, SPO11, is also vulnerable to forming equivalent SPO11 cleavage complexes (SPO11ccs), that are programmatically removed by the MRN complex[482,483]. Though defining only a limited subset of nevertheless critical cells, it was observed that ATM loss in mice accumulates SPO11ccs, ultimately requiring an MRN-independent DPC removal pathway[482]. Attributable to their heterogeneity, the repair of DPCs is also divided into many pathways and genes. The study and understanding of DPC repair draw helpful analogies to ICL repair in how they were discovered and classified.

1.3.6.2 | Replication-coupled DPC repair

It is a reasonable assumption that the barrier of DPCs to DNA replication (and transcription) might mean that some form of coupling might exist between these pathways and their repair. Like many breakthroughs in ICL repair, *Xenopus* egg extracts were able to reconstruct DPC repair in a cell-free system and provided evidence that a conservative proteolytic mechanism was stimulated by DNA replication, without any hint of replisome disassembly nor DSB generation[484]. Concurrently, the first DPC protease, Wss1 (SPRTN in humans) was discovered in yeast, and exhibited proteolysis of ssDNA-bound proteins[485]. Inactivating mutations in *SPRTN* are lethal for cells and hypomorphs in the germline and result in Ruijs-Aalfs syndrome, a progeria associated with genomic instability[486]. DPCs accumulate, culminating in further genome instability, through the disabling of replication forks[487]. The stimulation of DPC proteolysis, by SPRTN and/or the proteasome, is initiated by the stalling of CMG at the trapped protein, which triggers ubiquitylation of the protein adduct by TRAIP[488,489], promoting CMG bypass (assisted by the helicase, RTEL1) while also encouraging the recruitment of the proteasome[461]. TRAIP can also ubiquitylate an opposing CMG when forks converge on an ICL to either recruit the NEIL3 glycosylase

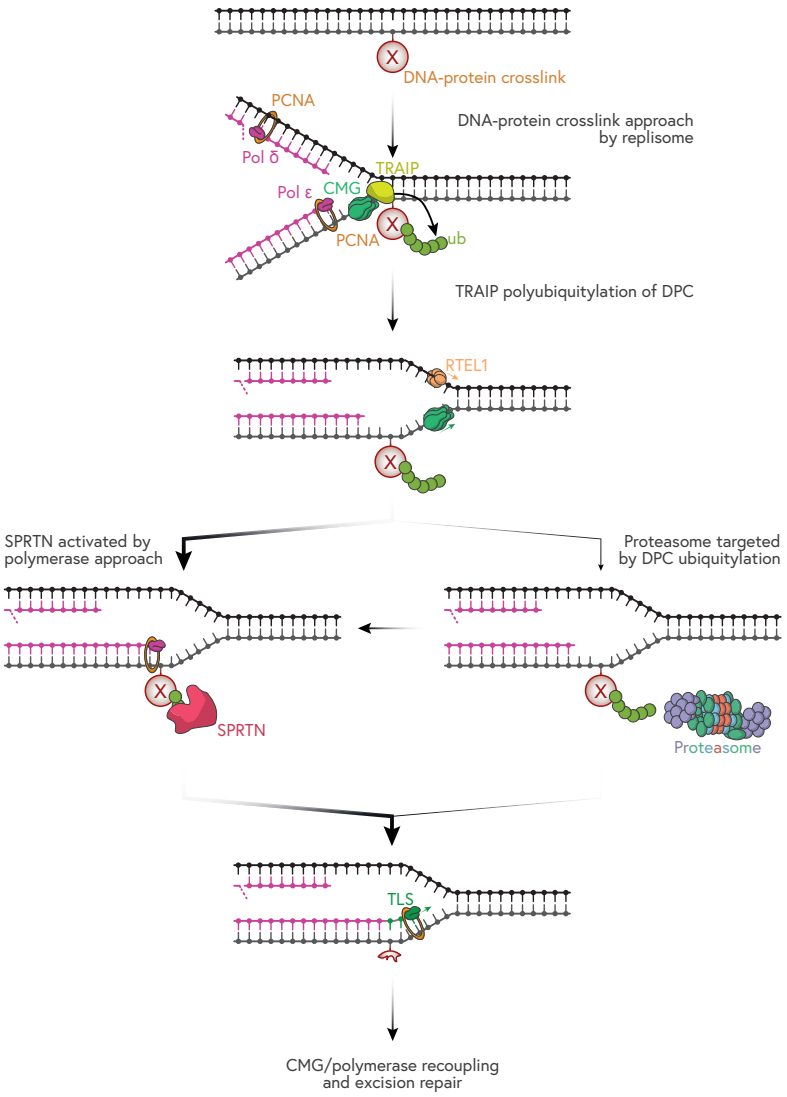


Figure 1.6 | Replication-coupled proteolysis of DNA-protein crosslinks (DPCs). DPCs present a substantial barrier to replication fork progression, and their repair is tightly coupled to replisome engagement. Upon encountering a DPC, the CMG helicase stalls, which triggers polyubiquitylation of the adducted protein by TRAIP. This event is aided by the accessory helicase RTEL1, which promotes fork progression past the lesion. Ubiquitylated DPCs are subsequently recognised by the proteasome or processed by SPRTN, which is activated upon the approach of a DNA polymerase and requires proximity to a DNA junction and interaction with PCNA for its activity. Both pathways function to degrade the protein adduct into a short peptide remnant, enabling TLS polymerases to bypass the lesion and allow CMG-polymerase recoupling and excision repair.

capable of unhooking certain ICLs or to initiate CMG removal *via* p97, which is a critical step in activating the Fanconi anaemia pathway[411] (Figure 1.6).

While ubiquitylation enhances SPRTN-proteolysis of the protein adduct, it is not totally necessary for this action, and SPRTN can degrade non-ubiquitylated DPCs, though at a reduced rate[490]. Where rapid recycling of non-covalent DNA-bound proteins, especially of the replisome, can be tolerated due to the abundance of these components (PCNA, polymerases, RPA), the CMG helicase requires timely licensing and activation and cannot be replaced like-for-like if processed unencumbered by SPRTN[463]. What protects from this event is that SPRTN requires a DNA polymerase to approach within a few nucleotides of the adduct, which also ensures leading and lagging strands are treated equally where DPCs persist[490]. It is an interaction between SPRTN and ubiquitylated PCNA *via* its PIP and UBZ domains that confines SPRTN activity only during the polymerase approach[491]. The prevailing theory is that SPRTN acts as the key protease, minimising the protein footprint of a DPC to a short peptide adduct that is amenable for CMG bypass[490]. The proteasome acts to back up this function by dealing with DPCs that are inaccessible to SPRTN proteolysis, like those with shielded ssDNA, which is a determinant of SPRTN activity[492]. Additional backup proteases include the DDI1/2, which targets polyubiquitylated DPCs[493], and the serine protease FAM111A, which safeguards productive fork progression in a manner dependent on its trypsin-like protease domain and involves interaction with PCNA and DNA[494]. Replication-associated DPC repair involves the combined efforts of enzymes acting either downstream of CMG to promote replication fork progression (TRAIP/RTEL1) or those acting behind to guarantee that gaps are filled in (SPRTN/REV1).

1.3.6.3 | Replication-independent DPC repair

NER deficiency is associated with some mild formaldehyde sensitivity, possibly suggesting NER promotes the removal of protein adducts on dsDNA (though could also hint at the generation of archetypal NER substrates)[495,496]. Within a cell-free system, however, NER is incapable of acting on DPCs of full protein-DNA links but instead can excise DNA with a short peptide attached (less than 10-15 kDa in size)[497,498]. It is plausible to conclude, then, that NER acts on the DNA-peptide scar abandoned by DPC proteolysis *via* SPRTN or the proteasome. Interestingly, only TC-NER components CSA and CSB, and not GG-NER

components XPC, CETN2, XAB2, DDB1 or RAD23A, reflect sensitivity to formaldehyde in various CRISPR-KO screens (Figure 3.1)[446,499,500]. While the understanding of transcription-coupled DPC repair remains in its early phases, analogy to TC-NER is likely relevant given recently reported reliance on CSA and CSB[501–504]. However, in these contexts, NER itself was dispensable[501,503].

In the case of TOP1ccs, it is proposed that trapped TOP1 is targeted for SUMOylation-dependent ubiquitylation *via* K48-linked chains, which subsequently directs the complex for proteolysis by either the proteasome or SPRTN[505]. The previously mentioned TDP1 enzyme then resolves the crosslink, leaving behind a $3'$ phosphate that can be processed by PNKP, thereby allowing single-strand break (SSB) repair machinery to complete the repair[506]. There is also evidence suggesting that XPF-ERCC1 may incise around TOP1ccs to remove the adduct[507]. These mechanisms illustrate how canonical single-stranded DNA (ssDNA) repair pathways, such as NER and BER, can take over once the bulky protein adduct is removed or reduced to a more manageable substrate. For DPC sequestering both DNA strands, as in the case of TOP2cc or SPO11cc, canonical double-strand break (DSB) repair pathways are employed. In TOP2cc repair, TOP2 trapping induces SUMOylation *via* ZATT, promoting the recruitment of TDP2[508], which is capable of resolving crosslinks tethering TOP2 to double-stranded DNA. This action generates readily ligatable DNA ends that are subsequently processed by non-homologous end joining (NHEJ) machinery[482]. Additionally, the MRN complex provides endonuclease activity that is enhanced in the presence of protein adducts at $5'$ DNA ends, enabling cleavage around either TOP2ccs or SPO11ccs[482]. The choice of DSB repair pathway may depend on whether the TOP2/SPO11ccs are engaged by MRN (promoting homologous recombination, HR) or by TDP2 (promoting NHEJ)[463,509]. DPC repair involves specialised mechanisms tailored to the nature of the crosslink and its associated DNA context. These pathways ensure that bulky protein adducts are efficiently resolved, leaving behind either a small protein scar or DNA break that canonical ssDNA or DSB repair pathways can repair to restore genome integrity.

1.4 | The SMX complex

1.4.1 | History of the SLX4 scaffold

Slx4 and its principal protein partner Slx1 were first identified as equally indispensable factors for the viability of budding yeast *Saccharomyces cerevisiae* lacking Sgs1^{BLM}, the yeast homolog of BLM[510]. Following this, in *Schizosaccharomyces pombe*, a stable interaction between Slx1 and Slx4 was shown to be crucial for viability given Rqh1^{FANCM} loss[511]. These epistatic relationships with helicases implicated Slx4-Slx1 in the resolution of joint DNA molecules, formed during replication and homologous recombination[512]. Additional nuclease binding partners for Slx4, Rad1^{XPF}-Rad10^{ERCC1} and Mus81-Mms4^{EME1}, were then discovered in yeast and these interactions were also confirmed for their human orthologues[513–517]. XPF-ERCC1 and MUS81-EME1 share a catalytic domain with common PD-(D/E)XK motifs, found within the nuclease-active proteins XPF and MUS81, respectively, while ERCC1 and EME1 play vital structural roles[518]. In contrast, SLX1's activity is more closely related to the GIY-YIG motif-containing homing endonucleases[519].

Comparing SLX4 domains throughout evolution provides a glimpse into the notable accessory factors, and hence binding sequences, that SLX4 has acquired (Figure 1.7). *S. pombe* expresses the most minimal SLX4 scaffold, hosting only an interaction with Slx1 through its conserved C-terminal domain (CCD), despite also possessing a SAP domain[511]. This minimal unit is only 419 amino acids long, less than a quarter of that in humans and just over half the size of *S. cerevisiae* Slx4 (Figure 1.7). The first stage of evolutionary acquisition is an N-terminal domain that governs interaction with Rad1^{XPF}-Rad10^{ERCC1}. The MUS312-MEI9 interaction-like region (MLR), spanning 500-558 in human SLX4, has not been annotated in *S. cerevisiae* Slx4, but given Slx4's ability to interact with Rad1^{XPF}-Rad10^{ERCC1}, in a manner that is mutually exclusive with Slx1, it is likely that origins of the MLR domain are laid therein[515]. The function of equipping Rad1^{XPF}-Rad10^{ERCC1} to Slx4's arsenal is to aid the repair of DSBs *via* single-strand annealing (SSA) by stimulating the excision of 3′-flap overhangs by Rad1^{XPF}-Rad10^{ERCC1}[520]. The SAP domains of *S. cerevisiae* or *S. pombe* do not directly bind to Mus81, as in metazoan, hinting that this function evolved later, but the SAP domain is required for accurate processing of 5′-flap DNA by Slx1 and is implicated in binding DNA itself[521].

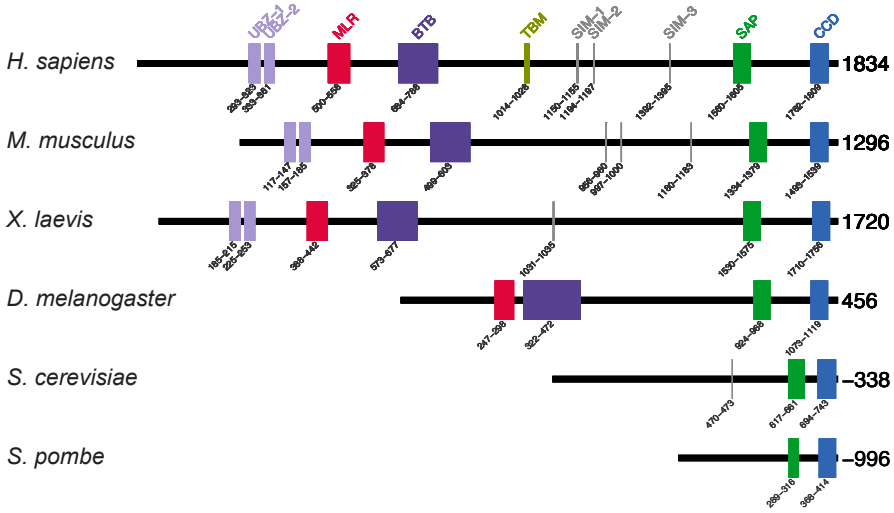


Figure 1.7 | Domain organisation of SLX4 orthologues across eukaryotes. The domain organisation of SLX4 orthologues across eukaryotes, adapted from[522]. All SLX4 orthologues are shown to scale with their lengths given to the right, and aligned residue boundaries for domains are shown. Abbreviations for domains: UBZ4, ubiquitin-binding zinc finger type 4; MLR, MUS312-MEI9 interaction-like region; BTB, Broad-Complex, Tramtrack and Bric a Brac; TBM, TRF2-binding motif; SIM, SUMO-interacting motif; SAP, SAF-A/B, Acinus and PIAS; CCD, conserved C-terminal domain. Organism names and their corresponding UniProt accession codes are as follows: *Homo sapiens* SLX4 = Q8IY92, *Mus musculus* = Q6P1D7, *X. laevis* = A0A8J0TU46, *Drosophila melanogaster* SLX4 = Q9VS48, *S. cerevisiae* = Q12098, *S. pombe* = Q9P6M0.

A further acquisition within the N-terminus of SLX4 was the evolution of an oligomerisation domain named the 'BTB' (Bric-a-brac, Tramtrack and Broad complex) domain, bestowing SLX4 a homodimerisation ability[523]. This dimerisation surface is saturated with hydrophobicity and likely means that all BTB-domain containing SLX4 orthologues are inherently dimeric to bury this hydrophobicity away from the solvent in cells. The BTB domain is necessary for forming SLX4 foci, *e.g.* at telomeres, with dimerisation-inactivating mutations detrimental to telomeric integrity[523]. The BTB domain is also required for efficient ICL repair, likely through a role in the appropriate binding of XPF[523,524]. However, the details of this interaction have not yet been determined and the BTB domain, though not necessarily dimerisation, is dispensable for ICL repair in one *Xenopus* egg extract system[525]. A missense mutation of a highly conserved residue in the BTB domain, G700R, has been associated with breast cancer, though in rare cases[526]. A final tool that evolved into SLX4's repertoire is the ability to bind ubiquitin or SUMO. Though research has only recently tackled the functional impact of this feature, it has been suggested that binding these related but disparate protein modifications funnels SLX4

and its associated binding partners into different repair pathways[527]. The N-terminal UBZ4 domains are critical for ICL repair and contribute to HR intermediate processing[443] (Section 1.4.5.1). The SUMO-interacting motifs (SIMs) forming part of the long linker that splits the ordered domains at the N- and C-termini are also involved in ICL repair but have been more fundamentally associated with a general response to replication stress and telomeric localisation[528–530]. Which SUMOylated or ubiquitylated factors SLX4 binds through these domains have not yet been determined. However, the SIMs also confer an interaction with the SUMO E2 conjugating enzyme UBC9[531]. SLX4 is associated with E3 SUMO ligase activity and drives SUMOylation of XPF and itself *in vivo*, in a manner requiring the SIMs and BTB domain for this activity[531]. Whether this is an intrinsic SLX4 activity or whether it accompanies an E3 ligase requires further investigation. SUMOylation of SLX4 has been shown to be promoted by NIP45, which aids the complex to target DNA catenanes into DSBs leading to the activation of the G2 damage checkpoint[532]. There are no links between *S. cerevisiae* Slx4 and SUMO or ubiquitin, but the *S. pombe* Slx1 protein has been shown to interact with SUMO *via* a conserved SIM, though the function of this interaction is poorly understood[533]. In *S. cerevisiae*, the SUMOylation of Slx4's binding partner, Saw1, forms a loop back to Slx4 *via* Slx1[534]. In HR, SLX4 is foremost involved in recruiting and influencing its associated nucleases to interact with 3′-flaps, or more complex D-loop and Holliday junction structures[535]. Emerging evidence points to yeast Slx4's involvement in end resection at DSBs (Section 1.3.4.2), even independent of any SSE partner[536,537].

SLX4 deficiency is not lethal in mice[538–541] nor in humans[7,524,542], but is incompatible with viability in chicken DT40 cells[442]. As with the depletion of essential HR genes, including RAD51, SLX4-deficient cells accumulate in G2 and show gross chromosomal instability, which is amplified by IR[442]. The essentiality of SLX4 in this context is not due to improper formation of an SLX4-MUS81 complex, since DT40 cells lack MUS81, but could be due to an association with XPF, also an essential gene in DT40 cells[543]. Full SLX4 knockout by CRISPR-Cas9 has not been achieved in many cancer cell lines despite efforts, and targeting the deletion of exons 2-15 of *Slx4* in mES cells resulted in only $Slx4^{-/+}$ clones despite an expected homozygous knockout rate of 5%, further suggesting the essentiality of SLX4[544]. SLX4 also trends towards essential in most cell lines within the CRISPR DepMap database, with a median gene effect of -0.3.

1.4.2 | SMX in SSA

In SSA, Slx4 is only recruited after Rad52-dependent displacement of a 3′-flap[520,545] but is not critical for *S. cerevisiae* Rad1^{XPF}-Rad10^{ERCC1} recruitment[520]. Instead, Saw1 fulfils this function by stimulating Rad1^{XPF}-Rad10^{ERCC1} activity on 3′-flap substrates[354,520], and the link between Slx4-Slx1 and Saw1 might, therefore, indicate that Slx4 indirectly recruits Rad1^{XPF}-Rad10^{ERCC1} in this manner. It is feasible that evolution streamlined this approach, bringing together the separate functions of *S. cerevisiae* Slx4 and Saw1 to a larger SLX4 unit capable of recruiting XPF-ERCC1 while also constitutively interacting with SLX1. The progression of SSA is licensed by Slx4 phosphorylation at T113 by Mec1^{ATR} or Tel1^{ATM}, and deficiency of these factors or expression of Slx4 phosphorylation-defective mutants impairs SSA[545]. Msh2-Msh3 is a well-established factor in SSA[546], stabilising annealed DNA intermediates during SSA, and interactions between human SLX4 and MSH2-MSH3 point further towards SLX4's role in SSA[323,547–549]. However, this interaction, now mapped to an N-terminal SHIP box of SLX4 (residues 210-220), is necessary for efficient MMR, though is dispensable for ICL repair[550]. This complex, MutSβ, also appears to stimulate HJ resolution by the SMX complex through its interaction with SLX4[549,551].

1.4.3 | SMX in HR

The consequence of lacking the BTR (BLM-TOP3A-RMI1-RMI2) complex that typically 'dissolves' the HR intermediary Holliday junction structure is to rely more heavily on nuclease-based 'resolvases' (Section 1.3.4.2). BTR loss is causative of Bloom syndrome, and patient-derived cells cannot then also lose resolvases[552]. HJ resolution in somatic cells is reliant on SLX4 and SLX1/MUS81, or GEN1, with GEN1 generating symmetric nicks on opposing strands like that for bacterial resolvases[553]. The SLX4 machinery appears to undertake a more complex procedure for dHJ resolution. An SLX1 nick on one strand is followed by an opposing second cut by MUS81[323,324,540,554]. Different partners of SLX4 are involved in different species for this precise role: the MUS312 (*D. melanogaster* orthologue of SLX4) interacts with the XPF orthologue MEI9 in the generation of meiotic crossovers[555], which is not reliant on MUS81[556]. In *Caenorhabditis*

elegans, SLX4's orthologue HIM-18 alongside any of its SSEs, XPF-1, SLX1-1, or MUS81-1, carries out the processing of recombination intermediates[557–559]. Despite this involvement in resolving meiotic crossovers, an SLX1 anti-crossover role has been described at centromeres[560]. In *S. cerevisiae*, Mlh1-Mlh3 endonucleases primarily process meiotic recombination intermediates, but some part of this task is carried out by Slx1-Slx4[561–564]. The influence of SMX in epistatic HJ processes was identified by specific observed phenotypes;[321,540,554,558,559,565] depleting SLX1 or MUS81 individually in Bloom syndrome cells reduces cell viability to a lesser extent than co-depletion, or sole depletion of SLX4[554] and in BLM-depleted cells that are deficient for SLX4 (FA-P), expression of SLX4$^{\Delta SAP}$ or SLX4$^{\Delta CCD}$ partially complements the reduction in SCEs[565], which together hint at non-epistasis in HJ resolution. The specific chromosome abnormality phenotype observed in FA-P cells lacking BLM or GEN1 are due to improperly processed HJs since it can be rescued by expressing bacterial RusA HJ resolvase, just as expression of SLX4$^{\Delta SAP}$ or SLX4$^{\Delta CCD}$ rescues[565]. The epistasis implied here does not disqualify these enzymes from acting on different substrates; however, genetic studies were followed up on by insect cell-purified SMX complexes. In this work, SMX was purified separately as SLX1-SLX4, MUS81-EME1, and XPF-ERCC1 and combined to evaluate their activities on model DNA structures. XPF was concluded to provide a more favourable SMX architecture for catalysis of HJ, but be non-catalytic otherwise[324]. However, in a purified biochemical system, certain *in vivo* contexts may be insufficiently modelled. Indeed, the SLX4-XPF interaction is not required for the survival of BLM-deficient cells nor contributes to the high SCE rate of certain SLX4 deficiencies[565]. It is the responsibility of SLX1 and MUS81 to enact the 'nick, counternick' mechanism for HJ resolution[324]. That only higher eukaryotes employ this nick-counternick mechanism, wherein MUS81 relies on SLX1 to perform an initial incision, is a matter of concern. *S. pombe* only uses Mus81-Eme1 for HJ resolution in the absence of Yen1[566,567], while Mus81-Mms4^{EME1} and Yen1 contribute to HJ resolution in *S. cerevisiae*[568–570]. Besides the work involving recombinantly-purified SLX1 mixed with MUS81-EME1 outside of cells[324], no involvement of lower eukaryotic SLX1 orthologues in HJ resolution has been described. There is loose evidence that SLX1 can act as a resolvase by itself, perhaps acting in place of MUS81-EME1, if required. The bacterially-produced SLX1-SLX4^{CCD} complex can resolve HJs *in vitro*, and in a

manner that generates symmetric nicks capable of rapid religatation[323,513]. The lack of full-length Slx4 in these preparations is a cause for concern, however, and limits certain interpretations of both full-length or human SLX4 activities, where other protein-protein interactions might influence the precise activities of SLX1. Insect cell-purified full-length SLX1-SLX4 is a promiscuous HJ nuclease, sometimes even cleaving a whole arm of an HJ[554], but whether the N-terminal extension and even the MLR present a misfolded and structurally distinct SLX4 that prevents its true folding around an HJ is plausible. The ultimate goal would be to express SMX as a whole entity fully, reconstituting the activities of XPF, MUS81, and SLX1 in near-native folds around SLX4 (Section 1.7). The study of mixed insect cell-purified SMX is a step towards this ideal reconstitution, but the lack of co-folding might transform the complex away from what exists *in vivo*[324]. SMX is likely not a fully assembled complex throughout the cell cycle. By visualising proximity-ligated foci between XPF and MUS81, progressive increases in the assembly of SMX have been inferred for cells transitioning from G2 to early mitosis, which decreases again from anaphase[571]. This limits complete SMX assembly to a short window, up to 45 min in length, despite constant subunit expression throughout the cell cycle and therefore points away from proteasomal control of SMX, towards one of cell-cycle-dependent post-translational modifications[571]. In yeast, the association of Mus81-Mms4^{EME1} is strictly cell-cycle regulated, peaking in G2/M phases that require efficient HJ resolution. Mms4^{EME1} is phosphorylated by both Cdc28$^{CDK1/2}$ and Cdc5^{PLK1}, resulting in stimulation of its HJ-resolving activity[570,572–574]. *S. pombe* Eme1 is instead phosphorylated by Cdc2^{CDK1} and Rad3^{ATR}, likely to prime the nuclease in a DNA-damage dependent fashion to achieve full activation[575]. Therefore, the restriction of MUS81 activity is achieved by overcoming cell cycle and DNA damage criteria, and this regulatory mechanism is critical for the survival of Rqh1^{BLM}-deficient cells[575]. What has been corroborated for human cells is that EME1 phosphorylation by CDK1 and PLK1 increases its resolvase activity on Holliday junctions from MUS81 immunoprecipitates[570], and promotes its interaction with SLX4[554,565]. Whether EME1 phosphorylation contributes to increased HJ resolution in human cells or how this would play out mechanistically is still unclear. However, MUS81 activity is contingent on CDK1-mediated phosphorylation of SLX4's SAP domain that drives its association with MUS81[576]. A recent report highlights the 10-fold increase

in affinity of the MUS81-EME1 interacting region (MIR) of SLX4 towards MUS81 by CDK1-dependent phosphorylation of residues within this SAP domain, T1544, T1561, and T1571[577,578]. The study also showed that phosphomimetic mutations increased the affinity of an SLX4-MIR peptide towards MUS81-EME1[577,578]. Reciprocal phosphorylation of MUS81 occurs, restricted to mitosis and suppressed in S phase, at S87 by CK2, which enhances its interaction with SLX4 and confers the activation of MUS81[577,579]. Despite this, insect cell-purified MUS81 and SLX4 interact when co-expressed, and the interaction does appear in Y2H experiments, though extra *in vivo* partners of SLX4 might serve to deplete this MUS81 interaction.

1.4.4 | SMX in ICL repair

As described earlier (Section 1.3.5.2), SLX4 and XPF are critical components of the FA pathway. Rad1^{XPF}-Rad10^{ERCC1} facilitates replication-coupled ICL repair *via* NER in G1 and replication-coupled, NER-independent mechanisms in S/G2 where it relies on interactions with Rpa1, Saw1, and Slx4[580]. After identifying SLX4's tandem UBZ4 domains[513], its function was easily presumed to be a pivotal link between upstream monoubiquitylation of FANCD2/FANCI[581] and downstream XPF-mediated SLX4-aided FA progression[435]. Partial deletion of UBZ4-1 and entire deletion of UBZ4-2 in patients were associated with ICL hypersensitivity and chromosomal aberrations[7,582]. Deleting the SLX4 UBZ4 domain in chicken DT40 cells prevents recruitment to ICL-induced damage foci and hypersensitises to crosslinking agents[442]. Deleting the tandem UBZ4 domain removes co-immunoprecipitation of monoubiquitylated FANCD2[442], and the recruitment of SLX4-XPF-ERCC1 is dependent on FANCD2 monoubiquitylation in *Xenopus* egg extracts during ICL repair[435]. A direct interaction, however, has been contested. SLX4's UBZ4 domains preferentially bind to K63-linked polyubiquitin chains over single ubiquitin molecules like those attached to FANCD2 during ICL repair[7,443]. Likewise, SLX4 recruitment to laser-induced ICL damage is not prevented in FANCD2-deficient cells[443]. There also exists a non-epistatic connection between deletion of the SLX4^{UBZ} and FANCC in DT40 cells, and laser-striped SLX4 recruitment to ICLs is independent of ubiquitin E3 ligases RNF8, RAD18, BRCA1[442]. In most of these works, however, the use of hypomorphic FANCD2 mutants means that some residual FANCD2

monoubiquitylation might permit SLX4 recruitment. Whether SLX4 directly binds a separate ubiquitylated entity (or entities) during ICL repair warrants further study, but SLX4-dispensable recruitment of XPF to ICLs by UHRF1 has been described[440]. Recently, a siRNA-screening approach with a GFP-tagged N-terminal SLX4 fragment (SLX4-N, 1-900) identified RNF168 as essential for MMC-induced SLX4 foci formation, which colocalise with RNF168 foci[583]. Laser-induced ICL recruitment of SLX4 depends on RNF168, and SLX4 and RNF168 show epistasis towards MMC treatment[583]. Overall, this study suggests the identity governing SLX4 recruitment to ICLs might be histone H2A, which is known to be ubiquitylated by RNF168 at K13/K15[583,584].

Despite no concrete attribution of how SLX4 is recruited to ICLs, its *bona fide* interaction with XPF-ERCC1 is undisputedly essential[400,585]. The MLR domain of SLX4 is essential in evading sensitivity to crosslinking agents and mediates the XPF interaction[524,538], with point mutations later identified that eliminated XPF-SLX4 interaction spanning 500 to 558 of SLX4[531,586] - FLW[531] and FY[546] are necessary for binding to XPF. XPF depletion in SLX4-deficient FA cells does not further perturb sensitivity to MMC[524], and complementation of *Slx4*$^{-/-}$ MEFs with a point mutant unable to interact with XPF worsens the chromosomal instability arising from MMC treatment[586]. Redundant routes for XPF recruitment have been described, such as *via* the UHRF1 scaffold, in a manner independent of SLX4[440], thus explaining how SLX4 absence is preferable to the expression of an SLX4 variant that fails to recruit XPF productively. SLX4 therefore recruits XPF-ERCC1 to either single replication forks or convergent forks, stalling at an ICL[435,587] and also enhances unhooking of an ICL by XPF-ERCC1-mediated incision(s)[435,541]. The precise position of this incision has been reported on the leading strand template 3′ to the ICL[541,588] but also 5′ to the ICL[444,589]. This 5′ cut is stimulated by RPA and suggested to introduce an entry point for the 5′-3′ exonuclease, SNM1A, which is capable of digestion through an ICL[444,445]. Supposing an exonuclease carries out the complete unhooking step, it is also possible that the SNM1A-related SNM1B/Apollo exonuclease can fulfil this function, as it has been shown to digest 5′ to 3′ through an ICL[590], which corroborates an immunoprecipitation-based interaction observed between SLX4 and Apollo[591]. Dual incision, which would be sufficient for total unhooking, has, however, been suggested for XPF-ERCC1[588,589], in a manner heavily influenced by

SLX4[541]. SLX4 depletion in *Xenopus* egg extracts inhibits unhooking incisions and overall replication-coupled ICL repair, with a transient interaction between the BTB domain of SLX4 and XPF required to place XPF-ERCC1 appropriately relative to the ICL[435,587]. In NER, the role of the second incision falls upon a second nuclease, XPG, and it is yet unclear whether *in vivo*, unhooking requires a second cut by *e.g.* SLX1 or FAN1, though SLX1 is not required for ICL repair in *Xenopus* egg extracts[525].

Mus81$^{-/-}$ and *Eme1*$^{-/-}$ mES cells are susceptible to crosslinking agents[592–595], but this level of sensitivity is only a fraction of that of *Ercc1*$^{-/-}$ cells[594], and MUS81's protective role against crosslinking agents is not dependent on an interaction with SLX4[540,596]. SLX4's SAP domain is not critical for resistance to ICLs, and MUS81 does not play a part in the replication-coupled ICL repair in *Xenopus* egg extracts[435]. However, in HCT116 *MUS81*$^{-/-}$ cells, an SLX4-binding mutant of MUS81 could not rescue MMC hypersensitivity[596]. It is possible to infer that the role of MUS81 in these genetic susceptibilities towards crosslinking agents is a consequence of a reduced proficiency in HJ resolution, which forms part of the eventual resolution of ICLs *via* HR. Nevertheless, DSBs caused by unhooking were found to be dependent on MUS81 when XPF-ERCC1 and SLX4 were depleted, perhaps suggesting this 3′-flap endonuclease can compensate for its related XPF counterpart[445]. SLX1 also appears necessary for ICL repair in MEFs in a manner not dependent on any HJ cleavage activity, given that SLX1 is not required for increased SCE formation by ICLs[540]. Still, human cells require MUS81-EME1 and SLX1, but not XPF, for elevated SCEs induced by cisplatin treatment[554] or MMC[565]. It is noteworthy that the majority of crosslinking agents used in these studies are promiscuous, generating mono-adducts at least as frequently as they might crosslink[597]. Therefore, the observed phenotypes might instead reflect MUS81-EME1 and SLX1 acting on forks that have been stalled by ssDNA adducts.

1.4.5 | The supernumerary functions of SLX4

1.4.5.1 | Replication stress response

Camptothecin-induced trapping of TOP1 generates DSBs that replication forks encounter and are stalled at[598]. SLX4 is intrinsically linked in this process[323,517], dependent on its association with MUS81, and to a lesser extent, SLX1[524]. Such dependencies

suggest MUS81 cleavage of toxic replication fork intermediates occurs following TOP1cc collisions[599,600], while the partial reliance on SLX1 possibly reflects the requirement of HJ processing after HR restarts replication forks. Abraxas, as part of the BRCA1-A complex, limits the build-up of single-ended DSBs arising from SLX4/MUS81 activity at TOP1cc-stalled replication forks and does so by counteracting K63-linked polyubiquitylation[601]. More recently, the Tus-Ter replication fork barrier developed by Scully and colleagues was applied to understand how SLX4 contributes to their resolution[544]. It is SLX4's interaction with XPF in this particular context, and not MUS81-EME1 nor SLX1, that is required for HR at DNA-protein stalled replication forks and SLX4-XPF cleavage also launches the induction of long-tract gene conversion by BIR from a DSB[544]. The requirement of SLX4-XPF was also needed for further downstream DDR factor recruitment at forks stalled by LacI binding to lacO arrays[602]. The SIMs of SLX4 play a part in permitting cleavage of CPT-induced replication intermediates[529], and these motifs are suggested to explain the observed presence of SLX4 at replisomes by iPOND (isolation of proteins on nascent DNA) experiments[603]. Mutating SIMs also diminishes HU-induced DSB formation[531] and prevents RNF4-dependent condensation of SLX4, which can also be observed for the SLX4^{F708R} dimerisation-mutant[604]. ATR inhibits the RAD18 monoubiquitylation of PCNA through phosphorylation near a PIP box, which prevents SLX4 recruitment to ub-PCNA and limits excessive stressed fork processing by SLX4-associated nucleases[605]. MUS81-mediated DSBs are critical for replication fork restart[83,594,606,607] and therefore SLX4 contributes to DSBs after HU treatment[531,608]. While it is often difficult to attribute which exact complex of MUS81 (whether alongside EME1 or EME2) enacts particular functions, it is clear that MUS81 is incapable of catalysis without one of the two partners. Nevertheless, the specific MUS81-EME2 complex appears more relevant at HU-stalled fork processing[83,607]. A BRCA1/SLX4-MUS81 complex has been attributed to the formation of DSBs in HCT116 cells treated with HU[609], which might be bridged by a BARD1-MUS81 interaction axis involving a phosphorylated BARD1 'MUSIC' motif, which has been connected to the resolution of joint molecules as a backup to BLM[610]. PLK1 kinase activity has been shown to regulate SLX4's interaction with MUS81[554,576,609]. The nature of SLX4-dependent (/MUS81-dependent) DSBs at replication forks is not always to the benefit of a cell; SLX4 or MUS81 depletion provides resistance to HU[531,611]. HU treatment

in combination with an ATR inhibitor results in further replication stress, DNA breaks and longer ssDNA tracts due to the derepression of origin firing[78]. In this study, breaks are created depending on SLX4, yet somehow not MUS81 nor XPF/SLX1. However, this may reflect redundancy between the three SLX4-scaffolded nucleases[612]. In *Candida albicans*, Slx4 deletion did not sensitise cells to HU but did to MMS[613], conflicting with the reports in human cells. In ATR-deficient cells, aphidicolin-induced DSBs are partly dependent on SLX4[614], downstream of RNF4-dependent E3 ubiquitin ligase activity and activation of the AURORA-PLK1 pathway, which might serve to enhance the SLX4-MUS81 interaction or help remodel forks such that they are more appropriate substrates for SLX4 nucleases. To further this hypothesis, FANCM depletion reduces the enrichment of SLX4-MUS81 observed in RAD51C-deficient cells at stalled forks after HU treatment[615]. Contrastingly, both SLX4, MUS81 and their mutual interaction are required for the DSBs that arise from HU-treated cells following CHK1 inhibition[608]. Yet, another study prescribes the response to replication stress to SLX1-SLX4 and XPF, but not MUS81[616]. Many interpretations surrounding SLX4-scaffolded nucleases and their impact on the replication stress response are currently conflicting, though they are likely to reflect compositional heterogeneity and redundancy between nucleases.

The action of several helicases such as SMARCAL1[82,617], FANCM[94], ZRANB3[89], and FBH1[611] can enact the reversal of stalled forks generating a more suitable substrate for MUS81 cleavage[324]. Indeed, MUS81 can incise reversed forks following Cyclin E/CDC25A overexpression-induced replicative stress[618] or in HU-treated BRCA2-deficient cells[83]. An established prerequisite for MUS81-dependent cleavage of reversed replication forks is the first exonucleolytic activity by MRE11 to trim annealed nascent strands into a ssDNA tail[83,619]. MRE11-dependent degradation of reversed forks contributes to MUS81-dependent cleavage in cells absent of fork-protective BRCA1/BRCA2, which loads RAD51 on ssDNA tails within reversed forks to protect them from nuclease digestion[88]. The ability of SLX4-MUS81 to produce DSBs means stalled replication forks sit on a knife's edge regarding cell viability: cleavage either aids fork revival where sufficient DSB repair machinery is available or forces their collapse and degradation otherwise.

SLX4 and MUS81 promote the survival of BRCA2-deficient cancer cells by initiating MiDAS by MUS81 DSB generation[620]. An alternative EZH2 (the catalytic component of

the polycomb repressive complex 2; PRC2)-MUS81 interaction promotes the enrichment of MUS81 at chromatin but impairs the viability of BRCA2-deficient cells[621]. WEE1 inhibition, through premature entry into mitosis and excessive origin firing, can manifest replication stress, which leads to DSBs and DNA damage in a manner dependent on SLX4 and MUS81-EME1 as well as MUS81-EME2[576,622]. The DNA damage in this context is a total pulverisation of chromosomes, referred to as chromothripsis, which, along with a phenotype of premature entry into mitosis, is prevented by depletion of SLX4, MUS81 or EME2[576]. WEE1 inhibition of CDK1 and PLK1 activity might prevent the association of MUS81 with SLX4, limiting the contribution of MUS81 to chromothripsis in this scenario[576].

1.4.5.2 | Dampening of cell cycle checkpoint signalling by SLX4

An interplay between yeast Slx4 and two BRCT-containing scaffold proteins, Rtt107^{PTIP} and Dpb11^{TopBP1} has revealed pivotal roles for Slx4 in the replication stress response, governed by phosphorylation and various protein-protein interactions[527]. These interactions enable Slx4 to restrain Rad53^{CHK2} activation while simultaneously promoting Mec1^{ATR} signalling[623,624]. In response to methyl methanesulfonate (MMS)-induced DNA damage, Slx4 forms a ternary complex with Rtt107^{PTIP} and Dpb11^{TopBP1}, stabilised by the recognition of H2A phosphorylation at S129 (equivalent to γH2AX) and Ddc1^{RAD9} phosphorylation at T602[625]. Ddc1^{RAD9}, Rad17^{RAD1} and Mec3^{HUS1} combine to form the PCNA-like sliding clamp that bears homology to the human 9-1-1 complex composed of RAD9-RAD1-HUS1[626]. This Rtt107^{PTIP}–Slx4–Dpb11^{TopBP1} complex dampens Rad53^{CHK2} activation by competing with Rad9^{53BP1} for binding to Dpb11^{TopBP1}, thereby limiting Rad53^{CHK2} hyperphosphorylation and mitigating its toxic effects on cell cycle progression and DNA repair[627,628]. Notably, Slx4 contributes to long-range DNA end resection (>10 kb) by displacing Rad9^{53BP1}, a molecular barrier to resection, and alleviates Rad53^{CHK2}-mediated inhibition of DNA repair factors such as Exo1[536,537]. An Rtt107^{PTIP}-Mms22^{MMS2L} axis appears to act redundantly with in the displacement of Rad9^{53BP1} from chromatin for the dampening of checkpoint activation[629].

In addition to its regulatory roles in Rad53^{CHK2} signalling, Slx4 facilitates Mec1^{ATR} activation behind stressed replication forks. Through the Rtt107^{PTIP}–Slx4–Dpb11^{TopBP1}–Ddc1^{RAD9}

complex, Slx4 enhances Mec1^{ATR}-dependent phosphorylation of its own components and downstream targets, promoting checkpoint signalling while allowing timely activation of the Mus81–Mms4^{EME1} endonuclease for joint molecule (JM) resolution[625,628]. These regulatory interactions are mediated by phosphorylation of Slx4 and Mms4^{EME1} by Cdc28$^{CDK1/2}$ and Cdc5^{PLK1}, which stimulate Mus81–Mms4^{EME1} enzymatic activity in G2/M[573,630]. Importantly, Slx4's role in JM resolution and SSA repair is independent of its Rad9^{53BP1} competition function, relying instead on its coordination of endonucleases such as Rad1^{XPF}-Rad10^{ERCC1}[515,545]. Together, these findings underscore Slx4's dual role in checkpoint regulation and DNA repair, balancing replication stress responses to maintain genomic stability while facilitating DNA repair and replication restart. Insights from yeast studies provide a framework for understanding similar mechanisms in human SLX4-TopBP1 interactions, which is also regulated by phosphorylation of SLX4 at T1260 by CDK1[628], though so far the only attribution of this link reflects SLX4-MUS81's role in MiDAS[631].

1.4.5.3 | Mitotic DNA synthesis (MiDAS)

Even with low-level aphidicolin leading to minimal replication stress, under-replicated CFSs (Section 1.2.2) persist into mitosis, requiring MUS81-EME1 for their breakage and subsequent repair. SLX4 colocalises to mitotic foci, and its depletion results in anaphase bridges, micronuclei and 53BP1 bodies in the following G1, hallmarks of CFS instability[529,531,632]. SLX4 forms foci in mitotic cells that colocalise with FANCD2 foci and recruits XPF-ERCC1 and MUS81-EME1[531,632]. Cleavage by SLX4 nucleases is not sufficient for genome stability, despite permitting chromatid separation at mitosis, since under-replication precedes instability in subsequent cell cycles. However, the breaks generated by SLX4-MUS81-EME1 prime a mitotic DNA synthesis process termed MiDAS, which is required for CFS maintenance[632]. Mechanistically, MiDAS resembles a break-induced replication (BIR) activity, requiring RAD52 and POLD3[632,633]. MiDAS or BIR-like conservative replication also applies at some telomeres and is referred to as alternative lengthening of telomeres (ALT), which maintains the repeat length of telomeres in a subset of cancer cells that do not reactivate telomerase[634,635]. SLX4's recruitment to CFSs might involve TopBP1, which also promotes MiDAS alongside generally maintaining

damaged chromatin through mitosis until more favourable repair environments in the next G1[636,637]. In MSI cancers, repetitive sequences like TA repeats frequently stall replication forks such that these sequences can restructure into cruciforms, vulnerable to cleavage by SLX1-SLX4[638] or SLX4-MUS81-EME1[181]. Despite the overlapping characteristics of CFSs with rare, fragile sites (RFSs), the specific induction of RFSs at the *FRAXA* locus by folate deprivation is resolved by a form of MiDAS dependent on POLD3, but in this case, also RAD51 and SLX1[639]. This dependency was not observed for MUS81 nor XPF but did require SLX4[639] and may reflect the at-baseline requirement of SLX1 in cells.

1.4.5.4 | SLX4 at telomeres

Telomeres, repetitive DNA sequences at chromosome ends, assemble the shelterin complex to protect chromosomes from degradation, illegitimate repair, and checkpoint activation[640]. Telomere shortening during replication leads to senescence, although cancer cells evade this by reactivating telomerase or using the ALT pathway, a recombination-based mechanism[641]. SLX4's role at telomeres emerged from its identification as a binding partner of TRF2, a key shelterin component[323,513,517]. TRF2 facilitates T-loop formation, shielding telomeres from ATM activation and end-processing enzymes, but must also regulate T-loop dismantlement during replication. SLX4, with SLX1 and MUS81-EME1, contributes to this disassembly, balancing telomere protection with replication[642,643]. SLX4 acts as a negative regulator of telomere length by mediating T-loop processing to generate T-circles and promoting telomeric homologous recombination (HR) events, such as sister chromatid exchanges (T-SCEs)[644,645]. In ALT cells, SLX4 is recruited to telomeres *via* TRF2, with its localisation correlating with telomere length and peaking in late S-phase[645–647]. TRF2, using its TRFH dimerisation domain, recruits SLX4 *via* a short linear TRF2-binding motif (TBM), which is crucial for SLX4's telomeric functions, including recruiting its associated SSEs like SLX1, XPF-ERCC1, and MUS81–EME1[523,644]. Interestingly, SLX1 is essential for preventing telomere elongation and replication-induced fragility, particularly in ALT cells[643,645]. Mouse SLX4 lacks this TBM, but somatic cells in mice do not repress telomerase[648] and may, therefore, be less reliant on TRF2-dependent telomere maintenance pathways or avoid the recruitment of nucleases to extending and thus deprotected telomeres.

SLX4 prevents telomeric fragility and dysfunction-induced foci (TIFs) in both ALT- and telomerase-positive cells. SLX4 deficiency leads to telomere fragility, particularly in cells

with long telomeres or replication stress-prone regions like CFSs as discussed earlier[643,646]. In ALT-positive cells, SLX4 depletion reduces T-SCEs and increases telomere dysfunction-induced foci formation, effects not rescued by mutants unable to bind TRF2 or SLX1[645]. TRF2 further protects telomeres by inhibiting PARP1 and SLX4 recruitment, preventing excessive telomere attrition and T-circle excision[649–651]. In mice, where SLX4 lacks the TBM, telomere protection involves alternative mechanisms, such as PARP1-mediated recruitment of SLX4 under dysfunctional conditions[652]. Functional interactions between SLX4, its SSEs, and helicases like RTEL1 and BLM are critical for telomere stability. RTEL1, recruited by TRF2 in S-phase, prevents T-loop excision by SLX4-associated nucleases and maintains telomere replication fidelity[653,654]. POLD3-dependent, BIR-like ALT to extend telomeres is promoted by BLM dissolution of recombination intermediates but opposed by SLX4-dependent T-loop processing, leading to limiting of ALT markers like APBs and C-circles[645,655]. In SMARCAL1-deficient cells, SLX4 contributes to telomeric damage and C-circle formation, highlighting distinct mechanisms of SLX4 action in telomerase-positive*versus* ALT contexts[656,657]. An ATR cascade inhibiting a RAD18-dependent PCNA-SLX4 interaction limits SLX4 accumulation at telomeres and preserves ALT telomere stability[605]. However, independent of either CSB or SLX4, XPF-ERCC1 is recruited to telomeres, for the instigation of break-induced lengthening of telomeres, dependent on the formation of RNA:DNA hybrids at these ALT telomeres called TERRA[658]. An SLX4-dependent SLX1 break has also been found to initiate BIR at telomeres, leading to fragile telomeres[659]. Therefore, different instances demand the action of each of the SSEs scaffolded by SLX4. Nevertheless, SLX4 plays a dual role in telomere regulation, acting as a mediator of T-loop dismantlement and a suppressor of excessive telomere processing. These functions underscore its importance in telomere length homeostasis, replication stress responses, and genome stability in ALT and telomerase-positive cells.

1.4.5.5 | Additional roles of SLX4 and its nucleases

An emerging aspect of XPF function does not require scaffolding by SLX4. R-loops can present a distinctive DNA damage scenario, but are perilous when the RNA:DNA helicase syntaxin (SETX) or BRCA1 are depleted, resulting in the release of DNA hybrids into the cytoplasm. These cytosol-expelled dsDNA triggers cGAS-immunogenic

signalling and eventual cell death[660] and cleavage by XPF and XPG are determinants of this pathway[660]. Conflictingly, however, an association between RTEL1 and SLX4 co-localises to transcribing RNA pol II, promoting FANCD2 co-localisation, in a yet-unclear mechanism that promotes the resolution of replication-transcription conflicts[661], and also MiDAS at G-quadruplex-associated R-loops[116]. Beyond the DNA damage response, SLX4 has surprising annotations within the literature relating to immunity and retrotransposition. In the context of HIV infection, SLX4 is hijacked by the HIV-1 accessory protein Vpr to induce G2/M cell cycle arrest and escape the innate immune response[662]. Vpr interacts with MUS81-EME1 and XPF-ERCC1 within the SLX4 complex, facilitating premature activation of the SLX4 complex, potentially targeting replication intermediates[662]. SLX4 also binds viral DNA and promotes its nucleolytic degradation to restrain innate immune sensing, with SLX4 deficiency linked to enhanced cytosolic accumulation of nucleic acids sensed by the cGAS-STING pathway[663]. Furthermore, SLX4 and MUS81 were shown to limit LINE-1 retrotransposition and inflammatory signalling, suggesting their role in maintaining genomic stability and suppressing chronic inflammation[663]. Recombinant SLX4-XPF-ERCC1 cleaves intermediates of retrotransposition, and loss of XPF-ERCC1 enhances LINE-1 retrotransposition[664]. Protection against retrotransposition during early development, where transcriptional silencing is weakened, might well contribute to toxic phenotypes of Fanconi anaemia in development.

1.4.6 | SLX4IP

SLX4IP is present only in vertebrates and was initially identified through mass spectrometry as co-immunoprecipitating with SLX1 and SLX4, though its structure and precise roles remain incompletely understood[323]. SLX4IP has a well-ordered N-terminus, composed of α-helices and anti-parallel β-sheets, and a highly disordered C-terminus, which, akin to SLX4, contains three hypothesised SIMs, although these motifs are critical for its interaction with SLX4[4,665–667]. It directly interacts with SLX4, primarily through SLX4's MLR domain, which also binds XPF, raising the possibility of mutually exclusive or cooperative interactions between SLX4IP and XPF within the SMX complexes[547,668]. This interaction may stabilise the macromolecular complex, as SLX4IP's stability decreases in the absence of SLX4 or XPF[547]. SLX4IP plays a critical role in the repair of ICLs

in non-ALT cells. Its disruption in HEK293 cells leads to hypersensitivity to MMC, but not to UV or IR or camptothecin, indicating a specific role in ICL repair[547]. Given its interactions with SLX4 and XPF, SLX4IP may stabilise the SLX1-SLX4-XPF-ERCC1 complex or assist in substrate positioning within the XPF-ERCC1 active site. This hypothesis aligns with observations that SLX4IP disruption modestly reduces crosslink unhooking efficiency *in vivo*[547]. SLX4IP may also modulate nuclease activities of the SMX complexes on downstream recombination intermediates, such as HJs, which warrants further investigation.

In ALT-positive cells, SLX4IP emerges as a regulator of the telomeric proteome. It correlates with SUMOylation of telomeric proteins like RAP1 and XPF and facilitates interactions between SLX4 and PIAS1, an E3 SUMO ligase[669]. SLX4IP-mediated SUMOylation of RAP1 modulates NF-κB signalling by promoting RAP1's interaction with the IKK complex, leading to nuclear translocation of NF-κ and activation of oncogenic pathways[669,670]. This regulation also influences Notch signalling and represses genes such as *TERT*, *ATRX*, and *DAXX*, hallmarks of ALT[669]. Interestingly, SLX4IP balances endonuclease and helicase activity at recombining telomeres, mediating the resolution *via* SLX4 and dissolution *via* BLM[668]. SLX4IP depletion exacerbates ALT phenotypes, increasing markers such as ALT-associated promyelocytic leukaemia bodies (APBs), telomere length heterogeneity, and telomeric DNA damage while reducing productive telomere maintenance[667,668]. Its roles in stabilising SLX4-containing complexes, modulating SUMOylation, and mediating interactions with helicases and endonucleases highlight its importance in genomic stability and telomere dynamics. Further studies are needed to elucidate its structural features, substrate specificity, and impact on DNA repair and recombination intermediates.

1.4.7 | Structural aspects of SLX4 and its nucleases

The current structural landscape of the SMX complex is summarised in Figure (Section 1.8). While significant progress has been made in the last decade, structural coverage encompasses an estimated 63% of all SMX proteins (51% if only mammalian homologs are considered) with particular shortcomings in structural data on SLX4, mammalian or otherwise (Figure 1.8b, top left panel) and mammalian SLX1 (Figure 1.8b, top right panel).

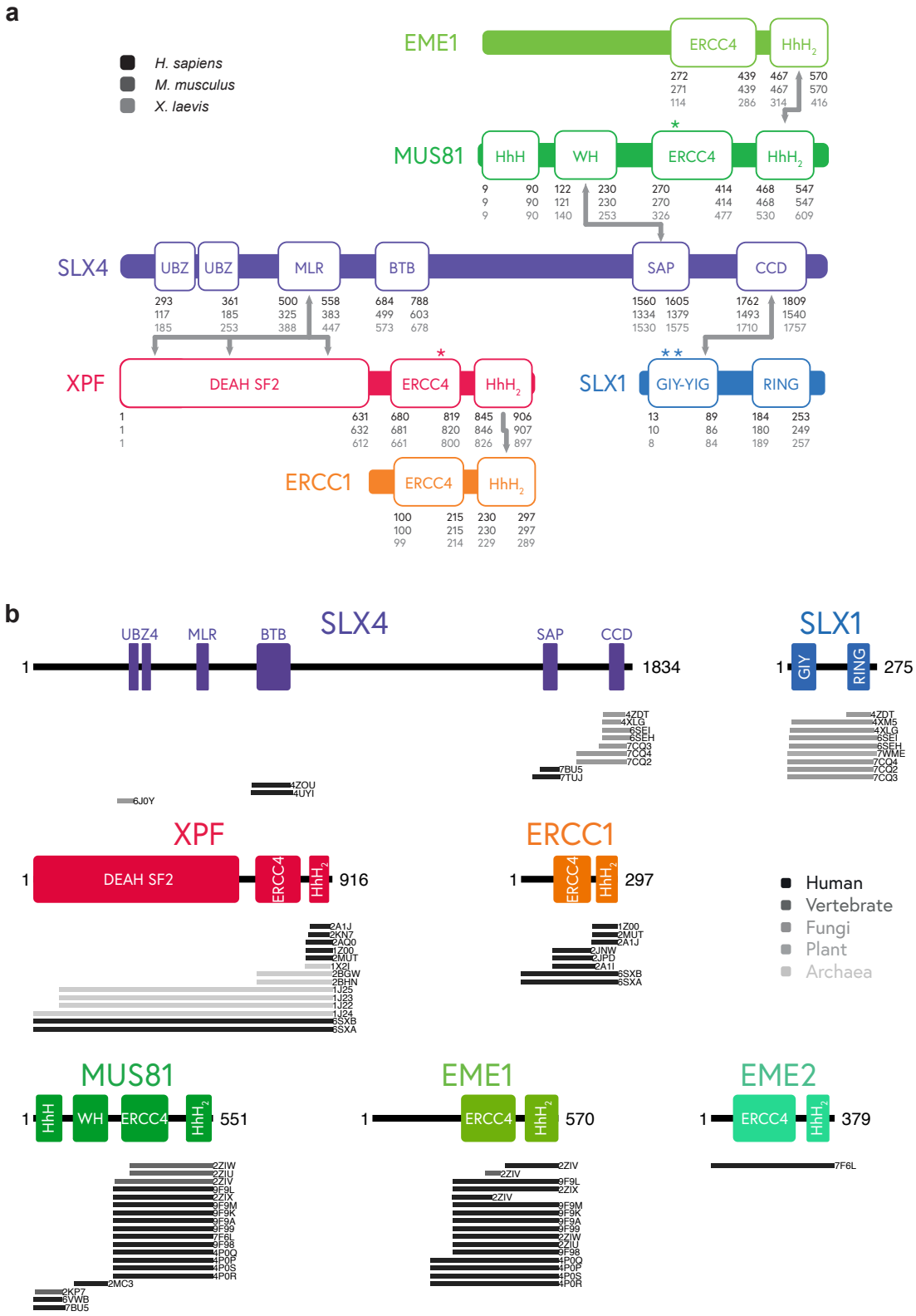

a
H. sapiens
M. musculus
X. laevis
EME1
ERCC4
HhH2
MUS81
HhH
WH
SLX4
UBZ
MLR
BTB
SAP
CCD
XPF
DEAH SF2
SLX1
GIY-YIG
RING
ERCC1
b
SLX4
UBZ4
1834
SLX1
275
XPF
916
ERCC1
297
Human
Vertebrate
Fungi
Plant
Archaea
MUS81
551
EME1
570
EME2
379

Figure 1.8 | The SMX complex and prior structural information. a, Domain organisation map of SMX adapted from ref[324]. Domains listed: UBZ – ubiquitin binding Zn finger; MLR – MUS312-MEI9 interaction-like region; BTB – broad complex, tramtrack and bric-a-brac; SAP – SAF-A/B, Acinus and PIAS; CCD – C-terminal conserved. DEAH SF2 - DEAH box containing superfamily 2; ERCC4 – excision repair cross-complementation group 4 catalytic domain; HhH2 – dual helix-hairpin-helix; GIY-YIG – GIY-YIG motif-containing nuclease domain of SLX1; RING – really interesting new gene Zn finger domain; WH – winged helix domain. Arrows indicate interacting regions, * denotes key catalytic residues. **b,** Structural coverage of the SMX complex. Mapped sequences of prior structural data for each SMX protein (and EME2) component in the PDB, aligned to the human protein sequence by the Biopython package function `localxx`. Black lines indicate human structures, with orthologues in other greys, as shown.

1.4.7.1 | The sparsely annotated architecture of SLX4

SLX4, much like the long-sought BRCA1 complexes or RAD51 paralogues (recently resolved structurally[671]), exemplifies a largely unstructured protein with multiple interaction partners critical for its stability and function. The structural characterisation of SLX4 remains limited, with only a few small, ordered, and highly conserved domains identified within its 1834 amino acid length. Notably, the BTB domain, essential for SLX4's nuclear localisation and foci formation, particularly at telomeres, has been resolved as a homodimer[523]. This structure revealed key contact residues, F681 and F708, critical for dimerisation and highlighted a buried hydrophobic surface, suggesting that homodimerisation might stabilise this region in cells, as no other oligomeric states were observed[523]. Interestingly, an unpublished crystal structure (PDB: 4UYI) reports a homotrimeric BTB domain. However, these findings may be influenced by crystal packing artefacts and might not fully represent the oligomeric state within cells. Additional insights come from the yeast Slx4, where the CCD forms a complex with Slx1, revealing how this region engages one dsDNA arm of a $5'$ DNA flap. Furthermore, work by Wyatt and colleagues demonstrated how the SAP domain of SLX4 alleviates an autoinhibitory interaction with the N-terminal HhH motif of MUS81[324]. Phosphorylation of the SAP domain was shown to induce structural compaction, enhancing its affinity for MUS81[577]. Together, these findings highlight SLX4's modular architecture and its dynamic and intrinsically disordered nature as a scaffold in DNA repair.

Inspection of AlphaFold's predictions for SLX4, or SLX4-nuclease(s) complexes expectedly offers little insight beyond what has been deposited in the PDB. A prediction of all components at equal stoichiometry highlights all established/theorised interactions (Figure 1.9a, b): 1, the GIY-YIG and RING domains of SLX1 and SLX4's CCD; 2,

the N-HhH domain of MUS81 and SLX4's SAP domain; 3, regions of XPF's SF2 helicase-like domain and ERCC4 domain and SLX4's MLR domain. A weaker interaction (AlphaBridge i-score of 0.45, usual cutoff 0.5) predicted between the ERCC1 HhH domain and SLX4's UBZ4 domain could suggest a novel point of XPF-ERCC1 regulation. While showing an overall compact fold and what appears to be interfaces between unexpected interactors (Figure 1.9c, *e.g.* between ERCC1 in orange and EME1 in light green), there is generally low confidence (pLDDT) in the majority of SLX4's structure and these non-canonical interfaces score poorly in overall PAE (Figure 1.9a), highlighting how experimental structural information is still indispensable for SLX4, even in the age of artificial intelligence-driven structure prediction.

1.4.7.2 | SLX1

SLX1 is a unique SLX4-guided structure-selective endonuclease, distinguished by its polarity in DNA incision. Unlike XPF and MUS81, which exhibit 3′-flap endonuclease activity, cleaving double-stranded DNA to produce 3′-ended single-stranded DNA (ssDNA) flaps, SLX1 harbours oppositely orientated 5′-flap endonuclease activity. This specificity is crucial for processing diverse DNA structures during replication, recombination, and repair. SLX1 comprises two key domains: the N-terminal GIY-YIG endonuclease domain (residues 13-89) and the C-terminal RING/FYVE/PHD-type zinc finger domain (residues 184-253)[522]. The GIY-YIG domain features compact "GIY" and "YIG" motifs, coordinated by a central arginine and C-terminal glutamate critical for divalent metal ion binding and catalytic activity[519]. ANKLE1, another member of the GIY-YIG nuclease family, shares structural features with SLX1 and plays roles in DNA damage repair, particularly in chromatin bridge cleavage during mitosis, mitochondrial DNA cleavage, and activation of mitophagy and STAT1 signalling[672,673]. However, the functional parallels between ANKLE1 and SLX1 remain largely unexplored. Structural studies of yeast homologs of human SLX1 provide insights into the organisation of the GIY-YIG domain, which comprises a network of five β-sheets surrounded by α-helices. The conserved GIY-YIG hairpin, formed between $\beta1$ and $\beta2$ strands, and catalytic residues — an arginine in $\alpha1$ and glutamate in $\alpha2$ — are central to its nuclease activity[517,540]. Comparative fungi structures reveal differences in domain organisation where a β-hairpin between $\beta4$ and $\beta5$ is present

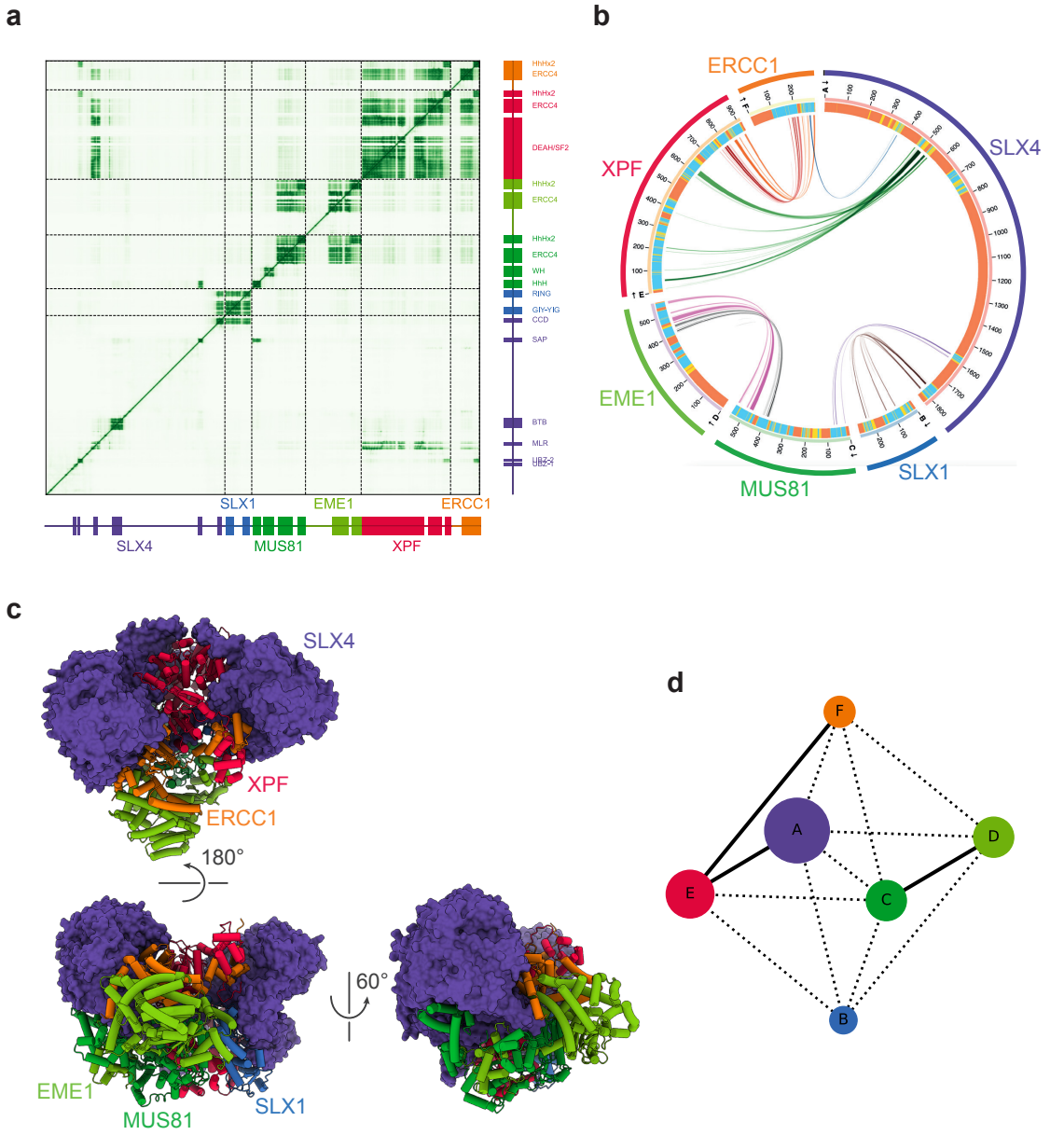


Figure 1.9 | AlphaFold3 prediction of the SMX complex. a, Per-residue predicted aligned error (PAE) for the top-ranked (based on average pLDDT) models of SLX4-SLX1-MUS81-EME1-XPF-ERCC1 (SMX) from AlphaFold3. **b,** Circos plot highlighting AlphaBridge-predicted interaction interfaces between SMX components using the model from **a**. **c,** Structural model of SMX using AlphaFold3. SLX4 is shown as a surface (purple) to highlight how each of XPF-ERCC1 (red-orange), MUS81-EME1 (greens) and SLX1 (blue) individually interact with SLX4. **d,** Contact map highlighting notable buried surface area between SMX components. Nodes reflect each component with their chain ID labelled. The edges between nodes are solid where chain-chain interfaces bury ≥ 300 Å^2, and dotted where the buried surface area is $<$ half of the greatest interacting area.

in apo forms of Slx1 but absent in Slx4-bound complexes, suggesting conformational changes upon Slx4 binding. A notable feature in the structure of Slx1 is a long α-helix that scaffolds the GIY-YIG and RING finger domains (Figure 1.10), with varying lengths observed in different fungi species. This variation's impact on Slx1 function remains unclear and warrants further investigation, especially since human SLX1 replaces this region with a long disordered sequence. The RING finger domain consists of α-helices and β-sheets with two zinc ions. Unlike RING domains in E3 ubiquitin ligases, SLX1's RING domain likely stabilises its interaction with SLX4 or mediates homodimerisation in the absence of SLX4[517]. In *Candida glabrata*, Slx1 forms a stable homodimer in the absence of Slx4, partially blocking nuclease activity by obstructing the active site. Prevailing theories suggest that Slx4's orientation presents a competitive opportunity for heterodimerisation instead. The dimerisation of Slx1 may reflect a fundamental principle of folding that necessitates the burial of polar surfaces, potentially leading to proteasomal degradation if SLX1 is forced to remain as a monomer without Slx4. A plant (*Arabidopsis thaliana*) SLX1-like homolog, *At*HIGLE, cleaves branched DNA substrates *in vitro* but without an SLX4 homolog and unlike the *C. glabrata* SLX1, exists as a functionally active dimer able to produce two symmetric nicks around a HJ[674]. Monomerisation can be achieved by removing *At*HIGLE's C-terminal region, which enhances and changes the nature of nicking[674]. However, in human cells, SLX4 is required for SLX1 stability, discrediting theories of regulation through homodimerisation[527]. SLX1's exhibition of nuclease activity on various DNA substrates, including replication forks, splayed arms, $5'$ and $3'$-flaps, and Holliday junctions, is reliant on DNA binding by three positively charged sites. Site I residues in $\alpha 1$ helix bind dsDNA near the branchpoint, critical for substrate positioning[675,676]. The second site is located in the crucial glutamate-containing helix and interacts with ssDNA in a manner essential for catalytic activity[675]. The final site involves the GIY-YIG hairpin loop and other structural elements, ensuring catalysis occurs only when the DNA is correctly aligned[676]. Mutations in Sites I and II impair DNA binding and catalytic activity, highlighting their cooperative roles in DNA incision[521].

A key challenge with structural models of SLX1 is the discrepancy between its perceived *in vitro* activity and the placement of active site residues on model substrates. *In vitro*, the primary cleavage product is typically an incision between the 3rd and 4th

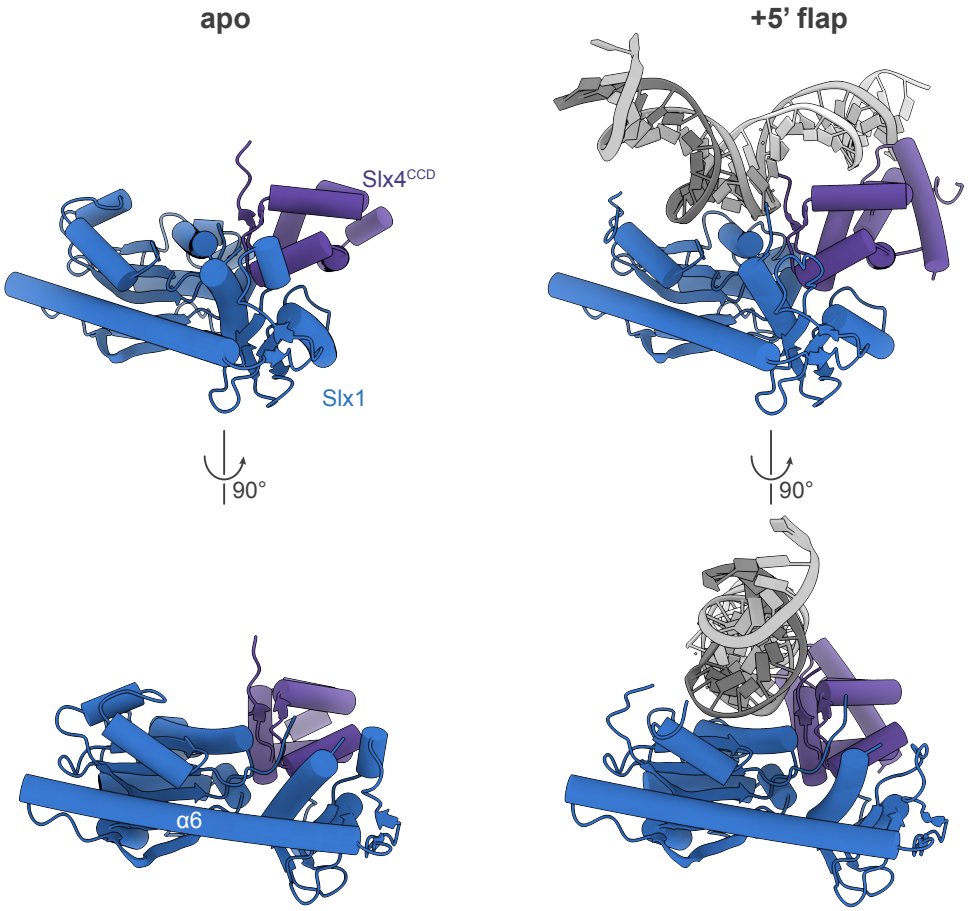


Figure 1.10 | Structures of Slx1-Slx4 in an apo and substrate-bound form. The overall structure of *S. cerevisiae* Slx1-Slx4^{CCD} in apo (left, PDB: 7CQ3) and 5′-flap-bound (right, PDB: 7CQ4) forms. The long α6 helix of Slx1 is highlighted.

nucleotide 3′ of the ss/dsDNA junction. Mechanistically, it was proposed that SLX1's action might resemble that of the Hpy188I restrictase, where the tyrosine residue in the GIY motif (specific residue in human SLX1 is to be confirmed) serves as a general base for activating a nucleophilic water molecule[675,677]. SLX1-SLX4 is a unique structure-selective endonuclease capable of cleaving various branched DNA structures. This versatility is critical for maintaining genome stability, playing roles in responding to replication stress, regulating telomere function, and participating in several DNA repair pathways, including homologous recombination and ICL repair.

1.4.7.3 | The common folds of MUS81-EME1/2 and XPF-ERCC1

As an obligate heterodimer, XPF-ERCC1 dimerises through helix-hairpin-helix motifs (HhH_2), similar to MUS81-EME1. The N-terminal region of the SLX4 scaffold interacts

with XPF-ERCC1 *via* its MLR[513,586], enabling replication-coupled ICL repair as part of the FA pathway. SLX4 enhances the nucleolytic activity of XPF-ERCC1 towards fork-like structures, especially containing an ICL[435]. XPF comprises a helicase-like domain at its N-terminus, which links to the HhH domain through the central catalytic, excision repair cross complementation group 4 (ERCC4) endonuclease domain[678]. This ERCC4 domain carries the critical catalytic residues within its central $\beta 3$ and $\beta 4$ strands that make up the GDX_nERKX_3D motif, which is similar to the $PDX_n(E/D)XK$ motif that define catalysis in type II restriction endonucleases[679]. Though ERCC1 doesn't participate in catalysis, it also adopts a central ERCC4-like domain, though the ^{725}ERKSISD residues that are present in XPF are mutated to ^{139}LFLSLRY in ERCC1[680]. This is a feature of the other two human complexes that also make use of ERCC4-like domains. MUS81, with catalytic residues intact, contains an ERCC4 domain, and so do both of its binding partners, EME1 and EME2[681]. Like ERCC1, catalytic residues in the ERCC4-like fold of EME1/2 have become non-functional. Though the overall domain structure also applies, the FANCM-FAAP24 complex breaks the asymmetric catalytic/non-catalytic heterodimer rule that pertains to XPF/MUS81 complexes. Both FANCM and FAAP24 have lost catalytic residues, though the divergence in FANCM is comparatively minimal, maintaining ^{1864}ERRSQSE in this motif[682]. Despite not all three complexes showing characteristic $3'$-flap-like nuclease activity, they all show an affinity for binding junction-containing DNA structures[680]. Unsurprisingly, the major domains of XPF/MUS81 employ helicase-like domains, since helicases can capture both dsDNA and ssDNA at some point in their catalytic cycles, which are often reversible[683].

XPF Structures of archaeal homologs of XPF, such as the *Pyrococcus furiosus* or *Aeropyrum pernix* Hef, form homodimers, providing valuable context to the evolutionary origins of ERCC4 domain-containing heterodimers[684,685]. DNA binding appears to compact the overall structure into an active conformation[685]. The HhH domain makes significant contacts with DNA, as it does in a *Sulfolobus solfataricus* XPF homolog[686], and a linker between the HhH domain and the ERCC4 domain facilitates domain movement to enable recognition and cleavage. The flexibility in this linker also allows the formation of an asymmetric dimer where one protomer is catalytically active, mimicking XPF-ERCC1 heterodimers. The structural model suggests that the branched DNA substrate binds the

HhH domain dimer in two orientations with linker flexion activating one nuclease domain to cleave the 3′ overhang[686]. Importantly, two separate sites engage the upstream duplex and ssDNA regions individually, but their orthogonal positioning necessitates a 90° bend at the junction, which is likely encouraged in cells by the stabilisation of TFIIH/XPA in NER[687] or SLX4 in the FA pathway[686]. It is possible FANCM's supposed remodelling of converged forks into a chicken-foot structure also promotes distorting to ∼90° at the ssDNA/dsDNA junction to facilitate XPF binding and cleavage[688].

Both the archaeal Hef proteins and human FANCM contain functional superfamily 2 (SF2) DEAH domains. Functional helicase activity in the equivalent domains of human XPF and MUS81 due to missing residues that bind and hydrolyse ATP[689,690], but the Hef helicase domain combines and enhances the endonuclease activity on fork structures[691]. Nevertheless, the DEAH-like domains in XPF and MUS81 contain a pair of classical RecA-like folds that enact DNA translocation, separated by an α-helical domain[692]. While inactive, deleting the helicase domain severely reduces the cleavage activity of XPF but does not change the substrate-specificity profile *in vitro*[693].

For human homologs, structures of the HhH domains of XPF and ERCC1 alone highlight how while both participate in heterodimerisation, only the HhH domain of ERCC1 binds dsDNA[694–696]. One set of full-length structures of XPF-ERCC1 was also made available more recently by the Neil McDonald group. These show XPF-ERCC1 in an auto-inhibited conformation whereby XPF's helical domain conceals the HhH domain of ERCC1 and precludes access to XPF's catalytic site[696]. In this state, XPF's HhH and ERCC1's ERCC4-like domain are fully solvent-accessible and able to interact with their predicted ssDNA region. With DNA bound, XPF's SF2 domain disassembles, which releases the HhH domains of XPF and ERCC1, and these domains can then coat the minor groove of the junction-adjacent dsDNA duplex[696], though primarily through interactions of ERCC1[694,695]. This state resembles the DNA-bound structures of Mus81-Eme1[697] and *A. pernix* XPF[686], highlighting the compaction of the XPF/MUS81 HhH domains against ERCC4-like domains of ERCC1/EME1. Displacement of the HhH domains uncovers the dsDNA-binding surface of ERCC1's Hhh domain and properly synergises both ss- and dsDNA-binding activities of ERCC1 by compacting the overall ERCC4-HhH fold[696]. However, the autoinhibitory apo state may be an artefact alleviated in cells through

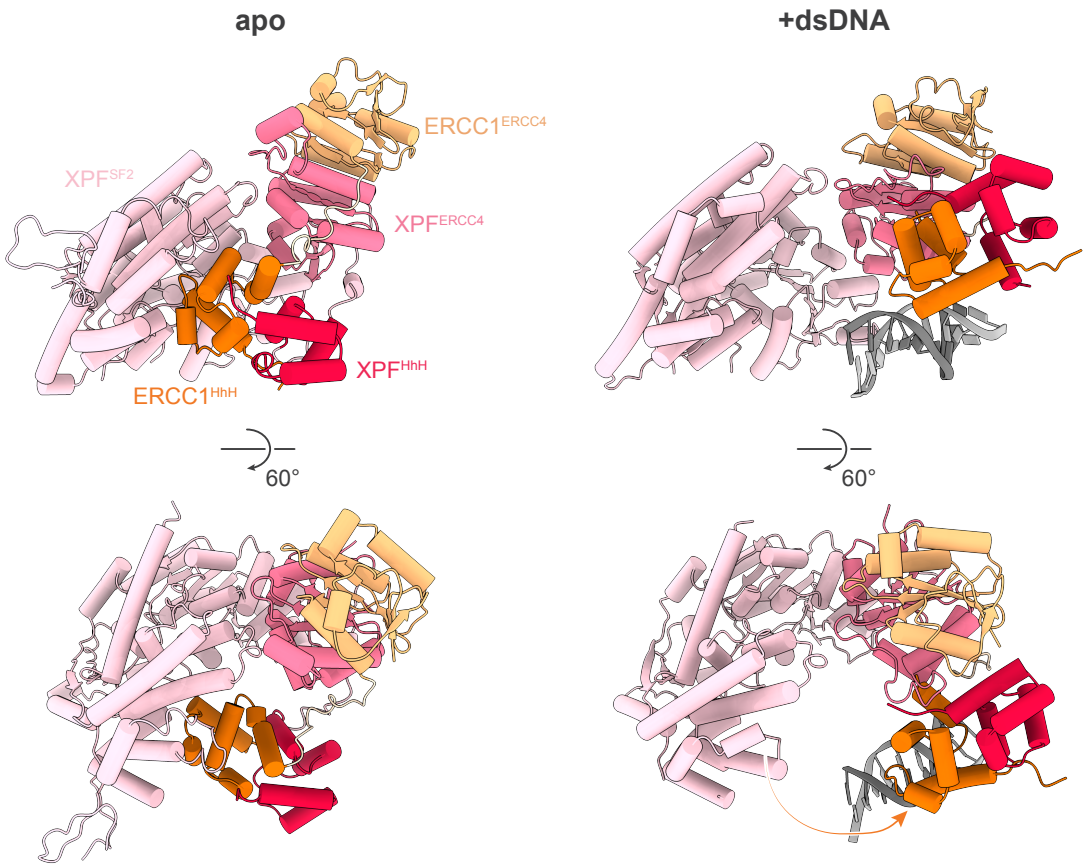


Figure 1.11 | **Structures of XPF-ERCC1 in apo and substrate-bound forms.** The overall structure of *H. sapiens* XPF-ERCC1 by cryoelectron microscopy in apo (left, PDB: 6SXA) and dsDNA-bound (right, PDB: 6SXB) forms. Upon substrate binding, the movement of the ERCC1 HhH domain away from the XPF SF2 helicase domain is shown by an arrow.

the licensing/activation of XPF-ERCC1 by SLX4 for ICL repair, or XPA for NER. A surprising finding was that XPF-ERCC1 preparations were abundantly constituted of a heterotetrameric state, which could not be resolved with improvement beyond 14 Å[696].

MUS81 Mus81 was identified through its cooperation with Rad54, a recombination/repair protein, in *S. cerevisiae* and the checkpoint kinase Cds1^{CHK2} in *S. pombe*[566,698,699], while its human counterpart was later discovered through homology searches[700]. MUS81 features separated HhH motifs at its N- and C-termini, with the C-terminal region also containing a pseudo-HhH motif whose functionality remains unclear[701,702]. The MUS81 ERCC4 nuclease domain also consists of a GDX_nERKX_3D motif with the sequence ^{333}ERKRLDD in MUS81. The non-catalytic partners, EME1 and EME2 in humans or Mms4/Eme1 in *S. cerevisiae*/*S. pombe*, also contain HhH motifs, contributing to

DNA binding and substrate specificity[607,703]. Biochemical studies have shown that MUS81 complexes preferentially cleave 3′-flap, replication fork, and nicked Holliday junction (nHJ) substrates, while intact Holliday junctions, splayed-arm DNA, and 5′-flaps are poor substrates, distinguishing MUS81 from other XPF family members like XPF-ERCC1[697,704]. Mechanistically, MUS81-EME1 and MUS81-EME2 introduce nicks near the branch points of 3′-flap structures, with nicking efficiencies influenced by the proximity of the 5′ terminus to the junction[607]. MUS81-EME2 was seen to showcase unique promiscuity lacking for MUS81-EME1 in cleaving 5′-flaps[607,704], but a follow-up study describes indistinguishable substrate, albeit requiring a nearly 10-fold greater molar ratio of enzyme:substrate[705]. A well-conserved winged helix domain (WH) differentiates MUS81 from other ERCC4-fold paralogues, which is necessary for meiotic recombination and DNA damage tolerance by impacting MUS81-EME1 cleavage patterns and overall activity[703]. MUS81-EME1 experiences substrate-induced conformational changes that include a hydrophobic wedge separating pre- and post-nick duplex DNA and binding pocket, which engages the 5′-nicked end. This facilitates the sharp bending of 3′-flap DNA and precise strand alignment at the active site[697]. A structure of MUS81-EME2 aligns well with the MUS81-EME1 structures[705].

MUS81 interacts with the SLX4 scaffold through the SAP domain of SLX4, which extends from residues 1560 to 1605, N-terminal only to the CCD that binds SLX1. *In vitro*, SLX4 influences MUS81 through stimulation of its catalytic activity and by expanding its substrate range[323,324]. It is believed that MUS81's involvement in SMX is confined to a perhaps transiently formed tri-nuclease complex with SLX1 and XPF-ERCC1[324]. The current model describes that full SMX formation occurs predominantly in mitosis[324]. SLX4 SAP's affinity for MUS81 grows 10-fold when phosphorylated by the mitotic CDK1 kinase[577], while phosphorylation of MUS81's N-HhH domain by CK2 is a prerequisite for disassembly of this autoinhibitory loop such that SLX4 can bind[320]. Six predicted CDK1 target sites, of which three (T1544, T1561, T1571) were confirmed, in SLX4's SAP domain are vital for interaction with MUS81, while phosphorylation of a single site, S87, in MUS81 determines interaction with SLX4[577]. Phosphorylation of S87 has been observed to increase steadily when cells are exposed to replication stress and while they progress through G2 into prophase, though disappearing sharply during

metaphase[579]. Presumably, this is a means to ensure no metaphase intermediates are cleaved prematurely. The untimely SMX formation, through inhibition of the CDK1-inactivating WEE1 kinase or genetic modification of Cdk1, leads to gross chromosomal instability and chromothripsis[579]. A phosphomimetic MUS81, S87D, phenocopies the harmful effects of premature SMX assembly[579].

1.5 | Background of particular methods in this study

Advancements in high-throughput sequencing technologies and CRISPR-Cas9 systems have transformed the study of DNA damage and repair, offering unprecedented precision in mapping damage and dissecting the genetic underpinnings of repair pathways. Sequencing-based DNA repair assays leverage the ability to locate and quantify DNA damage, mutations, and repair intermediates across the genome. CRISPR-Cas9-based genetic screens extend these capabilities by systematically perturbing genes involved in DNA damage response pathways[706]. The simplicity and versatility of CRISPR technologies have revolutionised forward and reverse genetic approaches, allowing genome-wide identification of genes critical for repair processes under diverse conditions. These screens have been instrumental in elucidating repair mechanisms, identifying synthetic lethal interactions, and uncovering potential therapeutic targets, especially in the context of cancer and genome instability.

1.5.1 | Sequencing-based DNA repair assays

In an age of ever-decreasing high-throughput sequencing costs, even assessment of DNA-damaging events can be measured by genomic studies, strengthening the findings of classical techniques evaluating bulk DNA damage. The locations of these adducts can be highly informative, assisting in understanding the precise mechanisms that govern how the damage forms and the molecular details of their repair. Initial forays into this field of 'adductomics' meant inferring where damage lies through whole genome sequencing and characterising the scars that define lesions after repair or replication. These scars were termed mutational signatures and have been painstakingly annotated over the decades[707,708]; as a notable example, cigarette smoke defines a particular signature that

displays gross pulmonary C>T and C>G mutations, possibly due to overactive APOBEC deaminase action from inflammatory reactions that arise in response to particulates[709]. Mutational hotspots and signatures have thus been well-defined for many genotoxins and those appearing from particular neoplasms[707]. As noted earlier, it is not the damage but how a cell sees that damage that defines most aspects of the underlying biology. Consequently, while DNA damage can be scattered (albeit with some occasional sequence specificity), their interactions with replication, transcription and DNA repair machinery influence how they are distributed through the genome[710].

Many methods exist to quantitatively assess DNA damage beyond classical phenotypic studies focusing on cytotoxicity. The comet assay measures the total abundance of DSBs through single-cell electrophoresis, though it can be adapted for SSBs and ICLs by first denaturing genomic DNA with or without *ex vivo* ionising radiation, respectively[711]. Mass spectrometric (MS) methods are considered the most sensitive for deciphering what particular damage is present in DNA. As an example, given that 100 different oxidative marks exist, the resolution offered by MS methods is key to deciphering damage beyond *e.g.* 'oxidative damage'[712]. However, MS techniques are often very involved with significant preparation steps, surpassing the resources of many laboratories and providing no information on the location of adducts. Proxies for the presence of damage, *e.g.* through immunochemical means, are ubiquitously used - damage is often quantified by Western blotting, flow cytometry, or immunofluorescence. In the latter, the intensity or number of punctate foci of damage-specific markers, such as γH2AX[713], 53BP1[714] or phospho-RPA2[715], can be probed. These proxies have been adapted to more 'omic' methods, mapping the location of particular damage across the genome. Chromatin immunoprecipitation (ChIP) and sequencing (ChIP-seq) studies of γH2AX have been performed to identify the locations of DSBs and how this damage-associated chromatin mark is propagated during damage sensing[374,716]. Damaged DNA immunoprecipitation (DDIP) is a catch-all term for using antibodies to pull down specified damage - a CPD-specific antibody can immunoprecipitate UV-induced damage and demonstrate the exceptionally high frequency of CPDs at telomeric sequences[717]. Enzymes fulfil a vital role in some sequencing techniques. T4 endonuclease acts as a cyclopyrimidine dimer-specific glycosylase, cleaving the phosphodiester bond between an intrastrand

crosslink and generating a strand break and in an antiquated approach, quantifying intrastrand crosslinks after UV damage involved Southern blotting of genomic DNA to compare the patterns of bands after treatment of genomic DNA with and without T4 endonuclease V (T4endoV)[718]. Instead, ligating adaptors to end-repaired fragments and cleaving with T4 endoV/APE1 permits an NGS approach for the identification of hypothesis-free intrastrand crosslink regions (CPD-seq)[719,720]. The practice whereby broken DNA acts as a starting point for adaptor ligation has proved fortuitous in break mapping. The first method to find locations of DSBs was BLESS (breaks labelling, enrichment on streptavidin and next-generation sequencing), which involved the labelling of *in situ* breaks after formaldehyde fixation and enrichment of broken DNA fragments on streptavidin before sequencing[721]. The technique was used to map aphidicolin-sensitive regions and highlight the many off-target DSBs produced by CRISPR-Cas9[721]. In efforts to circumvent the high background and lack of information regarding the structure of the DNA ends, the Nussenzweig lab developed END-seq[722,723], improving upon BLESS through the embedding of living cells in agarose plugs to limit high background artificial breaks in DNA caused by mechanical shearing during lysis or even more pertinently, through damage caused by formaldehyde during fixation[722]. With END-seq, or other similar DSB mapping techniques, DSB landscapes can readily be transformed into mapping SSBs instead by simply adding an enzymatic step that cleaves all ssDNA breaks into DSBs. This is achieved through the S1 nuclease, which cleaves gapped dsDNA into ends amenable for END-seq[724]. This technique unveiled the presence of non-B DNA cruciform structures that form from TA repeats, which accumulate in cancer cell lines with MSI[724].

Given that other forms of DNA damage are often funnelled through SSBs or DSBs as an intermediary step in their repair, identifying the location of DNA breaks can, in effect, be a proxy for the location of most other breaks. Therefore, with well-timed END-seq or S1-END-seq, it may be possible to map snapshots of any damage and their repair outcomes. Incorporating nucleotides capable of immobilisation has materially impacted techniques looking to map regions of synthesis. EdU incorporation, instead of thymidine, means that streptavidin can pull down DNA fragments after click-biotinylation of the EdU[725]. This discovery has meant that regions undergoing DNA replication can be mapped precisely, and following this on with precise timings after *e.g.* G1 arrest has meant that origins of

replication could be mapped[726]. Similarly, regions prone to MiDAS can be identified by a similar approach by collecting only mitotic cells (*via* mitotic shake-off) and sequencing EdU incorporated regions - sites of MiDAS were, as expected, enriched at all known CFSs[727]. It was observed that leading and lagging strand synthesis were not coupled during MiDAS, hinting further that the form of synthesis occurring in mitosis is a form of conservative break-induced replication[727]. The Nussenzweig group also developed an orthogonal method to view DNA damage in a highly specialised context. Where cells can be constrained to only G1, like in post-mitotic neurons or imatinib-induced G1-arrested Abelson pre-B cells, EdU incorporation *via* DNA synthesis only occurs in the context of repair[728]. Combined with S1 END-seq, the technique revealed how neurons accumulate SSBs, especially at neuronal enhancers or sites of DNA demethylation, requiring PARP1- and XRCC1-dependent SSB repair[728]. The theme of mapping DNA breaks has also allowed adaptation to the mapping of other DNA features, such as misincorporated bases like ribonucleotides. Recombinant RNase H2 treatment will cleave at embedded ribonucleotide loci, permitting mapping through a break-repair-sequencing technique like BLESS; this technique has been termed emRibo-seq[729].

1.5.2 | CRISPR-Cas9-based genetic screens

For many decades now, understanding protein molecular function has relied heavily on the alternating use of forward and reverse genetics[730]. Forward genetics, described simply, defines genotypes from phenotypes. How this worked in practice entailed random mutagenesis followed by phenotypic selection and genotyping of isolated mutants that conferred a particular phenotype. In the complementary reverse genetics model, arguably far more prevalent and relevant in this age, genes that lack defined phenotypic characterisation are perturbed, and phenotypes are then designated. Matched isogenic pairs of cell lines are crucial for this paradigm of cell biology, and comparisons between WT and gene-disrupted (loss *via* knockout or depletion *via* knockdown) allow a critical understanding of a gene's molecular function, its role in disease and its potential as a therapeutic target. What unites reverse and forward genetics are high-throughput screening approaches, such as through genome-wide sequencing[731] or large-scale proteomics[732]. These strategies can be particularly helpful in understanding vulnerable protein targets, identifying key

disease biomarkers or establishing the parameters of therapeutic success. CRISPR-Cas9 strategies have revolutionised both branches of performing these experiments in higher eukaryotes. The barrier to the generation of genetic knockouts has been greatly lowered by CRISPR-Cas9's simplicity, compared to previously inflexible technologies such as TALENs (transcription activator-like effector nucleases) or Zinc-finger nucleases[733,734]. Whereas these older techniques necessitated precise protein design to target genomic loci[735], the Cas9 nuclease targets a DSB at loci specified by a single guide RNA (sgRNA) sequence, which can readily be programmed akin to knockdown *via* siRNA or shRNA, though now offering total knockout[733]. One caveat to CRISPR-based gene targeting is the requirement of an adjacent NGG sequence named the protospacer adjacent motif (PAM), but given the incidence of this sequence being 1 in 16, this is generally not a significant barrier. The specified DSB can lead to the formation of indels when repaired by the mutagenic NHEJ pathway, occasionally arising in frameshifts that inevitably introduce stop codons. If, instead, an edit is required, the DSB can be chosen in a manner that promotes invasion into a user-provided DNA template with sufficient homology to trigger HR but harbouring the required mutation[736]. This method allows the introduction of specified mutations or endogenous gene tagging. CRISPR-Cas9 technologies have now surpassed far beyond the induction of a simple DSB. Fusing modular units to Cas9 has expanded its scope to a near-unlimited range of uses. Inactivating either the RuvC domain through D10A, or the HNH domain through H840A, converts Cas9 into a nickase (nCas9) capable of nicking either the sgRNA-targeted or the non-target PAM strand, respectively[737]. Inactivating both domains creates a nuclease-dead Cas9 (dCas9), and any fused proteins could, therefore, be targeted to a sgRNA-specified locus without Cas9 enacting any DNA breaks[738]. Early applications fused dCas9 to various effector proteins listed below:

1. Base editors, *e.g.* APOBEC, allowed the introduction of specific point mutations[739].
2. Transcriptional activators *e.g.* VP64-p65 - termed CRISPRa[740–742].
3. Transcriptional repressors, *e.g.* KRAB - termed CRISPRi (interference)[743].
4. Fluorescent reporters, *e.g.* EGFP[744].

5. Reverse transcriptase, *e.g.* MLLV-RT to copy mutations from an extended sgRNA (pegRNA) into the genome - termed prime editing[745].

Screens have long been a part of advancing biological research, especially concerning DNA damage and repair. RNAi-based phenotypic screens were instrumental in a range of discoveries, including PARPi[284], cisplatin[746] and mitomycin C[747]-sensitising factors. In these screens, different siRNAs or shRNAs were introduced into physically separate cell populations and their phenotypes were read out in a medium-to-high-throughput fashion, such as fluorimetry in a 96-well plate format[748]. What restricts this approach is the limited number of genes available for testing, the potential off-target risk of RNAi and the only partial reduction in gene expression[748]. Since only a single complementary sgRNA specifies a CRISPR-Cas9 target, the potential of CRISPR-Cas9 technologies far surpasses the ability of RNAi-based screens. Supposing a population of cells is created such that each cell contains a different sgRNA, perhaps targeting a single gene, it is possible to understand how the loss of any targeted gene affects fitness in various conditions. Using CRISPR knockout (CRISPR-KO) screens, a core and cell-line specific essentialome was generated[749], now available as a database in the CRISPR DepMap[750]. CRISPR-KO screens were soon adapted and complemented by CRISPRi and CRISPRa technologies, which adopt sgRNAs targeting the promoters of various genes[751]. Given that CRISPR-KO screens are predicated on the ability of DNA repair itself, *e.g.* the NHEJ machinery, to affect inactivating mutations in genes, the use of CRISPRi offers a more justifiable alternative for studies of DNA repair[751]. However, screens within DNA repair are now ubiquitous. They are usually present in two varieties: one uses paired mutant and WT cell lines, and the differences in surviving sgRNAs are compared; another compares the surviving sgRNAs where a drug is applied compared to untreated or its vehicle. Both types of screens have been used to draw an understanding of the PARP inhibition paradigm. PARP KO cells or PARPi using olaparib/talazoparib were compared by CRISPR-KO screens, and their differences were studied, showing that the BER and SSB repair factor, XRCC1, strongly sensitises to PARPi but promotes survival in PARP1 loss[752]. In BRCA1/2-deficient cells, CRISPR-KO screens were used to understand their hypersensitivity to PARPi, pulling out well-known factors, loss of which confers resistance, such as 53BP1 and RIF1, but also identified a key NHEJ-promoting complex referred

to as Shieldin, consisting of genes SHLD1-3 that drive PARPi resistance[394,396]. A more recent screen aimed to examine the mechanisms cells can resort to when deprived of functional NHEJ and HR, through loss of LIG4 and BRCA2, respectively, which identified the 9-1-1 complex and its interacting partner RHINO as key MMEJ factors[343]. The study also described the physiological impact of MMEJ in mitosis-persisting DSBs, potentially shedding light on the synthetic lethal relationship between the MMEJ *POLQ* gene and BRCA1/2[343]. This further impacts the clinic as it helps explain the efficacy of POLQ inhibitors alongside PARPi. The genetic response of RPE-1 $TP53^{-/-}$ cells was evaluated under a CRISPR-KO screen against 27 genotoxic agents, which provided a landscape of genetic determinants that govern cytotoxicity[446]. One significant caveat in genome-wide CRISPR-KO screens is the p53 status of cells, which can limit the survival of Cas9-mediated DSBs by sensing and inducing apoptosis when this occurs[753] but can be avoided by use of cell lines with mutant or knockout *TP53*, with the caveat this does also produce intrinsic changes to the DNA damage response.

The classical CRISPR-KO screen involves packaging lentiviruses with plasmids containing a distinct sgRNA sequence per lentivirus. Within the ensuing pool of packaged lentiviruses, multiple sgRNAs are encoded for each target gene to average off-target effects and variation in editing efficiencies. The screen can be performed in cells stably expressing Cas9, circumventing the need for Cas9 to be included in the plasmid of each lentivirus; however, many libraries also encode Cas9. The TKOv3 library[754,755] expresses Cas9 within each lentiviral construct. A low multiplicity of infection ensures that each cell is infected with only one sgRNA-lentiviral particle. A coverage of 200 or 500 is often adopted for each sgRNA to ensure depletion/enrichment events are identifiable. Since sgRNAs act as a unique barcode, but the plasmids encoding them are otherwise invariant, a single PCR can amplify around the sgRNA, and each is sequenced. Once a stably integrated population of cells is established, an initial timepoint (t0) sample is harvested to account for the initial library representation, and time points are taken throughout the screen, lasting at least 15 divisions, such that even slower turnover proteins can be depleted. The sgRNAs which have been lost are compared to those that make up the initial t0 sample, with under-represented sgRNAs denoting essential genes combating a perturbation, while any enrichment suggests genes that promote survival.

1.6 | Outstanding questions and contextual framework

In the expanse of DNA-damaging events, ICLs are uniquely cytotoxic lesions that completely obstruct DNA replication and transcription, requiring complex and often replication-coupled repair pathways for their resolution (see Section 1.3.5). While much of our mechanistic understanding of ICL repair has been shaped by studies using exogenous agents such as mitomycin C[448], cisplatin[756], and psoralens[589], the nature and repair of endogenous crosslinks remain far less well defined. This gap in knowledge is significant given the evolutionary conservation and complexity of the FA repair pathway, suggesting a persistent physiological need to resolve such damage. Among the most compelling candidates for endogenous ICLs are small, reactive aldehydes, particularly formaldehyde and acetaldehyde, generated through diverse cellular processes[9,10]. Formaldehyde, in particular, is released during histone demethylation and is tightly associated with epigenetic reprogramming events, which occur during differentiation and development. If not properly detoxified, formaldehyde can form DPCs[495] or even direct ICLs[757], leading to replication fork stalling and genome instability. One outstanding question explored in this thesis is whether endogenous aldehydes are a physiologically significant source of ICLs, particularly in stem or progenitor cells undergoing differentiation. During such transitions, bursts of formaldehyde from histone demethylation could render the genome vulnerable, especially if aldehyde detoxification (*e.g.*, *via* ADH5/ALDH2) or FA-mediated ICL repair mechanisms are overwhelmed or deficient[59]. Conversely, the consequences of formaldehyde damage in post-mitotic cells are also of interest. While such cells are not actively replicating and typically show low FA pathway activity, they remain epigenetically dynamic. In neurons, for example, synaptic plasticity involves local histone demethylation[758], which could transiently increase formaldehyde levels. Whether formaldehyde-induced lesions accumulate in such cells and how they are resolved, perhaps *via* transcription-coupled nucleotide excision repair (TC-NER)[756], remains unclear.

In addition to aldehydes, other endogenous metabolites may contribute to crosslinking lesions. Products of lipid peroxidation, ROS, and spontaneous oxidation or deamination can all yield DNA adducts that result in strand-crosslinking. These structurally diverse lesions may not all engage classical FA repair, raising the possibility that other complexes

such as SMX (Section 1.4), traditionally linked to replication-coupled ICL resolution, may function more broadly in alternative repair contexts, including transcription-coupled or post-replicative settings.

Addressing these questions is central to the goals of this thesis. The project investigates the interplay between aldehyde metabolism, endogenous ICL formation, and DNA repair through a combination of cellular and structural approaches. Specifically, the work seeks to understand: what biochemical and cellular strategies can improve purification and structural analysis of the SMX complex; whether knockout models of SMX components (XPF and SLX1) can provide insight into the biological consequences of SMX dysfunction; and whether differentiation-induced formaldehyde generation is a major contributor to the genomic instability seen in diseases like Fanconi anaemia. Moreover, it explores whether this instability stems from the replication burden of stem cell proliferation or from formaldehyde release during the reprogramming that accompanies differentiation. Finally, the overarching aim is to better define the architectural and functional properties of the SMX complex, which has long eluded even low-resolution structural characterisation. A full understanding of its dynamic assembly and nuclease coordination could provide key mechanistic insights into ICL resolution and suggest therapeutic strategies for targeting genome instability in cancer and rare genetic disorders.

1.7 | Aims

Structurally, there is limited information available on the individual components and full assemblies of the SMX (SLX4-MUS81-XPF) complex. High-resolution structures of full-length SLX1-SLX4 or SLX4-MUS81 complexes do not currently exist but are necessary to understand the molecular mechanisms of their catalysis. However, obtaining such structures is challenging due to the inherent flexibility and lack of any well-defined structure in SLX4, which appears to form dynamic and transient interactions with its partner nucleases. While genetic data provide insights into the roles of SMX in DNA repair, the structural basis for its assembly and function remains poorly understood. One central question is whether ICL repair constitutes a primary physiological role of the SMX complex and if the deficiency in this pathway accounts for the reduced fitness observed in SMX-null cells or patients with inactivating SLX4 mutations. **Chapter 3** examines

whether aldehydes generated during epigenetic reprogramming can induce crosslinks that require either detoxification or resolution *via* ICL repair pathways. By elucidating the interplay between aldehyde metabolism and DNA repair, this chapter aims to shed light on how endogenous aldehydes contribute to genome instability and whether SMX complex activity is required to mitigate this damage.

To further understand SMX biology, **Chapter 4** focuses on developing CRISPR-Cas9 knockout models targeting XPF and SLX1 in human cells. These models provide a platform to study how individual components of the SMX complex contribute to DNA repair processes. Additionally, paired wild-type (WT) and knockout (KO) cell lines were subjected to CRISPR-Cas9 screens in the presence of crosslinking agents. This approach seeks to identify synthetic lethal interactions and uncover additional pathways that compensate for the loss of SMX components. These studies aim to reveal the genetic networks underlying SMX function and inform potential therapeutic strategies for targeting DNA repair deficiencies in cancer. **Chapter 5** addresses the paucity of structural data that has hindered mechanistic understanding and drug development targeting SLX4 and its associated nucleases. Structural characterisation of SLX4 remains a fundamentally challenging endeavour, as evidenced by the limited annotation provided by even cutting-edge predictive tools like AlphaFold (Figure 1.9). To attempt to overcome these challenges, this project adopted a multi-faceted approach, including the expression and purification of chimaeric SMX complexes from insect cells, mouse SMX complexes from insect cells and mammalian cells and finally, human complexes expressed in mammalian cells. These efforts move towards the eventual end goal of a full-length SMX structure. Alongside this, preliminary functional assays on tractable sub-complexes were completed to corroborate and identify the post-translational modifications (PTMs) critical for SMX assembly and function. Mechanistic biochemistry has already revealed insights into the roles of PTMs in SMX complex stability and activity. These modifications may regulate the dynamic assembly of SMX in response to cellular needs, particularly under conditions of DNA damage stress.

Overall, this project combines genetic, structural, and biochemical approaches to advance our understanding of the SMX complex and its contributions to DNA repair, genome stability, and potential therapeutic interventions.

2 | Materials and methods

Contents

2.1 | Experimental model and subject details

2.1.1 | Human cancer cell lines

Human cell lines (RPE-1, 293FT, HeLa, A549, U2OS, HAP1) were grown in DMEM containing 4.5 g/l D-glucose and 110 mg/l sodium pyruvate (ThermoFisher Scientific, Gibco), supplemented with 10% fetal bovine serum (FBS; Gibco, A5256701), and 100 U/ml penicillin, 100 U/ml streptomycin (P/S; Gibco, 15140). HT-29 cell lines were grown in RPMI 1640 (Gibco, 21875), supplemented with 10% FBS and P/S as above. All above cells were cultured at 37 °C in a humidified atmosphere with 5% CO_2, with routine trypsin-lifted (TrypLE; Gibco, 12604) passaging when up to 90% confluency was reached. Expi293F cells (Gibco, A14527) were cultured in Expi293 Expression Medium (Gibco, A14351), at 8% CO_2 under 75% humidity with 120 rpm shaking rate to maintain suspension, or as adherent cells in a static incubator with 5% CO_2 and 100% humidity in FreeStyle 293 Expression Medium (Gibco, 12338) with 10% FBS. Mycoplasma testing was routinely performed on all cell lines using the MycoAlert Mycoplasma Detection Kit (Lonza, LT07-318), according to the manufacturer's instructions.

2.1.2 | iPSC culture

Induced pluripotent stem (iPS) cells used were available in the Nussenzweig group and described in detail previously[759,760]. Low-passage iPS cells were cultured in 10 ml sterile-filtered Essential 8 Medium (E8, A1517001, including E8 supplement) in a Matrigel hESC-Qualified Matrix (Corning, 354277)-coated 10 cm dish. Routine passaging was done at 80% confluency; cells were washed with Dulbecco's phosphate-buffered saline (DPBS; Corning, 21-031-CV) prior to Accutase (Innovative Cell Technologies, AT104) treatment for 5-7 min at 37 °C. Accutase was quenched with 5 ml of DPBS and collected in a 15 ml conical tube; cells were rinsed with another 5 ml of DPBS before centrifugation at 300 *g*, 5 min. The supernatant was aspirated, and cells were resuspended in 5 ml DPBS and counted using acridine orange and propidium iodide (AOPI; VWR, 89139-110

& 89411-112) staining. 5-10$\times10^6$ cells were seeded on a Matrigel-coated 10 cm dish with 10 ml of E8 containing 10 nM Y-27632 dihydrochloride (ROCK inhibitor, ROCKi; Selleckchem, S1049). Media containing ROCKi was aspirated the next day and replaced with E8; E8 was changed daily.

2.1.2.1 | iPSC differentiation into i^{3}Neurons

iPS cell differentiation was followed as previously described[728], with an outline provided here. On day 0, 20×10^6 iPS cells were plated onto a 15 cm plate in N2 medium (KnockOut DMEM/F12; Gibco, 12660) containing N2 supplement (Gibco, 17502), 1$\times$ GlutaMAX (Gibco, 35050), 1$\times$ MEM non-essential amino acids (NEAA; Gibco, 11140), 10 µM ROCKi, and 2 µg/ml doxycycline (Clontech, 631311), with medium replaced each of the next two days. On day 3, cells were re-plated on dishes coated with 0.1 mg/ml poly-L-ornithine (PLO; Sigma-Aldrich, P3655) in i^{3}Neuron medium: BrainPhys medium (STEMCELL Technologies, 05790), supplemented with 1$\times$ B27 Supplement (Gibco, A3582801), 10 ng/ml BDNF (PeproTech, 450-02), 10 ng/ml NT-3 (PeproTech, 450-03), 1 mg/ml mouse laminin (Sigma-Aldrich, L2020), and 2 µg/ml doxycycline. Half-medium changes were carried out on day 6, before the collection of cell samples on day 7.

2.1.3 | Insect cells

Spodoptera frugiperda (Sf9) and *Trichoplusia ni* (BTI-TN-5B1-4, Hi5) were cultured in Sf-900 II serum-free medium (Gibco, 10902) and ExpressFive serum-free medium (Gibco, 10486), at 27 °C in ambient CO_2, respectively.

2.1.4 | Bacterial strains

For screening of plasmids generated by all cloning methods, besides Gateway cloning, Stellar Competent *E. coli* (TAKARA, 636763), MAX Efficiency DH5α Competent *E. coli* (DH5α; Invitrogen, 18258012) and 10-beta Competent *E. coli* (10-beta; NEB, C3019) were transformed. Routine plasmid propagation was achieved by transforming 10-beta or DH5α. Gateway-cloned vectors were transformed into One Shot *ccd*B Survival 2 T1^R Competent Cells (Invitrogen, A10460). For the generation of bacmids, MAX Efficiency DH10Bac Competent *E. coli* (Gibco, 10361012) or DH10EMBacY Competent *E. coli*

(Geneva Biotech) were transformed. BL21 DE3 pLysS *E. coli* (Invitrogen, C6060) was transformed for the expression of recombinant proteins in bacteria (Section 2.3.3.3).

2.2 | Molecular biology

2.2.1 | Restriction digestion and ligation cloning

PCR amplification of an insert was performed to add unique restriction enzyme sites to either end so that they contain compatible sticky ends with the target vector. The target vector was either amplified by PCR, adding these same restriction enzyme sites or digested with the two restriction enzymes if the sites were already present. Following digestion with these enzymes and gel-purification, the cut insert was mixed with the linearised target vector (after thermosensitive alkaline phosphatase (TSAP; Promega, M9910) treatment), and ligated using T4 DNA ligase (TAKARA, 2011B) according to manufacturer's instructions, before transformation.

2.2.2 | InFusion HD cloning

PCR amplification of a target was performed to add flanking regions with 20 bp overlap with the target vector, which was amplified by PCR such that these homology regions would be at the ends of the linearised vector. This was sometimes also achieved through digestion with restriction enzyme(s). Both amplified/cut insert/vector of the expected size were excised from a 2% agarose gel, purified using Monarch kit Monarch DNA Gel Extraction Kit (NEB, T1020) and then combined with InFusion Snap Assembly (TAKARA, 638947), as per manufacturer's instructions.

2.2.3 | Golden Gate Assembly

For CRISPR-Cas9 sgRNA plasmids: pX458, pX458-mRuby2, pX459; sgRNAs were introduced by BbsI-targeted Golden Gate assembly. For chosen sgRNA sequences copied as DNA, a G was placed at the start unless the sgRNA already consisted of a G at the 5′ end. Oligonucleotides (sequences outlined below) were ordered with 5′ phosphates and annealed:

5′	CACC	*G*	sgRNA-sequence	
3′		*C*	sgRNA-complement	CCCA

Golden Gate reactions were performed by mixing annealed oligonucleotides with pX458 or pX459 vectors alongside T4 DNA ligase (NEB, M0202) and BbsI (NEB, R3539) and incubated for 10 cycles of 5 min at 37 °C and 10 min at 16 °C, before a final round of 37 °C for 1 h with the addition of fresh BbsI. Products were transformed into 10-beta Competent *E. coli*.

2.2.4 | Gateway cloning

An LR recombination reaction was performed by mixing 100 ng of the attL-containing entry clone plasmids and 300 ng attR-containing destination vectors (pCDNA6.2-V5-PL-DEST) and incubating with LR Clonase II mix (ThermoFisher Scientific, 11791) as per manufacturer's instructions.

2.2.5 | Site-directed mutagenesis

Site-directed mutagenesis was performed with long-range PCR reactions using CloneAmp HiFi PCR Premix (TAKARA, 639298) as per manufacturer's instructions. PCR products were run on a 2% agarose gel, excised and purified using Monarch® DNA Gel Extraction Kit, then treated with KLD enzyme mix (NEB, M0554) and transformed into 10-beta Competent *E. coli*.

2.2.6 | Plasmids and cloning protocols

Plasmids were cloned as detailed below, and all plasmids used in this study are summarised in Table A.1. In general, plasmids were either cloned through the design of synthetic genes within an intermediate vector, or as gene fragments, which were then PCR-amplified with homology to a destination vector and finally InFusion-cloned using these homology arms. Unless otherwise stated above, cloning reaction products were transformed into DH5α Competent *E. coli* or 10-beta Competent *E. coli*, with colonies selected on appropriate antibiotics: 50 µg/ml kanamycin (Sigma-Aldrich, K4000), 100 µg/ml carbenicillin (ThermoFisher Scientific, 10438923), 20 µg/ml gentamicin (Sigma-Aldrich, G1264), 50 µg/ml spectinomycin (Sigma-Aldrich, S4014), 10 µg/ml tetracycline (Sigma-Aldrich, T7660). Plasmids were extracted from colonies by miniprep (Qiagen, 27104) and sequence-verified

by Sanger sequencing/whole-plasmid nanopore sequencing (Source BioScience). Where required, plasmids were isolated from 200 ml LB cultures using the GenElute HP Plasmid DNA Maxiprep Kit (Sigma-Aldrich, NA0310).

2.2.7 | Insect cells plasmids

2.2.7.1 | Chimaeric SMX plasmids

The cDNA sequences for *X. laevis* (x) SLX4, xXPF, *H. sapiens* (h) ERCC1 were previously cloned into pDONR, meanwhile hMUS81, hEME1, xSLX1 was cloned into pAceBac1 (gifts of Dr. Denis Ptchelkine) and used to generate pKL-PBac:χSXE. Firstly, a platform was designed and synthesised into pUC57mini (GenScript) that consisted of TEV protease cleavage sites, separated by restriction enzyme sites, SalI-SpeI, XbaI-SphI, BamHI-XhoI, and flanked by BstEII-RsrII. From the above plasmids, the xSLX4 sequence was amplified, including Sal1-SpeI sites, xXPF sequence was amplified, including XbaI-SphI, and hERCC1 was amplified, including BamHI-XhoI and all sequentially cloned into the pKL-PBac backbone by restriction enzyme digestion-ligation cloning. pAceBac1:χEMS was previously cloned by Dr. Denis Ptchelkine.

2.2.7.2 | Mouse SMX plasmids

Peptide sequences for *M. musculus* SLX4, XPF, ERCC1, MUS81, EME1, and SLX1 were reverse-translated, and codon-optimised (SnapGene) for *S. fruigiperda* and constructed as synthetic genes (GenScript) to be cloned by InFusion into pAceBac1 containing an N-terminal tobacco-etch virus (TEV) protease (TEVP) sequence (gift from Dr. Abimael Cruz-Migoni). This cloning was achieved by including pAceBac1 homologous regions flanking the synthetic genes using primers 81-82 to form pAceBac1:mSXE (SLX4-XPF-ERCC1) and pAceBac1:mEMS (EME1-MUS81-SLX1).

Plasmid pFastBac-Dual:mSXE,EMS was cloned as follows: Plasmid pAceBac1:mSXE was PCR-amplified using primers 265-266 and pFastBac-Dual was PCR-amplified using primers 261-262 and the two were combined alongside a synthetic gene encompassing a Halo-tag with homology to immediately upstream of mXPF (GenScript) by InFusion cloning. This intermediate was amplified with primers 263-264 and combined with pAceBac1:mEMS amplified using primers 270-271 by InFusion cloning. This generated pFastBac-Dual:mSXE,EMS.

Variations to mouse SMX plasmids SLX4's central disordered region was predicted based on sequence alignments and domain mapping to encompass 759-1252 (Figure 1.7), which was deleted in pAceBac1 by site-directed mutagenesis using primers 75-76 to form pAceBac1:mSΔXE.

Nuclease-inactive mutants XPF^{D706A}, MUS81^{D307A}, and SLX1^{R38A,E79A} were made through site-directed mutagenesis using primers 39-40, 112-113, 43-44 (for R38A) and 45-46 (for E79A), respectively (yielding pAceBac1:mSX^{D706A}E and pAceBac1:mSΔX^{D706A}E, pAceBac1:mEM^{D307A}S and pAceBac1:mEMS^{R38A,E79A}). Plasmid pAceBac1:mEM^{D307A}S^{R38A,E79A} was generated from pAceBac1:mEMS^{R38A,E79A} using primers 112-113.

2.2.7.3 | Human SMX plasmids

Plasmids with a pTLCV2 backbone, containing fluorescent markers, were provided as a gift by Prof. Ying Dong - these include the fluorescent markers TagBFP, mCherry, and EGFP, all downstream of an IRES2 sequence that follows a multiple cloning site (MCS). pTLCV2:MCS-mOrange2 was generated by synthesising a mOrange2 (*H. sapiens* codon-optimised) gene fragment (Twist Bioscience) containing overhangs with homology to pTLCV2:MCS-mCherry, which was amplified by PCR using primers 253-254, and combined together by InFusion cloning. Similarly, pTLCV2:MCS-BirA-TagBFP was generated by synthesising BirA (*H. sapiens* codon-optimised) gene fragment (Twist Bioscience) containing overhangs with homology to pTLCV2:MCS-TagBFP, which was amplified by PCR using primers 255-256, and combined by InFusion cloning.

Where gene fragments were used (for the hSLX4-hSLX1, hMUS81-hEME1, and hXPF-hERCC1 constructs below), the full desired insert protein sequence was reverse-translated *in silico* with *H. sapiens* codon-optimisation, and the resulting DNA sequence was split up into fragments <1800 bp if the required recombinant sequence was longer than the maximum available synthesising fragment by Twist (>1800 bp). These split recombinant sequences each included 15-20 bp homology to the next fragment, and the first and final fragments consisted of 20 bp homology to the appropriate destination vector. All target vectors were amplified by PCR using primers 257-258 and inserts combined by InFusion cloning:

1. pTLCV2:hXE,BirA-TagBFP

- hXPF-P2A-hERCC1-3c-AviTag-t-His_6 fragments were cloned into pTLCV2:MCS-BirA-TagBFP

2. pTLCV2:hSS-EGFP

   - SF-t-hSLX4 fragments were cloned into pTLCV2:MCS-EGFP
   - hSLX1 fragment was cloned into this intermediate plasmid

3. pTLCV2:hME,mCherry

   - hMUS81-T2A-HA-HaloTag-t-hEME1 fragments were cloned into pTLCV2:MCS-mCherry

4. pTLCV2:SLX4IP-mOrange2

   - A His_6-SUMO-hSLX4IP fragment was cloned into pTLCV2:MCS-mOrange2

For transient overexpression plasmids, synthetic genes encoding hMUS81-T2A-HA-HaloTag-hEME1 were re-synthesised and cloned into pTwistCMV-Puro by Twist Bioscience to form pTwistCMV-Puro:hME. To generate pCDNA3.1-Hygro:hSS, the sequence spanning SF-t-hSLX4-IRES2-SLX1-t-EGFP from pTLCV2:hSS-EGFP was subcloned into pCDNA3.1-Hygro (GenScript). For pCDNA3.1-Zeo:hXE, the sequence spanning hXPF-P2A-hERCC1-3c-AviTag-t-His_6 was subcloned into pCDNA3.1-Zeo (GenScript). Mutations D705A in XPF, D706A in MUS81, and E41A;R82A in SLX1 were generated by the site-directed mutagenesis service provided by GenScript.

2.2.7.4 | Bacterial cell plasmids

GFP and mCherry nanobody plasmids Peptide sequences for the GFP-binding fragment of a single-chain camelid antibody (nanobody)[761], and an mCherry-binding nanobody[762] were reverse-translated with *E. coli* codon optimisation, and cloned from a synthetic gene into pET28a (Twist bioscience) using NcoI and XhoI for restriction digest-ligation cloning to generate pET28a:GFP-Nb and pET28a:mCherry-Nb, respectively.

2.2.7.5 | Human cells plasmids

CRISPR-Cas9 and sgRNA plasmids CRISPR-Cas9 knockout of *ERCC4* required plasmids: pX458:sg*ERCC4*-1 and pX458-mRuby2:sg*ERCC4*-2. These were generated by Golden Gate assembly using BbsI, to introduce sgRNAs: `CAACACGCAGCCGGCCGAGG`, and `GTACTCCAGCAGCGGCGCCA`, respectively, which, together, directed partial deletion of *ERCC4* exon 1, resulting in a frameshift. Note: cloning of sgRNAs into these plasmids was performed by Dr. Philip Hublitz.

For CRISPR-Cas9 knockout of *SLX1*, plasmids below were generated through BbsI-dependent Golden Gate assembly of pX458 with sgRNA sequences shown:

1. pX458:sg*SLX1*-ex1-1: `GTACCGGGGCCGCGTCTACG`
2. pX458:sg*SLX1*-ex1-2: `CGTAGACGCGGCCCCGGTAC`
3. pX458:sg*SLX1*-int1-1: `CATCTCCGCGGCGGAACTCA`
4. pX458:sg*SLX1*-int1-2: `GGGCTGCGACAGGGACATCA`
5. pX458-mRuby:sg*SLX1*-int3-1: `AGAGCCATCCATTCCTCGGG`
6. pX458-mRuby:sg*SLX1*-int3-2: `TAGAGCCATCCATTCCTCGG`

For CRISPR-Cas9 knock-in at the N-terminus of *SLX1*, the plasmid pX459:sg*SLX1*-N1 was required, wherein a sgRNA targeting the start codon (`AAGGCCCGATGGGTCCCGCG`) was introduced by Golden Gate assembly. Four donor sequences were designed in this study - each included the same 750 bp homology on either side of the *SLX1* (LH and RH donor sequences). Donor 1, pMA:*SLX1*donor1, designed to add EGFP-P2A-FKBP-TwinStrep to the *SLX1* reading frame, made use of a *H. sapiens* codon-optimised sequence that was synthesised and cloned into pMA (Invitrogen). Donor 2, pCDNA6.2-V5-PL-DEST:*SLX1*donor2, designed to add BleoR-P2A-EGFP-T2A-FKBP-SF-t to the *SLX1* reading frame, made use of a *H. sapiens* codon-optimised sequence that was synthesised and cloned into pTwistENTR (Twist Bioscience), which was then Gateway cloned into pCDNA6.2-V5-PL-DEST. Donor 3, pCI-Neo:*SLX1*donor3, designed to add BleoR-EGFP-P2A-HA-HA-FKBP to the *SLX1* reading frame, made use of a *H. sapiens* codon-optimised sequence that was synthesised and cloned into pCI-Neo using XbaI and MfeI, so that the

CMV promoter was lost (GenScript). Donor 4, pcDNA3.1-Zeo:*SLX1*donor4, designed to add HA-HA-3×FLAG-HiBiT-HaloTag-t to the *SLX1* reading frame, made use of a *H. sapiens* codon-optimised sequence that was synthesised and cloned into pCDNA3.1-Zeo, in a reverse orientation to the CMV promoter (Invitrogen).

Plasmids for immunoprecipitation pCDNA5-FRTTO expressing EGFP-hSLX4 (pCDNA5-FRTTO:EGFP-hSLX4), and their derived truncation mutants (pCDNA5-FRTTO:EGFP-hSLX4(D1), pCDNA5-FRTTO:EGFP-hSLX4(D2), pCDNA5-FRTTO:EGFP-hSLX4(D3), pCDNA5-FRTTO:EGFP-hSLX4(D4), , pCDNA5-FRTTO:EGFP-hSLX4(ΔSAP), pCDNA5-FRTTO:EGFP-hSLX4(ΔSBD)) were provided as a gift by Dr. Kaima Tsukada. pTLCV2 expressing EGFP-hSLX4 doxycycline-inducible control (pTLCV2:EGFP-hSLX4) was also provided as a gift by Dr. Kaima Tsukada.

CRISPRf plasmids Plasmids for the site-specific generation of formaldehyde (Section 3.5) were generated as follows from the pLenti-tetON:KRAB-dCas9-DHFR backbone[763]. Sequences for *H. sapiens* ADH1B and ADH1G were codon-optimised for *H. sapiens* cell expression and synthesised with the inclusion of 20 bp homology regions at either end, and cloned into pUC57mini (GenScript). The synthetic genes were PCR-amplified from these plasmids using primers 240-241 and 242-243, for ADH1B and ADH1G, respectively and gel-purified. For LSD1-dCas9 fusion, LSD1 from plasmid pX330a:dCas9-LSD1[764] was PCR-amplified using primers 244-245 and gel-purified. The backbone vector, pLenti-tetON:KRAB-dCas9-DHFR, was PCR-amplified using primers 238-239 and gel-purified. Each insert, (ADH1B, ADH1G, or LSD1) was mixed with the amplified vector and cloned *via* InFusion cloning to generate pLenti-tetON:ADH1B-dCas9-DHFR, pLenti-tetON:ADH1G-dCas9-DHFR, and pLenti-tetON:LSD1-dCas9-DHFR, respectively.

Table 2.1 | **Primer Sequences Used in the Study.**

ID	Long Name	Type	Sequence
39	MmXPF-D706A-fw	Mutagenesis	CGT GGT ATC GCT ATC GAG CCT GTG
40	MmXPF-D706A-rv	Mutagenesis	ACG GTG GAT CAG GGA AGG

Continued on the next page...

ID	Long Name	Type	Sequence
43	MmSLX1-R38A-fw	Mutagenesis	CCC TGC TCG TGC TGT GCG TCA GC
44	MmSLX1-R38A-rv	Mutagenesis	TTC ACG GTG AAA CCC ACG
45	MmSLX1-E79A-fw	Mutagenesis	CTG CGT TTC GCT TGG GCT TGG CAG CAC
46	MmSLX1-E79A-rv	Mutagenesis	AGC AGC CAC AGC GGA AGG
75	P21_SLX4_del759-1253_fw1	Mutagenesis	CAC TCC CCT GGT CTG GGT ACC ACC
76	P21_SLX4_del759-1253_rv1	Mutagenesis	GTC CTG AGC CTC AGC CTT ACC AGC
81	SXEsyn-InF_rv	Mutagenesis	GTA GGC CTT TGA ATT TTA ACG AGG CAC CTT CAG GAA AGG C
82	SXEsyn-InF_fw	Mutagenesis	CCT GAC TGG GCT GGT AAG GCT G
112	MmMUS81-D307A-fw2	Mutagenesis	CTT TCG TGT GGG TGG CTC AGG AG
113	MmMUS81-D307A-rv2	Mutagenesis	CAC CCA CGT GCA GCT TAC GC
261	pFD_RS1-2_f	Mutagenesis	CCT ACG TCG ACG AGC TCA CTA G
262	pFD_RS1-2_r	Mutagenesis	CCT TTG AAT TCC GCG CGC
263	pFD_RS3-4_f	Mutagenesis	CTG CTA GCA CCA TGG CTC G
264	pFD_RS3-4_r	Mutagenesis	CTG ATG CAT AGC ATG CGG TAC CG
265	pA1_SX_InF-pFD_f	Mutagenesis	GTC CGA AGC GCG CGG AAT TCA AAG GAT GGG AGG CGA ATC CCT CTT CAA AG
266	pA1_SX_InF-pFD_r	Mutagenesis	CCT TGG ACA CCA TAG AGT CGA CAC CGG AAC
270	pA1_TEVP_EMS_InF-pFD_r	Mutagenesis	CCC GGT ACC GCA TGC TAT GCA TCA GCT TTG AAT TCT TAT TAA TGG TGA TGG TGA TGG TGA
271	pA1_EMS_InF-pFD_f	Mutagenesis	GGA TCT CGA GCC ATG GTG CTA GCA GAT GGC TCT GCG TCG TC
182	BOLA2-f1	KO Screening	GCG GCT CAA CTC ACC ACA T
185	SLX1-ex2-r1	KO Screening	GAA GCC GTG CAC GAC GAG

Continued on the next page...

ID	Long Name	Type	Sequence
204	ADH5ko_5-f1	KO Screening	GGC CAA ACC CCT CAT TTT AGC
206	ADH5ko_3-r1	KO Screening	TTG TGG CAT TTC AAC CGG TT
208	ALDH2ko_5-f1	KO Screening	TGA CCT TGG TCC ATT TCC CA
210	ALDH2ko_3-r1	KO Screening	ACA AGC TGA AGA GGA GGG AC
246	CRISPRi_ki_f1	KO Screening	CCG TAC CAC TTC CTA CCC TC
247	CRISPRi_ki_r1	KO Screening	CGT TGG GCA GGT TCT TAT CG
319	V3_2_F	Library	CTG CGT GCG CCA ATT CTG
320	V3_1_R2	Library	AGA ACC GGT CCT GTG TTC TG
322	D706_R	Library	CAA GCA GAA GAC GGC ATA CGA GAT ACG AAT TCG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT ACT TGC TAT TTC TAG CTC TAA AAC
323	D705_R	Library	CAA GCA GAA GAC GGC ATA CGA GAT TTC TGA ATG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT ACT TGC TAT TTC TAG CTC TAA AAC
325	D702_R	Library	CAA GCA GAA GAC GGC ATA CGA GAT TCT CCG GAG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT ACT TGC TAT TTC TAG CTC TAA AAC
326	D701_R	Library	CAA GCA GAA GAC GGC ATA CGA GAT CGA GTA ATG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT ACT TGC TAT TTC TAG CTC TAA AAC
329	D504_F	Library	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG GCT CTG AAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TGC ATT GTG GAA AGG ACG AAA CAC CG

Continued on the next page...

ID	Long Name	Type	Sequence
330	D503_F	Library	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACC CTA TCC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCG TTG TGG AAA GGA CGA AAC ACC G
331	D502_F	Library	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACA TAG AGG CAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAT TGT GGA AAG GAC GAA ACA CCG
332	D501_F	Library	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT ATA GCC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TTT GTG GAA AGG ACG AAA CAC CG

2.3 | Cell biology methods

2.3.1 | Generation of cell lines

2.3.1.1 | Generation of *ERCC4* knockout cell lines

3×10^5 A549, HT-29, and U2OS cells were seeded in each well of a 12-well plate. The next day, 2.7 µg of pX458:sg*ERCC4*-ex1fw and pX458-mRuby:sg*ERCC4*-ex1rv were transfected by Lipofectamine 3000 (Invitrogen, L3000), as per manufacturer's instruction. Medium was replaced the following day, and cells were allowed to grow for another day. Cells were sorted using a FACSAria Fusion Flow Cytometer (FACS; BD Biosciences) for EGFP and mRuby2 expression into single wells of a 96-well plate and grown in medium supplemented with an additional 10% FBS. Once a sizable colony had grown, cells were expanded to a 24-well plate, then a T75 dish, before freezing (freezing stocks were made by passage 4 at the latest). From the initial 96-well plate, genomic DNA was extracted using QuickExtract DNA Extraction Solution (Lucigen, QE09050) as per the manufacturer's instructions. PCR was performed with primers 276-277 using MyTaq HS Red Mix (Bioline, BIO-25048), with cycling conditions: 95 °C 2 min, 30× cycles

of 95 °C 15 s, 60 °C 15 s, 72 °C 30 s (for 2 kb product), and a final extension of 72 °C 2 min. Products of clones showing a putative knockout band were purified using a QIAquick PCR Purification Kit (Qiagen, 28104), and verified by Sanger sequencing (Source BioScience). Western blotting was performed to verify the loss of expression of XPF. Note: HT-29 and A549 Δ*ERCC4* cells were generated by Dr. Ceren Yalaz. U2OS Δ*ERCC4* cells were generated by Dr. Sanja Brolih.

2.3.1.2 | Generation of *SLX1* knockout cell lines

4×10^5 RPE-1, RPE-1 *TP53*$^{-/-}$, HAP1 and 293FT cells were seeded in each well of a 6-well plate with 2 ml of medium. The next day, 1 µg of pX458:sg*SLX1*-int1-1 and pX458-mRuby2:sg*SLX1*-int3-1 were transfected using Lipofectamine 3000. Media was replaced 24 h later, and cells were sorted two days later by FACS for EGFP and mRuby2 expression and plated as single cells into a 96-well plate containing media with 20% FBS. As above, clones were expanded, genomic DNA was extracted for PCR (using primers 334-336), and PCR-positive clones were maintained and frozen by passage 4. For confirmation by Western blotting, cells were harvested from a confluent 150 mm dish and either lysed *via* RIPA or processed for nuclear extracts and probed for SLX1.

2.3.1.3 | Generation of *SLX1* knock-in cell lines

3×10^5 RPE-1 cells were seeded in each well of a 6-well plate. The following day, 2.7 µg of pX459:sg*SLX1*-N1, or pX459 (control) alongside 2.7 µg of pTwistENTR donor sequence were transfected by Lipofectamine 3000. For the transfection of a sgRNA:Cas9 ribonucleoprotein (RNP), chemically protected sg*SLX1*-N (`AAGGCCCGAUGGGUCCCGCG`, Synthego) and SpCas9 2NLS (Synthego) were transfected by Lipofectamine CRISPRMAX Cas9 transfection reagent (Invitrogen, CMAX00). Medium was replaced 24 h later and supplemented with 10 µg/ml puromycin (Gibco, A1113803) and 3 µM DNA-PKi (Nu7441; SelleckChem, KU-57788) to promote homology-directed repair events; the high concentration of puromycin required for selection was due to partial resistance of RPE-1 cells to puromycin through their hTERT-immortalisation, which was transduced alongside a hygromycin (partial redundancy with puromycin) resistance-containing pGRN15 plasmid[765,766]. After 4 days of puromycin selection, the media was replaced without

puromycin but containing a starting dose of 100 µg/ml zeocin. After a further 3 days under minimal zeocin selection, the media was replaced with 200 µg/ml zeocin. After another week of zeocin selection (media containing 200 µg/ml zeocin replenished every 3 days), cells were sorted by FACS for GFP expression into single wells of a 96-well plate and grown as clones. Colonies from wells were expanded as above (Section 2.3.1.1) and frozen by passage 4. Genomic DNA was extracted as in (Section 2.3.1.1); with PCR performed using primers 182-185 or 284-285 (longer-range). PCR reactions containing a band corresponding to the expected knock-in size were purified using QIAquick PCR Purification Kit and verified by Sanger sequencing. To confirm knock-in by Western blotting, an 80% confluent well of a 6-well plate was grown overnight with or without dTag-13 (Sigma-Aldrich, SML2601), and expression of SLX1 was visualised by FLAG immunoblot.

2.3.1.4 | Generation of CRISPRi cell lines

For generation of CRISPRi cell lines (dox-inducibly expressing KRAB-dCas9) in HT-29 and HT-29 Δ*ERCC4*, 1 ml of lentivirus expressing dCas9-KRAB (addgene: 167935) was added to 5×10^6 cells, seeded in a 6-well plate 24 h prior. Media containing lentivirus was replaced with DMEM + 10% FBS + P/S, and cells returned to 37 °C for two further days. The resulting population was single-cell sorted for RFP^+ cells and grown in DMEM + 20% FBS + P/S until colonies formed. Colonies were then expanded and simultaneously split for genomic DNA extraction, where PCRs were performed using primers 246-247, to verify the presence of the dCas9-KRAB construct. PCR-positive clones were then confirmed by Western blotting for HA after overnight treatment with 2 µg/ml doxycycline and 50 nM trimethoprim (TMP).

2.3.1.5 | Generation of *ADH5*$^{-/-}$ and *ALDH2*$^{-/-}$ iPSCs

For the generation of *ADH5*$^{-/-}$ or *ALDH2*$^{-/-}$ iPSCs, 90 pmol of sgRNAs (chemically protected, from Synthego, Table 2.2) were mixed with 80 pmol Cas9 protein (gift of Dr. Raj Chari), and the RNP reaction was incubated at room temperature for 10 min and then stored on ice until electroporation. iPSCs were lifted from 10 cm dishes as per routine passaging (Section 2.1.2), and 5×10^5 cells per nucleofection were pelleted by centrifugation (300 *g*, 5 min, room temperature) and the supernatant removed. Cell pellets were resuspended

in 5 ml DMEM + F12 (Gibco, 11320), centrifuged, cell pellet washed twice with 5 ml OptiMEM. This final cell pellet was resuspended in 100 µl OptiMEM and mixed with the RNP mix as above and followed by careful transfer to a nucleofection cuvette. Cells were electroporated using an Amaxa Nuclefector II machine, program X-001, and the electroporated cells were immediately seeded on a 6-well plate pre-coated with Matrigel and containing 2 ml of E8 + ROCKi at a density of 1000 cells per well. The remaining cells from the electroporated mix were seeded in another Matrigel-coated well to account for the editing efficiency of the total population. ROCKi was removed by replacing the media with E8 the next day. Single-cell-derived colonies were picked and expanded into a 12-well plate, then a 6-well plate and finally a 10 cm dish before freezing using CELLBANKER® 1 Serum Containing Freezing Media (amsbio, 11888). Genomic DNA was extracted from clones using the DNeasy Blood & Tissue Kit (Qiagen, 69504) and screened by PCR using primers 204-206 and 208-210 for *ADH5* and *ALDH2*, respectively, for the desired genomic deletion and verified by Sanger sequencing (ACGT).

ADH5	sg*ADH5*-1	AUCCAUCGAAUUACCUCCAA
	sg*ADH5*-2	UGGUCUGGGAGGAGUCGGAU
ALDH2	sg*ALDH2*-1	AUCCUCGGCUACAUCAACAC
	sg*ALDH2*-2	UCCCCGUGUUGAUGUAGCCG

Table 2.2 | Sequences of sgRNAs for generation of *ADH5* and *ALDH2* knockouts in iPS cells.

2.3.2 | Cell culture techniques

2.3.2.1 | Transient transfection

Plasmids were transfected using Lipofectamine 2000 (Invitrogen, 11668027) as per the manufacturer's instructions.

2.3.2.2 | Lentivirus production and stable cell generation

For the assembly of lentiviruses, plasmids pHDM-Hgpm2 (SV40 GagPol, Addgene: 164441), pHDM-VSV-G (USVG, Addgene: 164440), pRC-CMV-Rev1b (Rev, Addgene: 164443), pHDM-Tat1b (Tat, Addgene: 164442) and the lentiviral transfer vector were used to transfect, at a ratio of 1:1:2:1:5, 4×10^6 293FT cells in 10 cm dishes, using Lipofectamine

2000 (Invitrogen, 11668027) as per manufacturer's instruction. Cell culture medium was replaced 4 h after transfection with media containing P/S. Viral supernatants were harvested 48 h after transfection, filtered through a 0.45 µm filter (ThermoFisher Scientific, 15216869), and used immediately for transduction or frozen and stored at -80 °C until use.

For transduction, cells were plated to reach 50% confluency the next day, without P/S, in 10 cm culture dishes and 1 ml of viral supernatant was added. Following another 24 h, culture media was replaced with media containing P/S, and selected with antibiotics, puromycin (Gibco, A11138) or G418 sulfate (Gibco, 10131035), where appropriate. FACS was used to select cells for EGFP or RFP expression, and sorted into a heterogeneous stable population or as single cells into wells of a 96-well plate.

2.3.2.3 | Freezing cell stocks

For the cryopreservation of cell lines, trypsinised cells were resuspended in freezing media (90% FBS, 10% DMSO) and transferred to cryotubes. Cells were slowly frozen in a pool of isopropanol to -80 °C and transferred to a liquid nitrogen tank for longer-term storage. To defrost cell lines, cryotubes were thawed quickly, centrifuged for 5 min, 500 *g* to remove DMSO, and plated in a T25 or T75 flask with fresh culture medium.

2.3.2.4 | Genetic screens using CRISPR-Cas9

120×10^6 HT-29 and HT-29 Δ*ERCC*4 cells, seeded at 10×10^6 cells per 15 cm dish, were independently transduced with the lentiviral TKOv3 library[749,755] at an MOI of 0.3, to initiate a library coverage of >500×. The TKOv3 library contains 71090 sgRNAs targeting 18056 genes, with approximately five sgRNAs per gene. Puromycin was added the next day, to a final concentration of 2 µg/ml, for the selection of positive transductants, which was maintained for a further two days; this was considered the initial time point (t0). Cells were then subcultured every 3-5 days until t22 (day 22), maintaining a library coverage of >500× and collecting 35×10^6 cells (t3, t6, t9, t14, t18). For all screens, SJG-136 treatments were applied according to the previous evaluation of LD_{20} concentration in uninfected HT-29 cells treated for 3 days. A total of 35×10^6 cells were frozen at t22 for gDNA extraction, storing pellets at -80 °C. Screens were performed in biological duplicates. DNA was extracted and prepared for Illumina sequencing as described (Section 2.5.4).

2.3.2.5 | Cell synchronisation

Cells were arrested in the G2/M phase through the addition of 0.1 µg/ml (unless otherwise stated) nocodazole (Sigma-Aldrich, SML1665) for 16 h.

2.3.2.6 | RNA interference

Cells were seeded at a density of 2×10^6 cells/ml and treated immediately with 20 nM siRNA using Lipofectamine RNAiMAX (Invitrogen, 13778) according to the manufacturer's instructions. Cells were forward transfected with 20 nM siRNA the next day and harvested/reseeded 24-48 h later. The siRNAs used are described in Table 2.3:

siRNA	Description	Type	Catalog No.	Sequences
siMUS81	ON-TARGETplus Human MUS81 siRNA SMARTpool, Dharmacon	Pool	L-016143-01-0005	CAGCCCUGGUGGAUCGAUA, CAUUAAGUGUGGGCGUCUA, UGACCCACACGGUGCGCAA, CUCAGGAGCCCGAGUGAUA
siSLX1-1	ON-TARGETplus Human SLX1 siRNA SMARTpool, Dharmacon	Pool	L-034933-01-0005	GCUAAGGGCCCAUGUGAUC, CAGAAUUAGAAGAGGCACA, GGACACUGAGAAAGAAGUA, GCUGGAGACCUGAUCCUCA
siSLX1-2	Chemically protected RNA oligonucleotide, ThermoFisher Scientific	Single	N/A	UGGACAGACCUGCUGGAGAUU
siBRCA1	ON-TARGETplus Human BRCA1 siRNA SMARTpool, Dharmacon	Pool	L-003461-00-0020	CAACAUGCCCACAGAUCAA, CCAAAGCGAGCAAGAGAAU, UGAUAAAGCUCCAGCAGGA, GAAGGAGCUUUCAUCAUUC

Table 2.3 | siRNAs used in the study.

2.3.2.7 | Incucyte growth assays

Cells were seeded as technical replicates of 3000 cells per well of a 96-well plate. Images were acquired at 20× magnification every 3 h, and the resulting percentage confluencies were quantified using Incucyte ZOOM 2018A software (Essen Bioscience). Doubling times were calculated in GraphPad Prism V.8 using the exponential growth equation.

2.3.2.8 | Colony survival experiments

Unless otherwise stated, for each biological repeat, 1000 cells were seeded in a 10 cm dish in triplicate for each different condition. Drug concentrations and treatments are outlined below (Section 2.3.2.10). 10-14 days following treatment, the growth medium was removed, cells were washed with PBS, and stained with Coomassie staining solution (0.1% Coomassie, 7% acetic acid, 50% methanol) for 30 min, followed by washing with ddH_2O and air-drying before scanning. The number of colonies was counted using a ColCount colony counter (Oxford Optronix). Representative dishes were selected for visualisation in Figures 3.3c, 3.7a, 4.7c, 4.8c.

2.3.2.9 | Resazurin/CellTiter-Glo cell viability assays

For each biological repeat, 300 cells per well were seeded in a 96-well clear-well plate with n = 6 technical repeats for each condition (n = 12 for untreated controls). Drug concentrations and treatments are outlined below (Section 2.3.2.10). 8 days following treatment, growth medium was replaced with a phenol-red-free version of the medium supplemented with 10 µg/ml resazurin, and cells incubated for 2-4 h until untreated control wells became a pink colour. Relative fluorescence at Ex/Em 530-560/590 nm was measured using a CLARIOstar plate reader (BMG LABTECH).

For CellTiter-Glo 2.0 (CTG) cell viability assays, cells were instead seeded at a density of 500 cells per well in a 96-well white plate (655083, Greiner). After 8-10 days following treatment, the medium was replaced with CTG, diluted 1 in 10 in PBS, and plates were shaken vigorously for 10 min before luminescence was measured using a CLARIOstar plate reader.

2.3.2.10 | Drug treatments

In general, cells were treated 16 h following seeding for both colony survival experiments and resazurin/CTG assays. For siRNA-treated cells re-seeded for a survival experiment, drug treatment was completed within 6-8 h of re-seeding. Unless otherwise stated within figure legends, cell treatments were as follows for cell viability/clonogenic assays. Cisplatin (CDDP, Teva pharmaceutical industries Ltd., Eastbourne UK Cat 51642169), formulated as a stock at 3.3 mM, was applied for 4 h before medium was replaced for the remainder of the assay. Trabectedin (Cambridge Bioscience, 34662-1 mg-CAY), diluted to a stock of 1 mM in DMSO, was applied for 2 h before medium was replaced for the remainder of the assay. ADH5 inhibitor, N6022 (Cayman, 21269), diluted in DMSO to a stock concentration of 50 mM; doxycycline (ThermoFisher Scientific, D9891), diluted in DMSO to a stock concentration of 50 mg/ml; formaldehyde (ThermoFisher Scientific, 28906), diluted fresh for each use in PBS to a stock concentration of 5 mM; olaparib (SelleckChem, S1060), diluted in DMSO to a stock concentration of 10 mM; were all applied continuously for the duration of the assay.

2.3.3 | Protein expression

2.3.3.1 | Baculovirus generation and amplification

Bacmid DNA and baculoviruses for insect cell-expressing plasmids were generated as described[767]. In brief, bacmid DNA was prepared by transforming DH10Bac Competent *E. coli* with 0.1 µg of the respective plasmid DNA with backbone pAceBac1 or pKL-PBac. Transformants were selected through blue(black)-white screening on LB-agar plates containing kanamycin, tetracyline, and gentamicin (Section 2.2.6) with 0.3 g/l S-gal, 0.03 g/l isopropyl β-D-1-thiogalactopyranoside (IPTG), 0.5 g/l ammonium citrate. Alkaline lysis of bacteria was followed by isopropanol precipitation of bacmid DNA, which were washed in 70% ethanol and resuspended in TE (10 mM Tris-HCl (pH 8), 1 mM EDTA) and used immediately or stored for <1 month at -20 °C. P0 baculovirus was generated through transfection of 5×10^6 Sf9 cells with ~50 µg of bacmid DNA using Cellfectin II (Gibco, 10362) – transfections were performed in T25 flasks; 20 µl of bacmid DNA (concentration range 1-4 µg/µl) was diluted with 10 µl Cellfectin II in 1 ml of SF900 II media, incubated for 30 min, and then added dropwise to cells. The transfection mix

was removed 4 h later and replaced with SF900 II medium. P1 baculovirus was harvested 96 h post-transfection by centrifugation of the growth medium (700 *g*, 5 min, 4 °C) and supplementing with 10% FBS. To generate P1 virus, 15×10^6 Sf9 cells in a T75 flask were infected with 1 ml of P0 virus, and P2 virus was harvested as above 72 h later. This same procedure, but using 30×10^6 Sf9 cells in a T175 flask, was repeated to generate P2, and sometimes P3/P4 viruses. All baculoviruses were protected from light and stored at 4 °C for <2 months; P0 virus aliquots were stored at -80 °C for <1 year.

2.3.3.2 | Protein expression in insect cells

Sf9 or HiFive cells cultured in suspension were diluted to a density of 2×10^6 cells/ml, infected with 1/100 (v/v) of either P3 or P4 baculovirus and incubated for 48-72 h. Cells were harvested by centrifugation (1500 *g*, 10 min, 4 °C), washed with cold PBS, centrifuged again and snap-frozen in dry ice prior to storage at -80 °C until purification.

2.3.3.3 | Protein expression in bacteria

For expression of GFP and mCherry nanobodies, plasmids were transformed into BL21 DE3 pLysS (Invitrogen, C6060) transformed with either pET28a:GFP-Nb or pET28a:mCherry-Nb. The genotype of this strain is: *E. coli* F^- *omp*T *hsd*SB (r_{B^-}, m_{B^-}) *gal dcm* (DE3) pLysE(CamR). Starter cultures from several colonies were grown together in LB containing 50 µg/ml kanamycin overnight at 37 °C with shaking - glycerol stocks were generated the next day in 30% glycerol, stored at -80 °C. 10 ml of starter culture was used to inoculate 2 l of LB containing antibiotic and grown at 37 °C with shaking, until an OD_{600} of 0.6-0.8 was reached. Cultures were cooled to 4 °C for 30 min prior to the addition of 0.5 mM IPTG (Sigma-Aldrich, I6758) for expression overnight at 18 °C for 16-20 h. Cells were harvested by centrifugation 6,000 *g*, 30 min, 4 °C and pellets frozen in dry ice before storing at -80 °C. Frozen pellets were used for purification within a month.

2.3.3.4 | Protein expression in mammalian cells

Expi293F cells were seeded at a density of 1.2×10^6 cells/ml on day 0. The following day, when density was estimated to be $\sim2.4\times10^6$ cells/ml, PEI-transfection mix was formulated as follows: tube A containing 60 µg/ml polyethyleneimine (stock at 2 mg/ml

PEI, ThermoFisher Scientific, 043896.01) and 1/20th (of culture) volume of OptiMEM was mixed with 1 µg (per ml of culture) plasmid in the same volume of OptiMEM. This mixture was added dropwise to Expi293F cells while swirling and cells returned to the shaking incubator. On day 2, 3.5 mM valproic acid (Sigma-Aldrich, P4543), and 1/100 v/v NEAA (Gibco, 11140) was added. Expressions were harvested either on day 3 or day 4 by centrifugation at 700 *g* for 15 min, resuspension in PBS (Gibco, 10010), centrifugation again and freezing in dry ice and storing at -80 °C until use.

2.4 | Protein methods

2.4.1 | Protein purifications of SMX and SLX4-containing complexes

For all SMX preparations (and derivatives like SLX4-C, including the various species preparations), after expression in insect cells as above (Section 2.3.3.2), thawed cell pellets were resuspended in buffer MS-L (20 mM HEPES-KOH (pH 8), 750 mM NaCl, 1 mM $MgCl_2$, 0.05% CHAPS, 0.5 mM TCEP) supplemented with 1× cOmplete mini EDTA-free cocktail (Roche, 11836170001) and 1 mM PMSF (ThermoFisher Scientific, 36978). Cells were lysed by sonication while on ice (30 s on, 1 min off, 3 min total on, 30% amplitude; Branson SFX550) and cell debris cleared by centrifugation (48,250 *g*, 1 h, 4 °C). The supernatant was incubated for 2 h at 4 °C with Strep-Tactin Sepharose beads (IBA Lifesciences) or StrepTrap HP 5 mL (Cytiva) equilibrated with buffer MS-L. Beads were then washed with buffer MS-L for >10 bead volumes, and bound protein was eluted with buffer MS-L + 2.5 mM desthiobiotin (IBA Lifesciences, 2-1000-002). Eluant was concentrated, unless otherwise stated, and loaded onto a Superose 6 Increase 10/300 GL (Cytiva) in buffer G (20 mM HEPES-KOH (pH 8), 150 mM NaCl, 1 mM $MgCl_2$, 0.5 mM TCEP). Peak elutions were pooled and concentrated and final concentrations determined by Nanodrop/Bradford assay (Bio-Rad, 5000111), before freezing in liquid nitrogen or dry ice and storing at -80 °C. Some buffers were adjusted between different preparations but any changes are stated for each experiment.

2.4.1.1 | GFP and mCherry nanobody purification and conjugation to agarose beads

Pellets of BL21(DE3) cells overexpressing GFP-Nb and mCherry-Nb (Section 2.3.3.3) were thawed, resuspended in buffer LNB (50 mM sodium phosphate (pH 7.5), 300 mM NaCl, 10% glycerol, 2 mM $MgCl_2$, 1 mM PMSF, 25 U/ml benzonase (Millipore, 70746), 5 mM imidazole), lysed by sonication (40% amplitude, 5 min total, 10 s on/20 s off). Lysed cells were centrifuged at 47,500 *g*, 1 h, 4 °C, filtered through 0.45 µm cutoff filters and loaded onto TALON metal affinity resin (635653, TAKARA), equilibrated with buffer WNB (50 mM Na-phosphate (pH 7.5), 300 mM NaCl, 10% glycerol, 5 mM imidazole) and incubated with shaking for 1 h at 4 °C. Resin was washed with 10 bead volumes of buffer WNB and eluted with buffer WNB containing 160 mM imidazole. Eluted protein was concentrated and loaded onto a Superdex 75 Increase 10/300 GL (GE Healthcare), equilibrated with buffer NB-G (50 mM sodium phosphate (pH 7.5), 150 mM NaCl). Peak GFB-Nb or mCherry-Nb fractions were concentrated, flash-frozen in liquid nitrogen and stored at -80 °C.

Conjugation of purified nanobodies to sepharose beads GFP and mCherry nanobodies were immobilised on sepharose beads through NHS-sepharose coupling, according to the manufacturer's instruction. Briefly, NHS-activated sepharose 4FF (Cytiva, 170906) was washed with 1 mM HCl and then equilibrated with buffer NB-G. After draining the solution, 34 mg of nanobody was added in a final 15 ml volume of buffer NB-G and the beads/nanobody mix rocked for 4 h at 4 °C, then 2 h at room temperature, and finally overnight at 4 °C. Unbound nanobody was collected by adding extra buffer NB-G. The remaining resin was washed with blocking buffer (50 mM sodium phosphate (pH 7.5), 200 mM NaCl, 200 mM ethanolamine) and kept in blocking buffer for 24 h at 4 °C. The resin was then washed and stored in 20% ethanol at 4 °C until use.

2.4.2 | Protein analysis methods

2.4.2.1 | Flow cytometry

For cell cycle analysis, the cell culture medium was supplemented with EdU (Invitrogen, A10044) at 10 µM for the final 15 min of culture. Cells were then harvested by centrifugation at 700 *g*, 5 min, 4 °C (after trypsinisation where appropriate), washed with PBS

and centrifuged again, ensuring that only a maximum of 1×10^7 cells were processed. Cells were fixed by dropwise addition of ice-cold 70% ethanol while vortexing and either processed immediately or stored at -20 °C for up to 1 month. Cells were washed with 1% BSA in PBS (PBS-BSA), then permeabilised in 0.5% Triton X-100 in PBS for 20 min, and washed twice in PBS-BSA. Click reactions were performed for 45 min in the dark with 100 µl/sample of a cocktail made up of 1× Click-iT reaction buffer (Invitrogen, C10276), 4 mM CuSO4, 10 µM AF647-picolyl-azide (Jena Bioscience, CLK-1300) and 1× Click-iT additive (Invitrogen, C10276). The cocktail was removed by centrifugation and the cell pellet was washed in PBS, before adding 200 µl/sample of PBS containing 1 µg/ml 4,6-diamidino-2-phenylindole (DAPI; BD, 564907) and incubating a further 10 min in the dark. Flow cytometry data was acquired using an Attune NxT (ThermoFisher Scientific) flow cytometer and analysed using FlowJo 10.10.

2.4.2.2 | Cell lysis for IP/Western blotting

For harvesting a whole cell extract, cells were washed with PBS, harvested by trypsinisation, washed again with PBS and centrifuged (1500 *g*, 10 min, 4 °C). Cells were then resuspended in buffer NLB (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM $MgCl_2$, 10% glycerol, 5 mM NaF, 0.2% NP-40, 1× protease inhibitor cocktail) supplemented with 25 U/ml benzonase, incubated for 5 min, and an equal volume of buffer SLB (62.5 mM Tris-HCl (pH 7.5), 10% glycerol, 2% SDS, 25 mM TCEP) added. Samples were subsequently boiled for 5 min, and centrifuged (20,000 *g*, 10 min, 4 °C). The supernatant was used as a cell extract, with protein concentration estimated by Nanodrop.

To obtain nuclear and cytoplasmic extracts, cells were harvested as above. Cell pellets were then lysed by hypotonic lysis for 20 min at 4 °C in 5× pellet volume of buffer HLB (50 mM Tris-HCl (pH 7.5), 10 mM NaCl, 10% Glycerol, 1.5 mM $MgCl_2$, 0.5 mM TCEP, Halt Protease and Phosphatase inhibitor cocktail (ThermoFisher Scientific, 78440), 1 mM PMSF). The cytoplasmic fraction was extracted as the supernatant after centrifugation (2,000 *g*, 5 min, 4 °C) with the pellet washed once in cold PBS and centrifuged again. This nuclear pellet was resuspended in buffer NEB (50 mM Tris-HCl (pH 8), 10% Glycerol, 0.5 mM TCEP, Halt Protease and Phosphatase inhibitor cocktail (ThermoFisher Scientific, 78440), 0.5% NP-40, 10 units benzonase/100 µl buffer, 2 mM $MgCl_2$), incubated on ice

for 20 min, then NaCl added to 400 mM and incubated for a further 10 min on ice. Any remaining debris was then removed by centrifugation (20,000 *g*, 20 min, 4 °C).

All extracts were used immediately or stored at -20 °C for <1 month before use.

2.4.2.3 | Immunoprecipitations

For the preparation of lysates for immunoprecipitations, cell pellets were washed in PBS, and lysed in RIPA (ThermoFisher Scientific, 89900), supplemented with Halt Protease and Phosphatase inhibitor cocktail, 1 mM PMSF, and unless otherwise specified 25 U/ml benzonase, for 45 min on ice. Lysates were then cleared by centrifugation at 20,000 *g*, 10 min, 4 °C, collecting the supernatant. The protein concentration was determined by use of the Pierce BCA Protein Assay Kit (ThermoFisher Scientific, 23225) and lysate concentrations normalised by dilution in buffer IPbase (40 mM HEPES-KOH (pH 8), 5% glycerol, 2 mM $MgCl_2$, 0.5 mM TCEP, 0.05% NP-40) containing 150 mM NaCl. GFPTrap magnetic agarose beads (ChromoTek, GTMA) were washed once in PBS, then 1% BSA in PBS with end-to-end mixing for 10 min at room temperature, then washed twice in buffer IPbase + 150 mM NaCl. Normalised lysate was then incubated with 12.5 µl of washed GFPTrap beads for 2 h with end-to-end mixing at 4 °C. Bound beads were washed thrice in successive buffers of IPbase containing 150 mM, then 250 mM and finally 500 mM. Elution was performed in a mix of 1× SDS sample buffer (ThermoFisher Scientific, NP0007), with 1× reducing agent (ThermoFisher Scientific, NP0009) and boiling for 5 min before SDS-PAGE analysis. Immunoprecipitations using Halo-tag magnetic beads (Promega, G7281) and Strep-Tactin XT magnetic beads (IBA Lifesciences, 2-5090-002) were performed as above. Halo-tagged proteins do not themselves elute from the beads due to a covalent linkage, but other bound proteins can still be visualised. For *in vitro* biotinylation of Halo-tagged protein (Figure 5.13), normalised lysate in RIPA/buffer IPbase + 150 mM NaCl was incubated with HaloTag PEG Biotin Ligand (Promega, G8591) at a final concentration of 1 µM for 4 h at 4 °C, then added to Dynabeads MyOne Streptavidin C1 (Invitrogen, 65001) and processed as above.

2.4.2.4 | Immunofluorescence of dCas9-LSD1 cells

10^6 RPE-1 FRT *TP53*$^{-/-}$ dCas9-LDS1 cells were seeded in 10 cm Petri dishes in DMEM 10% FBS 1× PenStrep. 4 hours after seeding, the Cas9 system was induced with 2 µg/ml doxycycline and 50 nM TMP or non-induced with corresponding controls (DMSO, water). The following day, 24 hours after seeding, cells were nucleofected. Briefly, cells were harvested, counted and 1.5×10^6 cells were nucleofected in P3 solution and in the presence – or not – of 20 pmol of gRNA (sgTelomeres: `TAGGGTTAGGGTTAGGGTTA`, sg*PDXDC1* promoter: `TGGTTGTAGCGTCATCGCAG`, Synthego), using programme EA-104. Cells were then re-seeded in a 12-well ibidi slide (ibidi, 81201) (6000 cells per well) in inducing or non-inducing medium. 24 hours after nucleofection, cells were fixed with 4% (w/v) paraformaldehyde (PFA) in PBS for 10 min at RT and washed 3 times in PBS. PFA-fixed cells were permeabilised in 0.5% Triton X-100 in PBS for 5 min at RT and washed 3 times in PBS. Cells were then incubated with blocking buffer (10% goat serum, 2% BSA, in PBS) for 30 min at RT and then incubated for one hour at RT with γH2AX antibody (Millipore 05-636) diluted 1:6000 in blocking buffer. Cells were washed once in 0.1% Triton X-100 in PBS, twice in PBS, and incubated with anti-mouse Alexa Fluor 568 diluted 1:1000 in blocking buffer for 45 min at RT. After washing once in 0.1% Triton X-100 in PBS, twice in PBS, cells were incubated for 20 min at RT in DAPI (1 µg/ml) in PBS, washed a final time twice in PBS and mounted with Prolong Gold (Invitrogen, P36930) coverslips. Images were acquired using a Cell Discoverer 7 (Zeiss) with a 50× water immersion lens. Image processing and analysis were carried out using the provided microscope software (Zen).

2.4.2.5 | Western blotting

Proteins were resolved by PAGE through a NuPAGE 4-12% Bis-Tris gels and transferred to a 0.45 µm PVDF (Millipore, IPVH00010) membrane in transfer buffer (25 mM Tris, 190 mM glycine, 20% methanol), overnight at 35 V at 4 °C. Membranes were dried for 1 h at room temperature, re-activated in 100% methanol, blocked for 1 h in either 5% milk + tris-buffered saline (TBS) or 5% BSA-TBS at room temperature, with rocking. Membranes were then incubated with primary antibodies, diluted in 1% BSA or milk in TBST (TBS containing 0.1% Tween) overnight at 4 °C with rocking. Membranes were washed three times in TBST for 10 min each at room temperature and then incubated

Antigen	Supplier	Species	Cat. No.	Dilution
ADH5	Proteintech	Rabbit	11051-1-AP	1/1000
β-Actin	Sigma-Aldrich	Mouse	A3854	1/5000
β-Tubulin	Sigma-Aldrich	Mouse	T4026	1/1000
BARD1	Bethyl Lab	Rabbit	A300-263A	1/4000
BRCA1	Santa Cruz	Mouse	sc-6954	1/200
CHK1	Santa Cruz	Mouse	sc-8408	1/1000
CHK1 pS345	Cell Signalling Technology	Rabbit	2348	1/1000
EME1	Novus	Mouse	NBP2-50386	1/1000
ERCC1	Santa Cruz	Mouse	sc-17809	1/250
FANCJ	Sigma-Aldrich	Rabbit	B1310	1/1000
FLAG	Sigma-Aldrich	Mouse	F1804	1/2000
GAPDH	Cell Signalling Technology	Rabbit	3683	1/2000
GFP	TAKARA	Mouse	632381	1/4000
H3 pS10	Abcam	Mouse	ab14955	1/5000
HA	Santa Cruz	Mouse	sc-57592	1/1000
His-tag	BioLegend	Mouse	652503	1/2000
Histone H2AX	Novus	Rabbit	NB100-383	1/5000
Histone H2AX pS139	Cell Signalling Technology	Rabbit	2577	1/5000
Histone H3	Proteintech	Rabbit	17168-1-AP	1/5000
Ku80	Cell Signalling Technology	Rabbit	2753	1/1000
MSH2	Santa Cruz	Mouse	sc-376384	1/500
MUC2	Invitrogen	Mouse	MA5-12345	1/1000
P53	Santa Cruz	Mouse	sc-126	1/3000
PCNA	Santa Cruz	Mouse	sc-56	1/500
RPA32	Millipore	Mouse	NA18	1/2000
SLX1	Proteintech	Rabbit	21158-I-AP	1/500
SLX4	Bethyl Lab	Rabbit	A302-270A	1/1000
SLX4IP	Bethyl Lab	Rabbit	A304-994A-M	1/1000
SNM1A	Bethyl Lab	Rabbit	A303-747A	1/2000
TRF2	Novus	Rabbit	NB110-57130	1/2000
XPF	Bethyl Lab	Rabbit	A301-315A	1/1000
Anti-mouse-HRP	Agilent	Goat	P044701-2	1/2000
Anti-rabbit-HRP	Agilent	Goat	P044801-2	1/2000

Table 2.4 | Antibodies used in the study.

with HRP-conjugated secondary antibodies diluted in TBST for 2 h at room temperature. Membranes were rewashed three times in TBST for 10 min each at room temperature. Membranes were imaged through chemiluminescence after incubation with SuperSignal West Pico PLUS Chemiluminescent substrate (ThermoFisher Scientific, 34580), and imaged on an iBright FL1000.

Antibodies used are summarised in Table 2.4.

2.5 | Nucleic acid methods

2.5.1 | Nuclease assays

To anneal DNAs, appropriate oligonucleotides (Table 2.5) were annealed in buffer A (10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA) to a final concentration of 1 µM by incubating at 100 °C for 3 min before slow cooling to room temperature. Reactions were assembled with two-fold serial dilutions of protein and a final concentration of 10 nM DNA substrate (as in Table 2.5) in buffer N (25 mM HEPES-KOH (pH 8), 40 mM NaCl, 10% glycerol, 0.5 mM DTT, 5 mg/ml BSA, 5 mM $MnCl_2$). Reaction mixtures were incubated at 37 °C for 1 h, then quenched with the addition of 0.5 volume of buffer S (95% formamide (v/v), 10 mM EDTA, 0.25% bromophenol blue), and analysed by 20% denaturing polyacrylamide gel electrophoresis (PAGE), containing 40% 19:1 acrylamide:bis-acrylamide (BioRad, 1610144) with 7 M Urea (VWR Chemicals, 28877.292) in 1×Tris-borate EDTA (TBE), run at 525 V for 1.25 h. Gels were scanned using a Typhoon 9400 (GE).

2.5.2 | Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

For gene expression analysis, cells were washed in PBS and lysed in buffer RLT (Qiagen) before RNA was isolated using the RNeasy Mini Kit (Qiagen, 74104). After quantification of RNA *via* nanodrop, 1 µg RNA was treated with 0.2 U of DNase I (NEB, M0303) for 30 min at 37 °C. cDNA was generated using SuperScript IV Reverse Transcriptase (ThermoFisher Scientific, 18090) as per the manufacturer's instructions using $Oligo(dT)_{20}$ primer (ThermoFisher Scientific, 18418020) and diluted to a final volume of 200 µl

Oligo	Sequence	Label
1	ATA AAT ATT TTT TAT TAA TAA TAG ATC ACC TTT CTT TCT CTT CTC CCC TT	3′-Cy5.5
2	TTC CCC TCC TCT CCT TCC TTC CTG ATC TAT TAT TAA TAA AAA ATA TTT AT	
3	AAG GGG AGA AGA GAA AGA AAG G	
4	GGA AGG AAG GAG AGG AGG GGA A	
5	TCC GTC CTA GGA AGG GGC TGC TAC CGG AAG	5′-Cy5.5
6	CTT CCG GTA GCA GCC TGA GCG GTG GTT GAA	
7	TTC AAC CAC CGC TCA ACT CAA ACT GCA GCT	
8	AGA CTG CAG TTG AGT CCT TCC TAG GAC GGA	

Table 2.5 | **Oligonucleotide sequences used in the study and structures formed from their combination.**

Structure name	Structure	Oligos used
Simple fork		**1 + 2**
5′-flap		**1 + 2 + 4**
Replication fork		**1 + 2 + 3 + 4**
Holliday junction		**5 + 6 + 7 + 8**

Table 2.6 | **Oligonucleotide combinations to generate DNA structures.**

in nuclease-free water. For each biological repeat, four technical repeat qPCRs were performed in a QuantStudio3 (ThermoFisher Scientific) using TaqMan Gene Expression Assay kits with probes *SLX1A* (TaqMan, Hs00225232_m1, exons 3-4) and *SLX1B* (TaqMan, Hs02341353_g1, exons 4-5) controlling for the expression of *GAPDH* (Taqman FAM-MGB, Hs02786624_g1, exon 8).

2.5.3 | SAR-seq library preparation

SAR-seq was performed as described[728]. i^{3}Neurons were treated with the following compounds: SJG-136 (1 nM), formaldehyde (100 µM), formaldehyde + ADH5i (N6022 at 10 µM; Cayman, 21269-25 mg-CAY), were included in the medium along with 20 µM EdU 18 h before harvesting. Cells were washed with PBS, lifted with accutase for 5 min, collected in PBS and pelleted at 200 *g* for 5 min before resuspending in 0.1% BSA in PBS. Ice-cold methanol (final concentration of 80%) was added dropwise with slow vortexing and samples were kept on ice for 20 min, and stored at -20 °C until processing.

Cells were then washed in PBS, permeabilised with 0.2% Triton X-100 in PBS for 10 min on ice, washed in PBS and then incubated in Click-iT cocktail (made up of 3 mM copper sulfate, 50 µM biotin azide (ThermoFisher Scientific, B10184), 1× Click-iT additive (ThermoFisher Scientific, C10424)) for 2 h with shaking. Cells were then washed in PBS, lysed in 50 mM Tris-HCl (pH 8) with 1% SDS and proteinase K overnight at 37 °C. DNA was extracted through precipitation *via* phenol:chloroform:isoamyl alcohol (25:24:1 v/v, Invitrogen, 15593031), and 2.5 volumes of ethanol and 1/10th volume sodium acetate. DNA pellets were resuspended in TE buffer and sheared to 150-200 bp fragments using a Covaris S220 sonicator at 10% duty cycle, 175 peak incident power, 200 cycles per burst, for 240 s. DNA was then again precipitated with the ethanol/sodium acetate mix as above, before resuspending in TE buffer. Biotinylated EdU DNA fragments were immobilised by adding an equal volume of DNA in TE buffer to MyOne Streptavidin C1 beads (ThermoFisher Scientific, 650-01), which were washed with wash buffer (10 mM Tris-HCl (pH 8), 1 mM EDTA, 1 M NaCl, 0.1% Tween20) and resuspended in 2× wash buffer. Dynabeads were incubated at 24 °C with shaking for 30 min, then washed three times in 1 ml of wash buffer, twice in 1 ml EB, and once in 1 ml of 1× T4 DNA ligase buffer (NEB, B0202). Dynabeads were resuspended in 50 µl of end-repair reaction mix

(1× T4 DNA ligase buffer, 0.4 mM dNTPs, 2.7 U T4 DNA polymerase (NEB, M0203), 9 U T4 polynucleotide kinase (NEB, M0201), and 1 U Klenow fragment (NEB, M0210)) and incubated at 24 °C with shaking for 30 min. Dynabeads were washed further in 1 ml wash buffer, twice in 1 ml EB, and once in 1 ml 1× NEBuffer 2 (NEB, B7002) and resuspended in 50 µl A-tailing reaction mix (1× dA-tailing buffer (NEB, B6059) and 20 U Klenow fragment exo- (NEB, M0212)), with incubation at 37 °C with shaking for 30 min. Dynabeads were washed once in 1 ml NEBuffer 2 and resuspended in 115 µl ligation reaction mix (1× Quick Ligase buffer (NEB, B6058), 6000 U Quick Ligase (NEB, M2200), 5 nM annealed TruSeq truncated adaptor) and incubated at 25 °C with shaking for 20 min. Ligation reactions were stopped by adding 50 mM EDTA and Dynabeads were then washed three times in 1 ml wash buffer, three times in 1 ml EB and resuspended in 8 µl EB + 10 µl 2x Kapa HiFi HotStart Ready Mix (Kapa Biosciences, 07958927001). Primers (10 mM) 5′-CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGG AGTTCAGACGTGTGCTCTTCCGATCT and 5′-AATGATACGGCGA CCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T (asterisk indicates a phosphothiorate bond and a NNNNNN TruSeq index sequence) were added with 37 µl PCR reaction mix (20 µl 2× Kapa HiFI HotStart Ready Mix, 17 µl H_2O) for a final volume of 60 µl. DNA was amplified using a programme of 45 s at 98 °C, 15 cycles of 15 min at 98 °C, 30 s 63 °C, 30 s 72 °C, and final extension of 5 min at 72 °C. PCR products were magnetically separated from Dynabeads and cleaned with 1.8× volume AMPure Beads XP (Beckman Coulter, A63880). These PCR products were run on 2% agarose gels, with 150-200 bp bands excised and purified using the QIAquick gel extraction kit (Qiagen, 28704). Library concentrations were quantified by KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems) and normalised to 2 nM before sequencing using an Illumina NextSeq 500/550 (75 bp single read) kit.

2.5.4 | Library preparation for CRISPR-Cas9 genetic screens

A CRISPR-KO screen in HT-29 cells was performed as follows, but is an ongoing experiment, and the results have not yet been analysed for inclusion within any results chapter.

Total genomic DNA (gDNA) from 3.5×10^6 cells was isolated using the QIAamp Blood Maxi Kit (Qiagen, 51194), and concentrated to >500 ng/µl through ethanol precipitation. Total gDNA content was quantified by Qubit BR assay (ThermoFisher Scientific, Q33265)

and diluted to 300 ng/µl. To establish the relative quantities of genetic modifications that each cell possessed, a 128 bp region of the stably integrated lentiviral cassette consisting of the sgRNA was PCR-amplified using Q5 Mastermix Next Ultra II (New England Biolabs, Cat M5044L) with primers 319-320: V3_2_F: 5′-CTGCGTGCGCCAATTCTG, reverse: V3_1_R2: 5′-AGAACCGGTCCTGTGTTCTG, as part of PCR-1. For each sample, 35 x 50 µl PCR reactions were performed. For each sample, the PCR products of each reaction were pooled, and 5 µl was run on a 2% agarose gel to confirm successful amplification. A second PCR reaction containing i5 and i7 multiplexing barcodes was carried out, with dual primers uniquely identifying each sample (Table 2.7). For each sample, 4 x 50 µl PCR reactions were performed. Again, all reactions for each sample were pooled and purified using the QIAquick PCR cleanup kit (Qiagen, 28104) eluting in 30 µl, which was all subsequently run on a preparative 2% agarose gel. Bands corresponding to a 206 bp product were excised and purified using the QIAquick gel extraction kit, the eluant of which was further purified through the QIAquick PCR cleanup kit. The quality of these final PCR products was established by D1000 ScreenTape kit (Agilent) and quantified by Qubit BR assay. Sample PCR products were multiplexed at equal molar ratios (4 nM). The amplified sgRNA sequences and indices were sequenced on an Illumina NextSeq500/550 kit as described previously[768]. Sequencing was performed by Dr. Timothy Rostrom at the WIMM.

Sample	Experiment	i5 index	i5 primer ID	i7 index	i7 primer ID
HT-29 day 0	A	D_501	332	D_701	326
HT-29 Δ*ERCC4* day 0	A	D_501	332	D_702	325
HT-29 day 0	B	D_502	331	D_705	323
HT-29 Δ*ERCC4* day 0	B	D_502	331	D_701	326
HT-29 day 22	A	D_503	330	D_701	326
HT-29 Δ*ERCC4* day 22	A	D_503	330	D_706	322
HT-29 day 22	B	D_504	329	D_705	323
HT-29 Δ*ERCC4* day 22	B	D_504	329	D_702	325
HT-29 day 22 + SJG-136	A	D_501	332	D_705	323
HT-29 Δ*ERCC4* day 22 + SJG-136	A	D_502	331	D_706	322

Table 2.7 | Primer indices used for next-generation sequencing of CRISPR-Cas9 screens.

2.5.5 | Biophysical methods

2.5.6 | Cryoelectron microscopy (cryo-EM)

Purified chimaeric SMX complex preparations, diluted in glycerol-free buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM TCEP) were prepared for EM using conventional negative staining protocol[769] with 0.2% uranyl acetate and imaged at room temperature on a Tecnai T12 electron microscope operated at 120 kV. Images were recorded at a magnification of 68,000× and defocus 50-5000 µm on a Gatan CCD camera. For purified mouse SMX, protein was thawed and immediately buffer-exchanged into a MSG-/sucrose-free buffer (20 mM HEPES-KOH (pH 8), 250 mM NaCl, 2 mM $MgCl_2$) using Zeba Spin Desalting Columns (ThermoFisher Scientific, 89877) and prepared on Quantifoil R 1.2/1.3 300 mesh Cu/Au (Quantifoil), which were glow-discharged PELCO easiGlow GlowDischarge system (Ted Pella Inc., USA) for 60 seconds. Glow-discharged grids were frozen in liquid ethane using an FEI Vitrobot Mark IV (ThermoFisher Scientific) under 4 °C and 100% relative humidity. The blot force was -2 and blot time was 2 s.

2.6 | Quantification and statistical analysis

2.6.1 | Single-particle analysis

A total of 26 micrographs were collected for purification products χSMX, imported to Relion-3.1 for manual-particle-picking after CTF-correction, where 1397 particles were picked. 2D class averaging and generation of an initial model was completed in Relion-3[770], with visualisation of final 3D models in ChimeraX[771].

2.6.2 | SAR-seq data processing

SAR-seq reads were aligned to the reference genome (hg19) using bowtie with the specified parameters `-n 2 -l 50 -m 1`[772]. The resulting alignment files as `.sam` format were converted into `.bam` format and sorted using the `view` and `sort` functions within Samtools[773]. To further process the data, the sorted `.bam` files were transformed into the `.bed` format using the command `bamToBed` (v.2.29.2) from Bedtools[774]. For intensity

comparison analysis, mitochondrial reads were excluded from the SAR-seq dataset. *Note: the analysis of SAR-seq data was performed by Dr. William Nathan.*

2.6.3 | Colony survival assays

Colony numbers were automatically quantified using the ColCount program, with parameters tailored to the specific methods for each cell line. To account for variations in plating efficiency, the counts were normalised to untreated controls. For each cell type, condition, and drug dosage, the mean value of three technical replicates was calculated. Additionally, the reported data represent the mean ± standard deviation (s.d.) from three biological replicates (n = 3), unless otherwise specified in figure legends.

2.6.4 | Resazurin/Cell-Titer Glo

The mean of 3-6 technical repeats after subtraction of background relative fluorescence was reported as the value for a single biological repeat. Unless otherwise stated, three biological repeats were performed for each experiment.

2.6.5 | RT-qPCR

Raw Ct values were obtained from QuantStudio analysis software and analysed through the $2^{-\Delta\Delta Ct}$ method and are reported as relative gene expression compared to the *GAPDH* control.

2.6.6 | Accession of online databases

The following resources/databases were used in this study: For predicting the correlative effects of CRISPR-KO based knockouts in various cell lines, expression, and copy number data were retrieved from the DepMap portal and were accessed through `https://depmap.org/portal/`. Analysis of the genetic landscape of sensitivity towards various DNA-damaging drugs was accessed from `https://durocher.shinyapps.io/GenotoxicScreens/`. PhosphositePlus was used to identify any annotated PTMs on various protein sequences and accessed from `https://www.phosphosite.org/homeAction.action`. The Johannes Walter lab's protein prediction tool was used to identify paired

interactions within the DDR and was accessed from `https://predictomes.org/`[775]. Protein predictions were made using AlphaFold `https://alphafoldserver.com/`, and the AlphaBridge tool (`https://alpha-bridge.eu/`) was used to identify protein-protein interfaces between chains. AlphaFold outputs were plotted using an in-house script to map domains onto the heatmap obtained from an average predicted aligned error for all five models.

2.6.7 | Statistical methods and reproducibility

All experiments, unless otherwise stated in figure legends (relating to screening experiments) are representative of three repeats. For any survival experiment, including statistical significance testing, data points represent the mean of biological repeats, each comprising $\geq$3 technical repeats. Unless otherwise stated, error bars indicate the mean ± standard deviation. Statistical tests were performed using GraphPad Prism or in R using the `rstatix` package, with statistical tests detailed further in figure legends.

3 | Systems to study formaldehyde-induced DNA damage

Contents

3.1 | Source of endogenous lesions requiring the FA pathway

A causative agent for age-associated DNA damage phenotypes has recently been attributed to endogenous aldehydes, which are capable of forming DNA crosslinks and other genotoxic lesions[757]. These lesions are particularly relevant in the context of DNA repair-deficient models such as the *Ercc1*$^{-/\Delta}$ mouse, which remarkably recapitulates hallmarks of premature ageing[776,777]. If these features were solely reflective of a nucleotide excision repair (NER) defect, it would imply that UV-induced lesions are the primary drivers of such degeneration. However, this explanation is insufficient to account for systemic ageing phenotypes in internal tissues with negligible UV exposure. A broader model involves endogenous DNA lesions that either require NER or implicate other repair pathways, such as interstrand crosslink (ICL) repair via the Fanconi anaemia (FA) pathway. XPF-ERCC1's central role in both NER and ICL repair may therefore underlie the observed progeroid

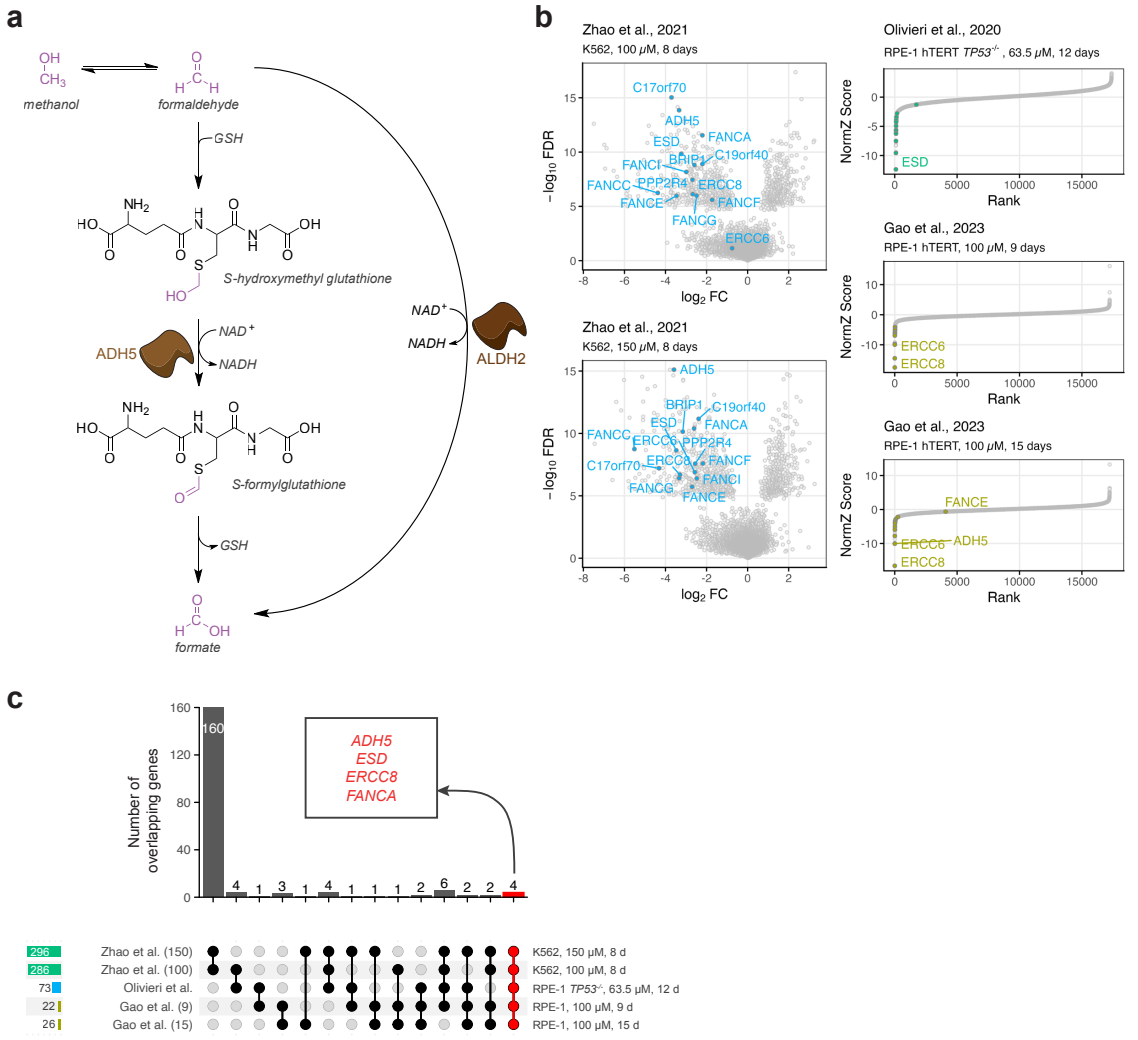


Figure 3.1 | Formaldehyde detoxification involves catabolism and the FA pathway. **a,** The biochemical pathway of formaldehyde metabolism, catalysed by ADH5 and ALDH2, converting formaldehyde to formate through intermediate S-hydroxymethyl glutathione. **b,** Meta-analysis of genetic screening data highlighting *ADH5* and DNA repair genes (e.g., *FANCA*, *ERCC8*) as significant hits across multiple CRISPR-Cas9 knockout screens for formaldehyde sensitivity. Volcano plots and ranked gene scores are adapted from Zhao et al. (ref[499]), Olivieri et al. (ref[446]), and Gao et al. (ref[500]) and indicate key genes associated with formaldehyde sensitivity. **c,** Overlap analysis of critical genes identified across datasets. The UpSet plot reveals the intersection of genes involved in formaldehyde resistance, emphasising *ADH5*, *ESD*, *ERCC8*, and *FANCA*.

syndromes in these mice and potentially reflect a broader class of homeostatic lesions not limited to canonical NER substrates.

Endogenous aldehydes, particularly formaldehyde, have emerged as potent sources of such genome damage. Detoxification of formaldehyde is governed by enzymes such as ADH5 and ALDH2 (Figure 3.1a), which have been repeatedly identified in CRISPR-Cas9 screens as key determinants of cellular resistance to formaldehyde-induced toxicity[9,778,779]. ADH5 loss sensitises cells to formaldehyde, while genes involved in

both aldehyde detoxification and DNA repair, including *FANCA*, *FANCG*, and *ERCC8*, are consistently implicated across five independent formaldehyde-focused CRISPR-Cas9 screens[446,499,500] (Figure 3.1b, c). Interestingly, *ALDH2* does not consistently appear as a hit, potentially due to redundancy with other mitochondrial aldehyde dehydrogenases or reflecting tissue-specific roles.

A clear physiological correlate of aldehyde-induced genome damage is seen in Fanconi anaemia (FA), a rare disorder of bone marrow failure (BMF) and developmental defects linked to ICL repair deficiency. In *Fancc*$^{-/-}$ mice, many haematologic features of human FA are absent under baseline conditions, despite cellular hypersensitivity to DNA crosslinking agents such as mitomycin C (MMC)[780–782]. However, BMF can be induced in these mice through acute or chronic MMC exposure[783], underscoring the context-specific vulnerability of haematopoietic stem and progenitor cells (HSPCs). In human FA patients, the CD34^{+} compartment is depleted, suggesting impaired stem cell maintenance or proliferative capacity[784]. ALDH2 is essential for the viability of *Fancd2*$^{-/-}$ embryos, and while heterozygous *Aldh2*$^{+/-}$ mothers can sustain development of *Fancd2*$^{-/-}$ *Aldh2*$^{-/-}$ offspring, such double-deficient mice exhibit spontaneous BMF highly reminiscent of human FA[785].

Traditionally, failure of the haematopoietic system in FA has been attributed to the high replicative output of HSCs, rendering them sensitive to replication-associated ICLs (Figure 3.2a, top inset). Indeed, the FA pathway has been implicated in the resolution of replication stress[786,787]. However, an emerging model proposes that differentiation, rather than proliferation, represents a critical window of susceptibility[59,401,788]. Epigenetic reprogramming during haematopoietic lineage commitment involves extensive chromatin remodelling, including the activity of histone demethylases such as LSD1[789]. These enzymes generate formaldehyde as a by-product of demethylation reactions, potentially leading to DNA crosslinking events precisely at the time when aldehyde detoxification and ICL repair are most essential (Figure 3.2a, bottom inset). Another possibility is that haematopoietic reprogramming includes a burst of transcriptional stress, with R-loop-derived damage reliant on aspects of the FA pathway[790]. This model posits that progenitor differentiation, driven by transcriptional rewiring, exposes cells to endogenous formaldehyde in a repair-deficient FA context, thereby exacerbating BMF (Figure 3.2a). Such a

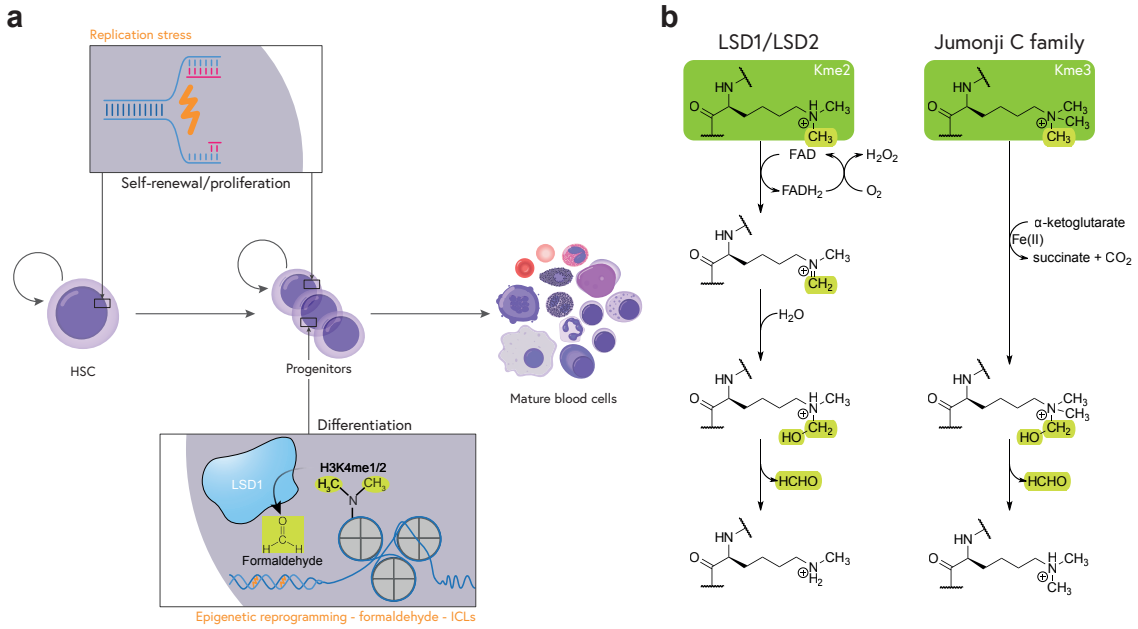


Figure 3.2 | Mechanisms of bone marrow failure in Fanconi anaemia. a, Schematic of DNA repair requirement during stages of haematopoietic stem cell (HSC) self-renewal, proliferation and differentiation into mature blood cells. During self-renewal and proliferation, HSCs experience replication stress, which can be detrimental in cells without a functional Fanconi anaemia (FA) pathway. As differentiation proceeds, epigenetic reprogramming occurs partly through bursts of histone demethylation, which generates endogenous crosslink-capable formaldehyde. This might then invoke the interstrand crosslink repair function of the FA pathway. **b,** The reactions of the enzymatic demethylation of histone lysine residues generate formaldehyde. Left: LSD1/LSD2 (flavin adenine dinucleotide (FAD)-dependent amine oxidases) catalyse the demethylation of lysine monomethyl and dimethyl groups (Kme2 depicted here), proceeding *via* oxidation and hydrolysis, releasing formaldehyde. Right: Jumonji C (JmjC) domain-containing demethylases use Fe(II) and α-ketoglutarate as cofactors to hydroxylate mono- to tri-methylated lysines (Kme to Kme3), also producing formaldehyde. Adapted from Kooistra & Helin (ref[56]).

perspective reframes the pathophysiology of FA as a disorder of disrupted differentiation, driven by aldehyde accumulation during chromatin remodelling, and further implicates FA proteins as key guardians of differentiation-linked genome integrity.

While formaldehyde release is a well-characterised consequence of histone demethylation, particularly through the enzymatic activity of LSD1/2 and members of the Jumonji domain-containing demethylase family (Figure 3.2b)[57], it remains unclear whether formaldehyde persists at sufficient concentrations or for a sufficient duration within the nuclear environment to induce covalent ICLs *in vivo*. Furthermore, the chemical feasibility of such crosslinking events under physiological conditions remains to be established. To directly assess whether the FA pathway is critical during differentiation, and whether formaldehyde generated through epigenetic reprogramming can indeed crosslink DNA, a multi-pronged strategy was employed. First, a stem cell model was established using

aldehyde-detoxification-deficient cells to determine whether such cells are particularly vulnerable during differentiation, especially in the context of a dysfunctional Fanconi anaemia pathway. Second, a sequencing-based approach (SAR-seq)[728] was adopted to map sites of DNA repair activity genome-wide and evaluate whether formaldehyde generates a repair footprint consistent with known interstrand crosslinking agents. Finally, a synthetic approach was used to direct histone demethylase activity to specific genomic loci using CRISPR-Cas9, enabling the interrogation of whether such targeted enzymatic activity induces localised DNA damage in a manner consistent with formaldehyde-dependent genotoxicity.

3.2 | Generating an aldehyde-sensitive differentiation model system

3.2.1 | iPS cell knockout of *ADH5* and *ALDH2*

In addressing whether epigenetic reprogramming might prove a source of endogenous formaldehyde, a hypothesis was proposed that stem cells unable to clear formaldehyde (e.g. deficient in ALDH2 or ADH5) would exhibit impaired differentiation or present increased markers of DNA damage during the differentiation process. To explore this, a collaboration with Andre Nussenzweig's laboratory was undertaken, leveraging their infrastructure and expertise with induced pluripotent stem cells (iPS cells). The aim was to generate *ADH5*$^{-/-}$ and *ALDH2*$^{-/-}$ iPS cells capable of differentiating into post-mitotic glutamatergic cortical neurons (i^{3}Neurons or i^{3}N)[728,759,760]. This system offers two advantages that are beneficial for this study. Generation of neural progenitor cells, which are precursors to mature neurons, typically spans 21-30 days[791], while active mature neurons require an additional month for differentiation from NPCs[792]. The iPS cells (referred to as illW iPSCs) doxycycline-inducibly express neurogenin-2 integrated into the safe-harbour AAVS1 locus, which aids the rapid conversion of iPS cells to neurons in a simple, two-step protocol requiring only six days[759]. Partial differentiation is achieved on Matrigel, after which re-plating onto dishes coated with poly-L-ornithine (PLO) completes neuronal maturation (Figure 3.6a)[760], as detailed further in methods (Section 2.1.2.1). This offers greater throughput in experimentation and accelerates any epigenetic change-induced

genome damage due to the comparatively faster differentiation rate. It was reasoned that the large-scale epigenetic changes occurring during this rapid differentiation[793] could be a significant source of DNA-damaging aldehydes.

These iPS cells have also previously been engineered to express dCas9-KRAB[728], enabling CRISPR interference (CRISPRi) screens to identify determinants of differentiation efficiency. However, this integrated and constitutively expressing dCas9-KRAB prevents delivery of Cas9/sgRNA in plasmid form for any CRISPR-based knockout strategies, lest to not risk any further integrations or interference between the two systems. Therefore, to generate *ADH5*$^{-/-}$ or *ALDH2*$^{-/-}$ iPS cells, sgRNAs were introduced as synthetic gRNA/Cas9 ribonucleoprotein (RNP) complexes (Section 2.3.1.5). For targeting of *ADH5*, sgRNAs creating a 1590 bp deletion between exons 6 and 7 were used (Figure 3.3a). Critical residues of ADH5 are encoded within exons 7 and 8, and mutations c.832C p.A278P or c.966delG p.W322* lead to aplastic anaemia, mental retardation, and short stature and microcephaly (dwarfism), also known as AMeD syndrome, a defect of formaldehyde clearance[778]. It was therefore concluded that disruption of the reading frame before these residues should result in a functional *ADH5* knockout despite the likely expression of half of the protein sequence. These sgRNAs were also chosen unbiasedly from an unpublished sgRNA library to minimise off-target effects (personal communications: Dr. Raj Chari). PCR around the cut sites was carried out (amplifying a 2.7 kb product in WT and a 1.1 kb product where the expected deletion was made) to estimate the efficiency of synthetic sgRNAs (Synthego), and confirmed that the combination of the two sgRNAs cut the *ADH5* gene in 70% of the bulk nucleofected iPS cells (Figure 3.3b). Since only a small deletion in the *ALDH2* gene was expected (Figure 3.4a), it was challenging to observe a PCR product size shift that could be visualised by electrophoresis during screening. Instead, PCR products from each colony, amplifying around the cut sites, were mixed with an equal amount of WT PCR product and a SURVEYOR-like assay was performed to search for any mismatches between the sequence in each colony and WT. The SURVEYOR assay involves melting of PCR-amplified DNA and random re-annealing such that some DNA duplexes contain mismatches from CRISPR-Cas9-dependent genome edits. T7 endonuclease I then cleaves any non-perfectly matched DNA, and the presence of fragmented PCR products is therefore an estimate of editing efficiency. For *ALDH2*,

the sgRNAs cut at a lower efficiency as measured by the SURVEYOR assay, with an estimated 25% efficiency (Figure 3.4b). Furthermore, to confirm the functional impact of using the sgRNAs, a colony survival assay was performed on illW iPS cells, pre-treated with sg*ADH5*, sg*ALDH2* and Cas9 alone. It appeared that iPS cells nucleofected with sg*ADH5* were more sensitive to formaldehyde than compared to those nucleofected with Cas9 alone or sg*ALDH2* (Figure 3.3c). However, as a preliminary verification experiment, this represented only a single biological repeat.

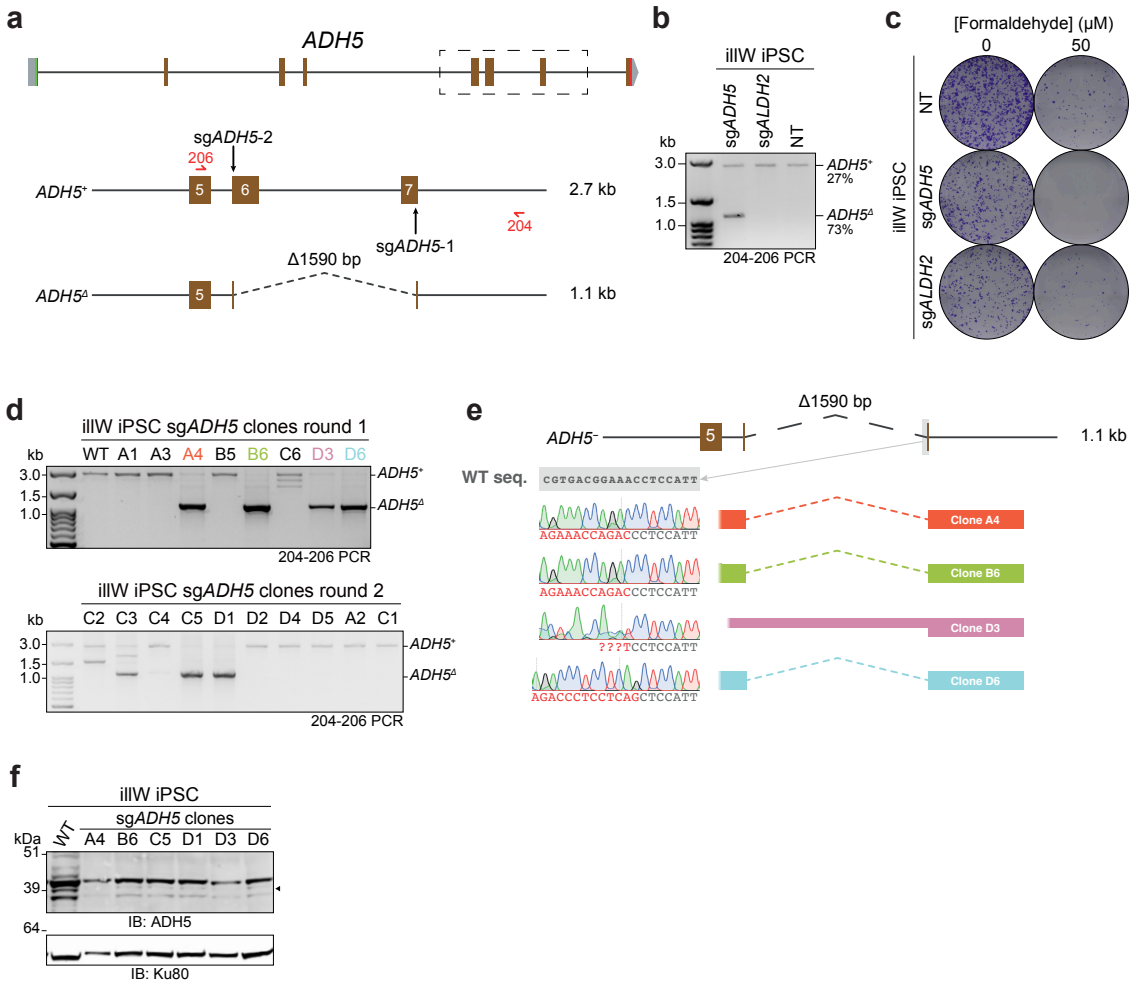


Figure 3.3 | Generation of *ADH5*$^{-/-}$ iPS cells. a, Schematic of the ADH5 gene editing strategy using two sgRNAs targeting exons 6 and 7 to induce a 1590 bp deletion. **b,** Validation of editing efficiency of the two sgRNAs used. iPS cells were nucleofected with sg*ADH5*, and sg*ALDH2*/no sgRNA as controls, and PCR was performed across the cut site to identify the proportion of cells with edited *ADH5*. **c,** Formaldehyde sensitivity assay comparing no sgRNA-Cas9-, sg*ADH5*-, and sg*ALDH2*-nucleofected iPS cells, revealing reduced colony survival in sg*ADH5* cells exposed to 50 μM formaldehyde. **d,** Colony screening by PCR for *ADH5*$^{-/-}$ clones, demonstrating the presence of the expected deletion in clones A4, B6, C5, D1 (upper panel), and D3, D6 (lower panel). **e,** Sanger sequencing of representative *ADH5*$^{-/-}$ clones (A4, B6, D3, D6) confirming the precise 1590 bp deletion. **f,** Western blot analysis showing loss of ADH5 protein expression in knockout clones (A4, B6, C5, D1, D3, D6); the expected band for ADH5 is marked.

3.2.1.1 | *ADH5* knockout

After nucleofection of iPS cells with both *ADH5*-targeting sgRNAs and Cas9, clones were manually picked and expanded. After genomic DNA extraction, PCR was performed as above to identify exon 6-7 deletion, which amplifies a 2.7 kb product in WT, and a 1.1 kb product where the expected deletion was made (Figure 3.3a). Of 30 single-cell clones expanded after nucleofection with sg*ADH5*, six clones were obtained with an expected deletion of 1590 bp within *ADH5* (Figure 3.3d). Clones A4, B6, D3, D6 appeared to be homozygous for the mutant allele of *ADH5* (Figure 3.3d) and Sanger sequencing of the products of these PCRs showed deletions across the 1.6 kb sequence, though not all presented with the same cut sites and clone D3 appeared to be heterozygous despite a single PCR band (Figure 3.3e). ADH5 protein status could not be definitively determined due to the presence of several non-specific bands by Western blotting. However, each of the six clones A4, B6, C5, D1, D3 and D6 appeared to lose the expected band at 39 kDa (Figure 3.3f).

3.2.1.2 | *ALDH2* knockout

34 clones were expanded from the pool of iPS cells nucleofected with sg*ALDH2*. Of these, six colonies were picked containing mismatches (Figure 3.4c), and two of these were sequence-verified to have homozygous indels that result in premature stop codons (Figure 3.4d). Both of these clones reflected cutting by only one of the sgRNAs used, sg*ALDH2*-1. In clone C3, the *ALDH2* ORF mistranslates from N375 and reads a premature stop at 393 (Figure 3.4d, middle panel). In clone D2, a N375K mutation is read until a premature stop also at 393 (Figure 3.4d, lower panel). Western blotting was performed to probe the loss of ALDH2, but there were, again, too many non-specific bands visible, hampering the identification of ALDH2 loss (data not shown).

3.2.1.3 | Functional characterisation of *ADH5*$^{-/-}$ and *ALDH2*$^{-/-}$ iPS cell clones

Short-term viability assays were carried out to verify the sensitivity of knockout clones to formaldehyde. By Western blotting, the most likely *ADH5*$^{-/-}$ iPS cell clone, A4, showed no sensitivity to formaldehyde alone or with ADH5 inhibitor (ADH5i) (Figure 3.5a). Of

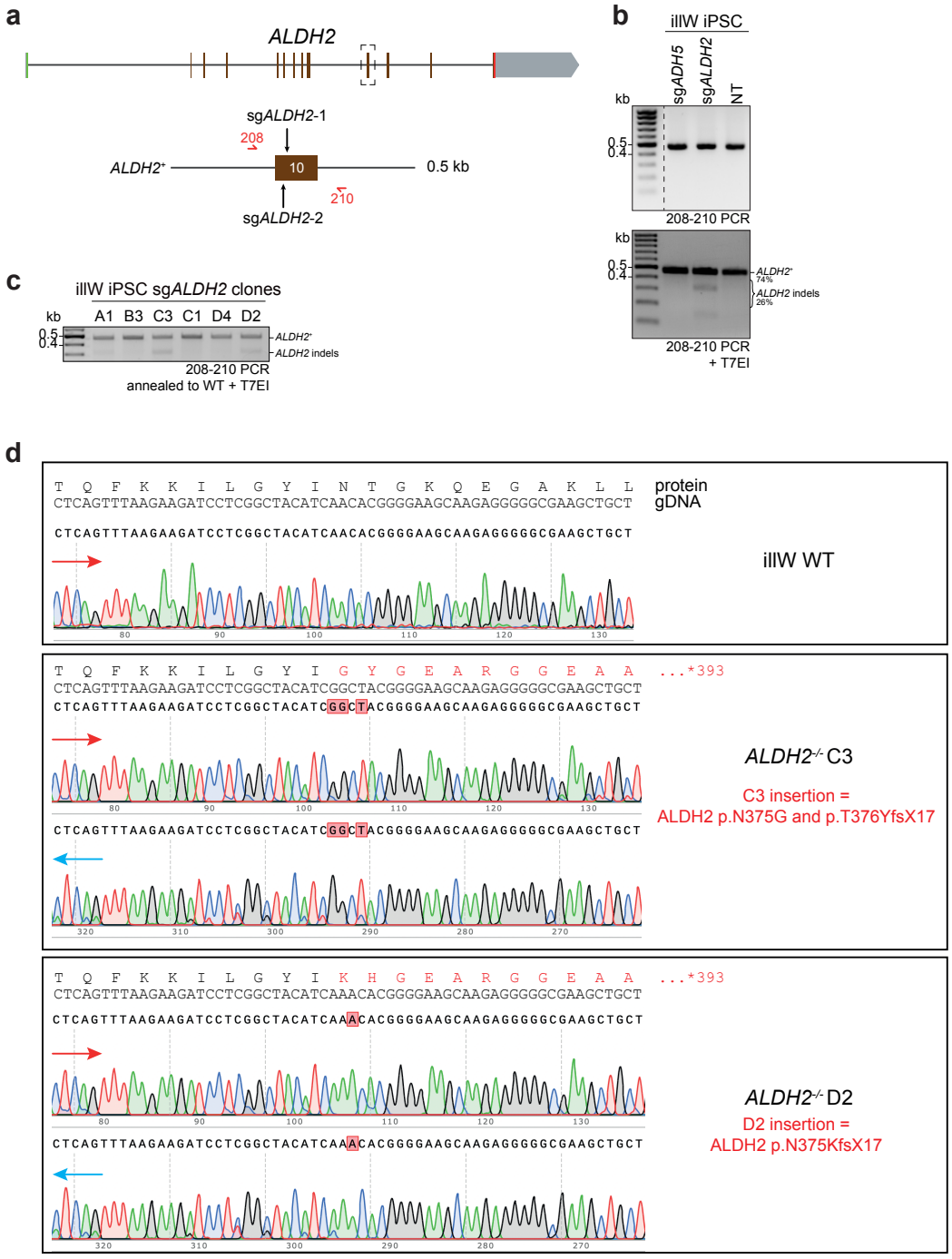


Figure 3.4 | **Generation of *ALDH2*$^{-/-}$ iPS cells.** (a) Schematic of the *ALDH2* gene editing strategy employing sgRNAs targeting exon 10, leading to small indels at the targeted site. **b,** PCR analysis confirming *ALDH2* editing in illW iPSCs, with the expected 0.5 kb amplicon and validation of indels by the SURVEYOR assay. **c,** Screening of *ALDH2*-edited clones (A1, B3, C3, C1, D4, D2) by PCR. PCR products from each clone were melted and annealed to WT PCR products and cleaved with T7 endonuclease I (T7EI), showing distinct cuts for edited alleles in each clone. **d,** Sanger sequencing of wild-type (WT) and representative *ALDH2*$^{-/-}$ clones (C3 and D2), highlighting a single base insertion leading to a frameshift that results in a premature stop codon. Clone C3 contains an insertion (p.N375G and p.T376YfsX17), while clone D2 shows another insertion (p.N375KfsX17).

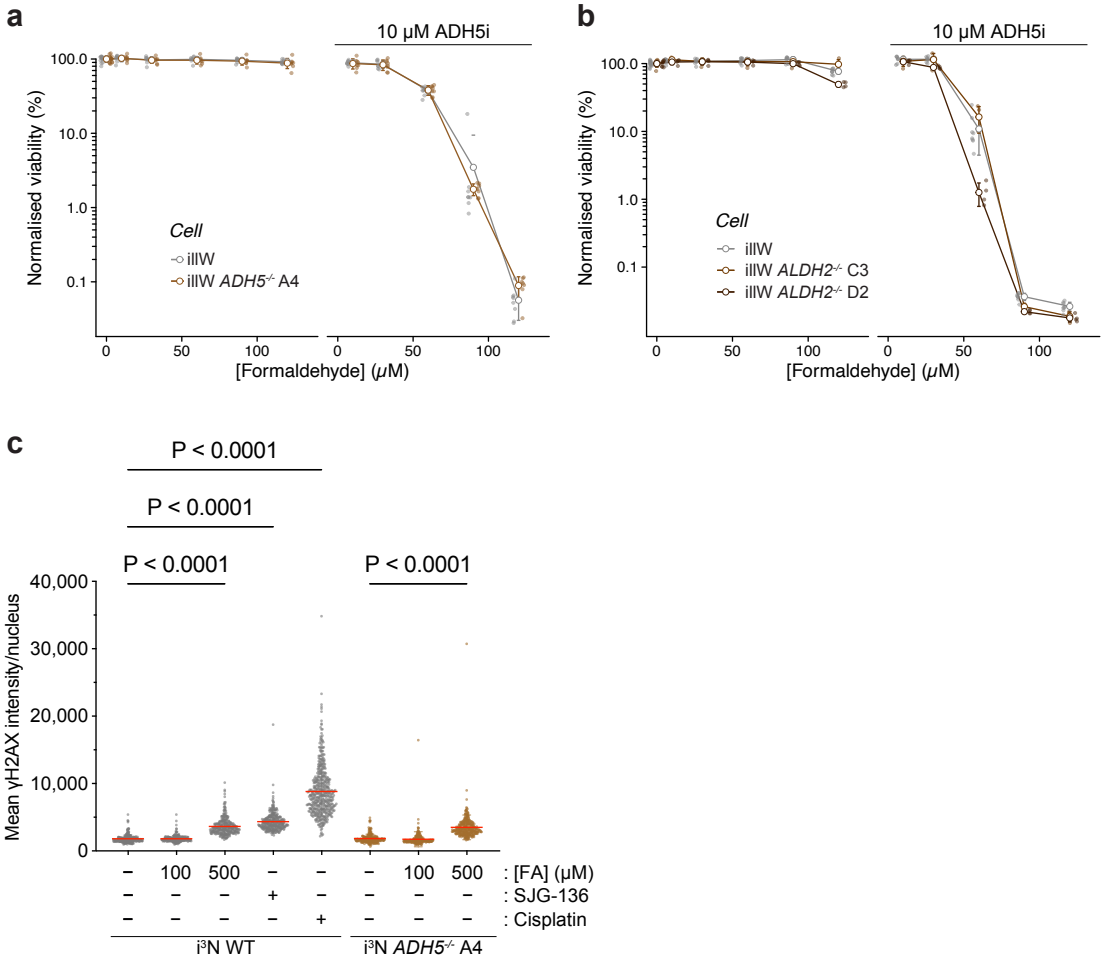


Figure 3.5 | Characterisation of *ADH5*$^{-/-}$ and *ALDH2*$^{-/-}$ iPS cells. **a,** Dose-response curve showing cell viability in wild-type (illW) and *ADH5* knockout (illW *ADH5*$^{-/-}$ A4) cells treated for 3 days with increasing concentrations of formaldehyde ± 10 μM ADH5 inhibitor (ADH5i). **b,** Dose-response curve for wild-type (illW) and illW *ALDH2*$^{-/-}$ C3 and D2 cells treated for 3 days with formaldehyde ± ADH5i. **c,** Quantification of γH2AX intensity in wild-type and *ADH5*$^{-/-}$ after differentiation into i^{3}N and then exposed for 18 h to varying concentrations of formaldehyde (FA), the DNA cross-linking agent SJG-136 (1 nM), or cisplatin (1 μM), represented mean of n = 1 biological repeat. Statistical significance was determined using a Kruskal-Wallis test with post-hoc Dunn's multiple comparisons test - comparisons between *ADH5*$^{-/-}$ and WT for each condition were carried out.

ALDH2$^{-/-}$ iPS cell clones, D2 showed only the slightest sensitivity to formaldehyde, with viability at 60% of WT iPS cells at the highest dose of 120 μM, but clone C3 did not recapitulate this phenotype (Figure 3.5b). It is possible that these clones do not represent true ALDH2 knockouts since the sgRNAs chosen are a considerable distance into the gene. It is also possible that ALDH2 loss alone is sufficient to observe notable formaldehyde sensitivity, given redundant and non-mitochondrial aldehyde detoxifiers such as ADH5, perhaps especially in a pluripotent cell state. To this end, we included an ADH5 inhibitor (N6022) and checked viability with increasing formaldehyde. Clone D2, though again not

clone C3, showed an increased sensitivity to formaldehyde compared to WT (Figure 3.5b).

Reasoning that these short-term assays were insufficient to separate differences in viability to formaldehyde between WT and aldehyde-detoxifying-deficient iPS cells, γH2AX abundance was measured to determine whether any increased damage was observed in *ADH5*$^{-/-}$ cells. As a proof-of-principle, neurons were observed to have elevated γH2AX intensity when treated with SJG-136 or cisplatin for a period of 18 hours (Figure 3.5c). Unfortunately, however, WT and *ADH5*$^{-/-}$ A4 i^{3}Ns had comparable levels of γH2AX, and no difference was observed under treatment with 100 µM formaldehyde or even up to 500 µM formaldehyde, despite induction of damage at this concentration (Section 3.5c).

3.3 | Mapping the DNA breaks generated by formaldehyde

Alongside generating these models, an orthogonal experiment set out to identify the chemical nature of formaldehyde damage to DNA. SJG-136 is a potent crosslinking agent and is considered to generate solely ICLs[794]. SAR-seq can map sites of DNA repair associated with DNA synthesis (Section 1.5.1). DNA polymerisation strictly reflects repair synthesis in post-mitotic neurons that have escaped replication — the incorporation of EdU, therefore, charts regions of the genome undergoing repair-associated replication (Figure 3.6a). Treatment of i^{3}N with SJG-136 results in SAR-seq peaks that localise to genes (Figure 3.6c, d). Such a pattern does not dictate that crosslinks only form at gene bodies but that any repair-associated DNA synthesis is localised at genes. The replication observed also shows the pronounced role of transcription-coupled crosslink repair in dealing with crosslinks in a non-FA-dependent pathway (Figure 3.6c, d). These results were used as a benchmark to compare treatment with formaldehyde, such that if formaldehyde acts as a source of interstrand crosslinking lesions, a similar SAR-seq pattern would be observed. The total DNA content within the amplified SAR library can measure the overall damage caused by an agent. Treatment of i^{3}Ns with methanol caused no apparent enrichment in overall SAR library content as seen for treatment with SJG-136 (Figure 3.6b). At the same time, there was a slight increase in SAR library abundance when i^{3}Ns were treated with formaldehyde, which was enhanced in the presence of ADH5i

(Figure 3.6b). After high-throughput sequencing of prepared SAR libraries, formaldehyde treatment appeared not to result in the typified enrichment pattern seen for SJG-136 crosslinking (Figure 3.6c, d). However, formaldehyde alongside ADH5i was more akin to SJG-136 treatment and showed more notable SAR peaks at gene bodies (Figure 3.6c, d).

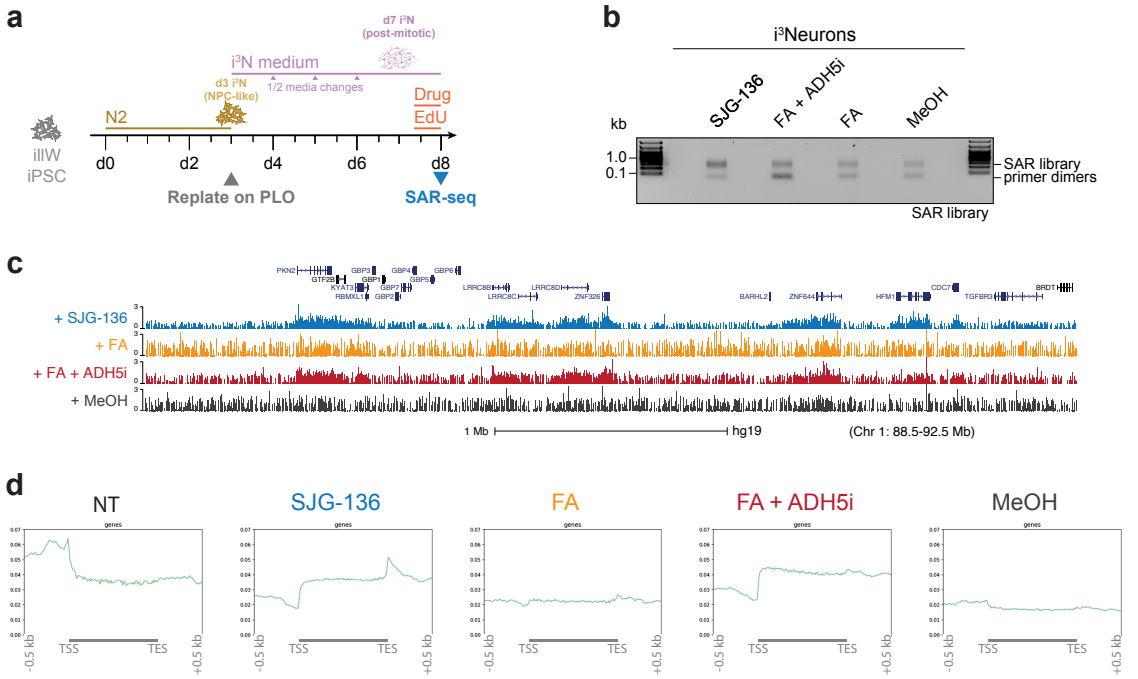


Figure 3.6 | SAR-seq comparisons of formaldehyde treatment with a potent crosslinker, SJG-136. a, Timeline of illW iPS cells differentiation to i^{3}Neurons and SAR-seq experimental setup. Cells were differentiated from illW iPSCs, and glutamatergic i^{3}Neurons at d7 were exposed to indicated conditions (100 µM formaldehyde (HCHO), 100 µM HCHO + 10 µM ADH5 inhibitor (ADH5i), 1 nM SJG-136, 0.1% methanol (MeOH)) alongside 10 µM EdU for 18 h prior to harvesting. **b,** Validation of SAR library preparation for i^{3}Ns under different treatments, showing amplified SAR library products. **c,** SAR-seq coverage tracks depicting enrichment at exons within chromosome 1 (88.5–92.5 Mb region, hg19). **d,** Heat maps of SAR-seq signal for 0.5 kb on either side of gene boundaries in i^{3}Neurons, ordered by RNA-seq TPKM[728].

3.4 | An alternative model for differentiation in the cancer cell line HT-29

The observations that formaldehyde exposure results in an ICL-like repair signature prompted further investigation into whether differentiation itself, rather than proliferation, might represent a state of increased vulnerability to DNA damage in the absence of repair factors such as XPF. While iPSCs represent a powerful system for probing this hypothesis, particularly due to the physiological relevance of their differentiation trajectories, time

constraints within the collaborating laboratory limited the extent of experiments conducted in that system. It would also be valuable in future studies to determine whether aldehyde-induced genotoxicity during differentiation is a generalisable phenomenon across multiple lineages beyond the haematopoietic system.

Fortunately, an observation made while performing colony survival studies of HT-29 WT and Δ*ERCC4* cells (further detailed in chapter 4.1) was that the growth of HT-29 in DMEM resulted in colonies that were different in morphology and were only weakly attached to the plate such that colonies would become detached even with careful PBS washes (Figure 3.7a, c). Based on this characteristic, colony formation numbers after plating in DMEM were significantly reduced compared to plating in RPMI (Figure 3.7b). For the specific purpose of colony survival assays, a switch to RPMI solved this issue entirely, and colonies could be maintained and stained well (Figure 3.7a, b). This observation prompted a further look into the unique characteristics of HT-29 cells that govern this difference. HT-29 cells are a human colon adenocarcinoma cell line, which is phenotypically stem-like under standard growth conditions[795]. Treatment with forskolin, colchicine, nocodazole and taxol induces the differentiation and apoptosis of HT-29 cells[795]. The differentiated phenotype resembles mature intestinal cells with a strict monolayer formed that holds tight junctions between cells, akin to an apical brush border[796].

One report describes how HT-29 differentiation can be obtained through a simple switch of growth medium from RPMI to DMEM[797], likely explaining the different staining proficiencies when grown in the two mediums. The switch from RPMI to DMEM resulted in a different morphology of HT-29 cells, which appear to form more epithelial-shaped cells within colonies (Figure 3.7c). Furthermore, the expression of the goblet cell-specific, MUC2 marker[798], increased when HT-29 cells were cultured in DMEM (Figure 3.7d), and this induction was observed from day 2 in DMEM (Figure 3.7d). Therefore, as with the iPSC study (Section 3.2), whether differentiation is slowed or leads to increased DNA damage, especially in the context of XPF loss, could be studied in this more tractable HT-29 model system.

While the expression of MUC2 increased gradually over 5 days of incubation in DMEM for HT-29 cells, the overall level of MUC2 expression was markedly less in HT-29 Δ*ERCC4* cells (Figure 3.8a). In the presence of ADH5i, the effect on MUC2 expression

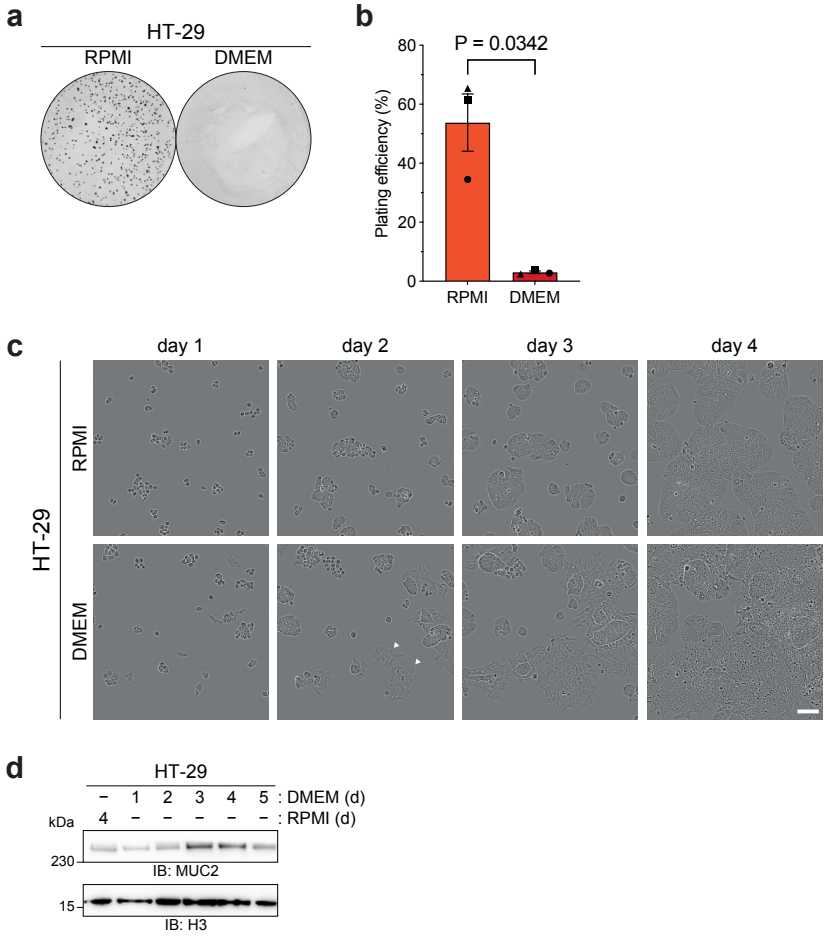


Figure 3.7 | **Differentiation of HT-29 cells.** **a,** Colony formation assay showing significantly reduced colony staining in HT-29 cells cultured in DMEM compared to RPMI. **b,** Quantification of plating efficiency for HT-29 cells in RPMI and DMEM media, represented as mean ± s.d. of n = 3 biological replicates, each consisting of three technical replicates. Statistical significance was determined using a two-tailed, paired t-test. **c,** Phase-contrast microscopy images of HT-29 cells cultured in RPMI or DMEM over four days, illustrating distinct morphological differences and reduced proliferation in DMEM. Scale bar: 100 µm. Epithelial-like morphology, which is different to the expected HT-29 morphology, is highlighted by white triangles. **d,** Western blot analysis of MUC2 protein expression over five days of culture in RPMI or DMEM.

in HT-29 Δ*ERCC4* cells was even more pronounced, while having little effect in WT cells (Figure 3.8b). DMEM conditions also support higher overall maximal confluencies compared to RPMI in HT-29 WT cells, and the increased maximal confluency is not affected by the presence of ADH5i (Figure 3.8c, left panel). Maximal confluency of HT-29 Δ*ERCC4* cells in DMEM is, however, affected by the presence of ADH5i but not when cultured in RPMI (Figure 3.8c, right panel). The overall growth rate, measured by the area under the curve (AUC) of Figure 3.8c, is significantly lower in the presence of ADH5i only for XPF-depleted HT-29 cells in DMEM (Figure 3.8d). These data reflect how differentiation (occurring in DMEM) is diminished and leads to reduced colony

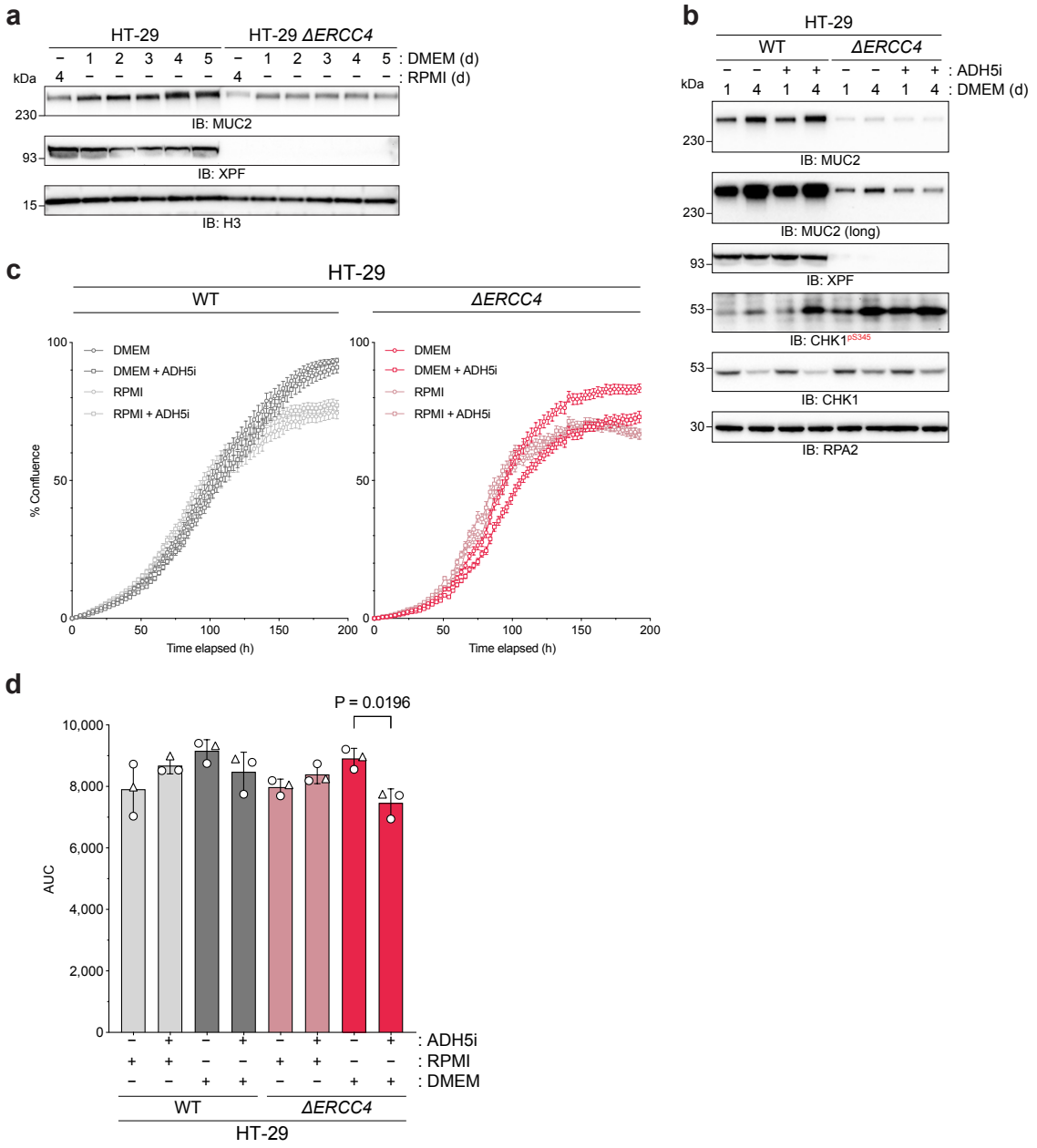


Figure 3.8 | Loss of XPF hampers HT-29 differentiation into goblet cells, with indications of increased DNA damage. **a,** Western blot analysis of MUC2 expression in wild-type (WT) and Δ*ERCC4* HT-29 cells cultured in RPMI or DMEM media over five days. **b,** Western blot analysis of WT and Δ*ERCC4* HT-29 cells cultured in DMEM with or without 25 µM ADH5 inhibitor (ADH5i) for four days. **c,** Cell confluence analysis of WT and Δ*ERCC4* HT-29 cells grown in RPMI or DMEM media over 8 days, with or without 25 µM ADH5i. Each curve represents the mean ± s.d. of n = 3 biological replicates, each consisting of 4 technical replicates. **d,** Quantification of the area under the curve (AUC) for cell confluence data in **c**, represented as the mean ± s.d. of the n = 3 biological replication as per **c**. Point shape represents matched experiments in order: circle, square, triangle. Statistical significance was determined by Brown-Forsythe and Welch ANOVA tests, with post-hoc Dunnett's T3 test for multiple comparisons. Comparisons made as follows: For each of WT and Δ*ERCC4* cells, RPMI vs RPMI + ADH5i, and DMEM vs DMEM + ADH5i.

outgrowth where XPF is absent. This phenotype is exacerbated by ADH5 inhibition. These experiments suggest that epigenetic remodelling during differentiation is a source of endogenous, DNA-crosslinking formaldehyde, requiring the action of ADH5 for detoxification or XPF failing that.

3.5 Induction of formaldehyde at defined genomic loci

The studies described above rely on the ability to differentiate cells and monitor their phenotypes. An alternative approach to investigate the effects of endogenous aldehydes in DNA ICL repair involves generating crosslinking agents *in situ*. This method would employ a system that enzymatically releases a crosslinking agent, such as formaldehyde, to create DNA ICLs at specified genomic locations. The precise targeted release of these crosslinking agents to defined genomic regions can be achieved using the CRISPR-Cas9 system.

Aldehydes are capable of crosslinking DNA and are generated in direct proximity to DNA during the reversal of certain histone marks[799]. Histone demethylation proceeds *via* the release of formaldehyde molecules, and therefore, likely increases the local concentration of DNA-proximal formaldehyde[800]. Given this hypothesis, does the fusion of a histone demethylase to dCas9 prove sufficient for a local formaldehyde concentration capable of forming DNA adducts? Can this system then be evolved to exploit further DNA-damaging chemistries? As an example, does the fusion of an alcohol dehydrogenase release formaldehyde or acetaldehyde if a cell is burdened with an excess of methanol or ethanol? It is foreseeable that such a system can be adapted to 'activate' a harmless pro-drug near a target genomic locus by a dCas9-fused enzyme, which might open an interesting avenue for therapeutics. As a start, however, constructs were designed that would fuse dCas9 to LSD1, a H3K4me1/2 demethylase (Figure 3.9a)[801,802]. Alongside the androgen receptor, AR, LSD1 can also demethylate the repressive H3K9 methyl marks[803] and reportedly catalyses the demethylation of non-histone substrates such as DNMT1 and p53[804]. The fusion was also considered with a DNA demethylase such as TET1. However, the DNA demethylation pathway appears to avoid the generation of reactive aldehydes, instead opting to shuttle formaldehyde through formate, which is then resolved by BER (Section 1.3.1)[805]. Thus, a DNA demethylase-dCas9 fusion would not be effective in generating isolated aldehydes.

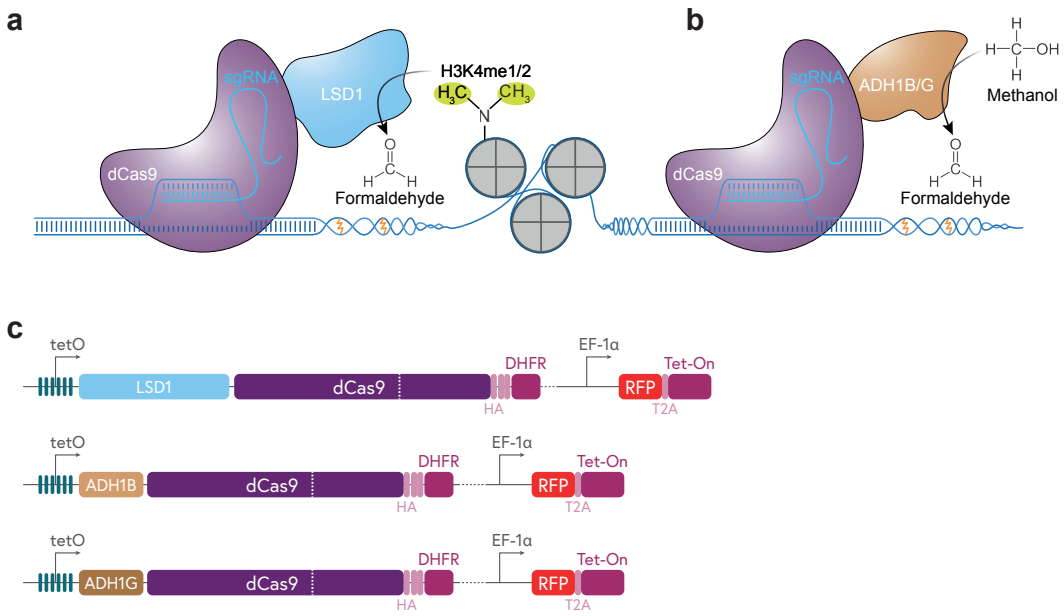


Figure 3.9 | Design of dCas9-based systems for formaldehyde induction near the genome. a, Schematic of the dCas9-LSD1 system, designed to target formaldehyde-induced H3K4 methylation sites. LSD1 demethylates H3K4me1/2 to release formaldehyde[801]. **b,** Schematic of the dCas9-ADH1B/G system, engineered to produce formaldehyde from methanol *via* the enzymatic activity of ADH1B or ADH1G, enabling modulation of intracellular formaldehyde levels. **c,** Genetic constructs for dCas9 systems under tetO promoter regulation. Constructs include LSD1-dCas9, ADH1B-dCas9, and ADH1G-dCas9 fusions, with DHFR (dihydrofolate reductase), and EF-1α-driven RFP-Tet-On to control expression from the tetO array.

However, the enzymatic activity of alcohol dehydrogenases as the fused target was considered. Alcohol dehydrogenases cofactor with NAD^+ to aid the oxidation of toxic alcohols to aldehydes and ketones[806]. Seven genes encode human ADHs, which all exist as dimers[807]. The predominant ADH belongs to class 1 (of five evolutionary ADH classes), and is encoded by genes *ADH1A*, *ADH1B* and *ADH1C*, which produce α, β, and γ subunits, respectively[807,808]. The remaining ADH classes, II to V, are encoded by genes ADH4, ADH5, ADH7 and ADH6, respectively[807]. These genes enable humans to metabolise and tolerate alcoholic beverages. However, their evolutionary derivation likely originates from processing alcohol traces in either food or that which is produced by fermentation within some gut bacteria. The substrate specificity of ADHs is comparatively relaxed[809], ceding to even the complex alcohol retinol, which is oxidised to retinal and then retinoic acid, an agonist influencing gene expression changes[810]. This enzymatic promiscuity is vital in enabling this study as fused ADHs could accept and convert the simplest alcohol, methanol, to formaldehyde (Figure 3.9b)[811,812]. As a candidate for the fusion to dCas9, at first, orthologues of ADHs from other species were considered, such as equine ADH1E,

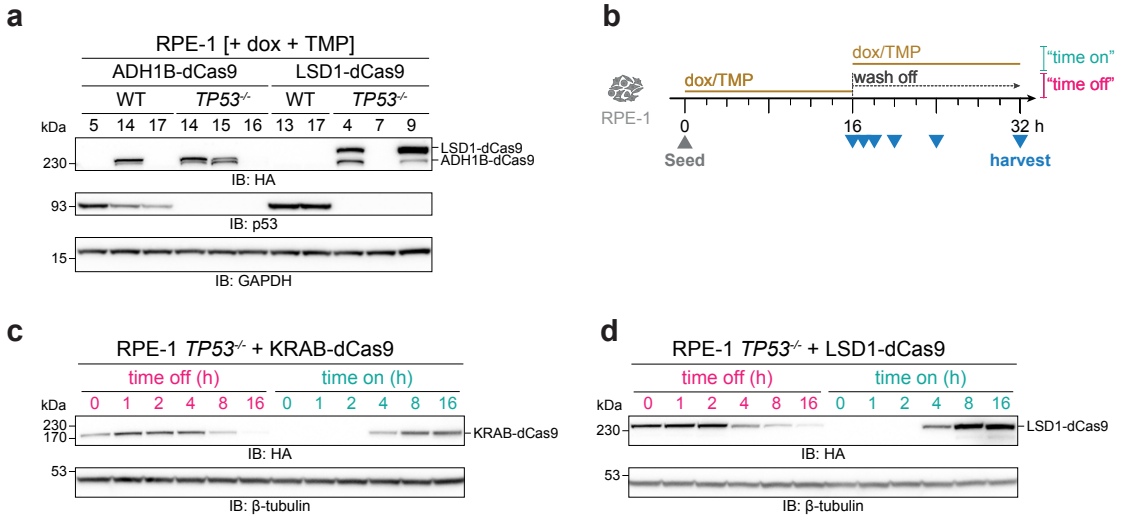


Figure 3.10 | **Regulation of dCas9 fusion protein expression in RPE-1 cells. a,** Western blot analysis of ADH1B-dCas9 and LSD1-dCas9 expression (HA) in wild-type (WT) and *TP53*$^{-/-}$ RPE-1 cells treated with 2 µg/ml Doxycycline (Dox) and 50 nM trimethoprim (TMP). **b,** Experimental timeline illustrating the induction ("time on") and washout ("time off") of Dox/TMP treatment in RPE-1 cells. Cells were harvested at defined intervals for analysis. **c,** Western blot analysis of KRAB-dCas9 (HA) expression in RPE-1 *TP53*$^{-/-}$ cells during "time off" and "time on" phases. **d,** Western blot analysis of LSD1-dCas9 (HA) expression in RPE-1 *TP53*$^{-/-}$ cells during "time off" and "time on" phases, showing similar regulation dynamics as KRAB-dCas9.

which has been extensively studied[813,814]. However, a suggestion was made that the human ADH1B and ADH1G enzymes have greater efficiency for methanol oxidation (personal communication with Prof. Bryce Plapp, University of Iowa) (Figure 3.9b). As part of design considerations, it was envisioned that a constitutively expressing dCas9-fusion system would allow time for adaptation in the cells to local increases in formaldehyde, perhaps leading to the development of tolerance mechanisms that hinder these studies. Instead, an inducible system was chosen. However, any low-level expression of the fused construct could result in the same adaptive changes over time, which ruled out the use of a Dox-inducible TetON system alone; leaky expression has been noted for this system due to the presence of tetracycline in medium (though, this could be prevented by the use of tetracycline-free medium)[815,816]. A solution is offered by a construct from the Emma Rawlins group, who developed a dual-inducible CRISPRi system (KRAB-dCas), which combines Dox-inducible TetON with a small-molecule (trimethoprim, TMP)-targeted degron, DHFR, offering robust and well-controlled induction[763].

In these lentiviral constructs, KRAB was replaced with genes encoding LSD1, ADH1B, and ADH1G (Figure 3.9b). Lentiviruses produced from these constructs were transduced

into RPE-1 cells, single-cell sorted for TagRFP+ (Appendix B.1, Figure B.1), and grown as single colonies. No colonies could be expanded further from ADH1G-dCas9 transduced cells, but several were obtained for ADH1B-dCas9 and LSD1-dCas9. Of these, clones were maintained and verified for the Dox- and TMP-inducible expression of the dCas9-fusion, which could be monitored by Western blotting for HA (Figure 3.9b). ADH1B-dCas9 cells were obtained in a WT RPE-1 background (clone 14), and in *TP53*$^{-/-}$ cells (clones 14 and 15), while LSD1-dCas9 cells were obtained in WT RPE-1 cells (clones 3 and 8, only shown in Figure 3.10a), and in *TP53*$^{-/-}$ cells (clones 4 and 7) (Figure 3.10a). Despite having established clones of RPE-1 cells that can inducibly express ADH1B-dCas9, from this point, the LSD1-dCas9 cells were chosen for further study. This was to ensure that a more physiologically relevant system could be developed first. Given that LSD1, and not ADH1B, exerts its catalytic function more in the vicinity of DNA under normal physiology, LSD1-dCas9 provides a better model of how formaldehyde might be generated near and crosslink DNA. To identify the minimum time required for Dox/TMP-based induction of or degradation of the fusion construct, LSD1-dCas9 expression was monitored by Western blotting for HA after varying lengths of Dox/TMP treatment and Dox/TMP wash-off (Figure 3.10b, c). Similar to RPE-1 *TP53*$^{-/-}$ cells expressing KRAB-dCas9 (as a control), Dox and TMP induces expression of the dCas9-fusion in as little as 4 h, though its presence lingers for a further 16 h after wash-off (Figure 3.10c, d).

Initially, general DNA damage was visualised through the induction of γH2AX as measured by Western blotting. In two separate RPE-1 LSD1-dCas9 clones (clones 3 and 8) and one RPE-1 *TP53*$^{-/-}$ LSD1-dCas9 clone, no induction of γH2AX was observed after induction of fusion construct expression for 24 h (Figure 3.11a). This was an expected but necessary verification to ensure that, without sgRNA-dependent targeting to particular loci, the expressed LSD1-dCas9 could not result in DNA damage alone. It is worth noting that there was a slight increase in p53 within the WT RPE-1 cells when induced for the expression of LSD1-dCas9 (Figure 3.11a). However, the RPE-1 *TP53*$^{-/-}$ LSD1-dCas9 were used from this point forward. This cell line, and the parental RPE-1 *TP53*$^{-/-}$ were nucleofected with synthetic RNAs resulting in expression of sgRNAs targeting the promoter region of *PDXDC1* and telomeres. Induction of LSD1-dCas9 expression in cells nucleofected with sg*PDXDC1*-promoter and sgTelomeres resulted in increased γH2AX

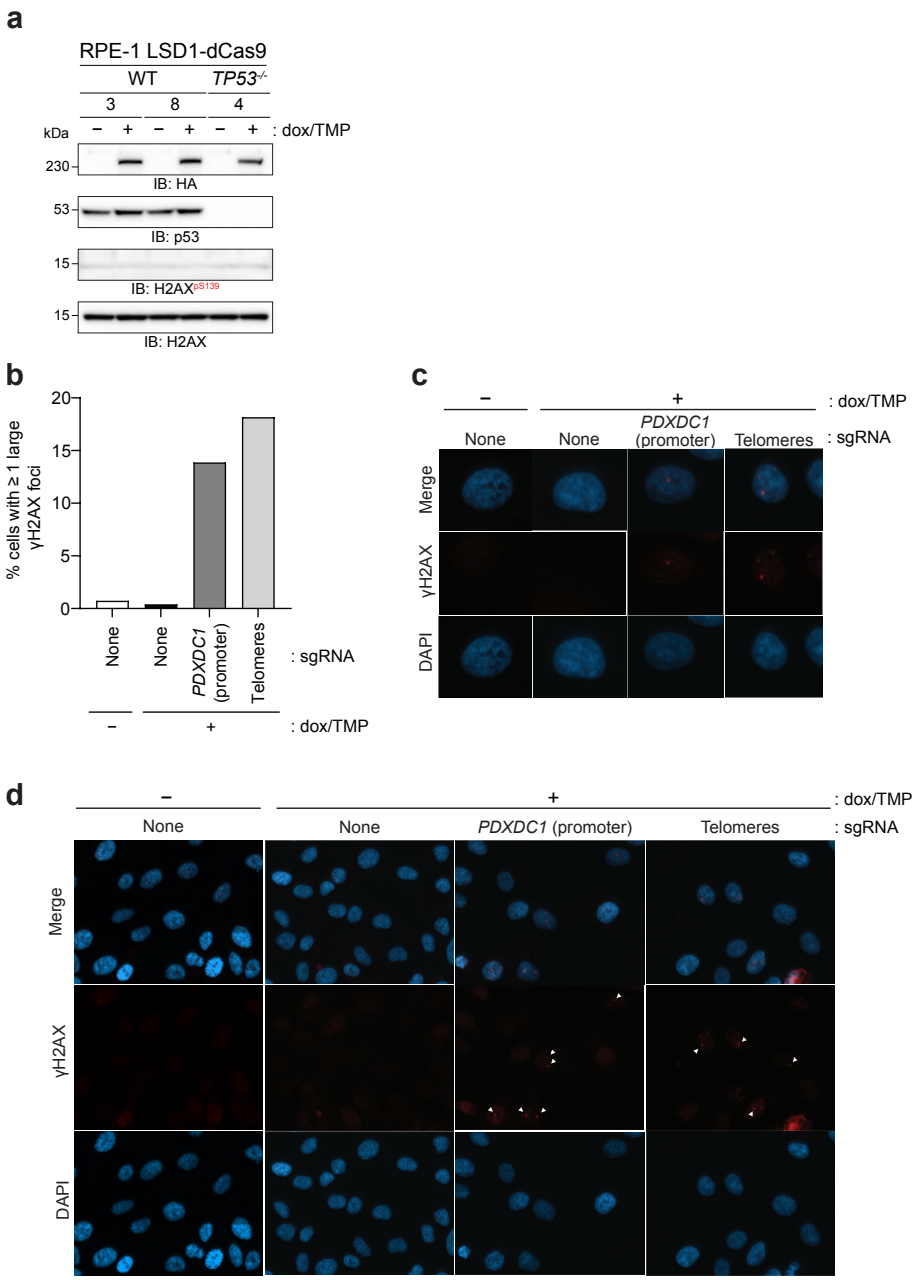


Figure 3.11 | Analysis of DNA damage response in RPE-1 LSD1-dCas9 cells following targeted induction. **a,** Western blot analysis of HA-tagged LSD1-dCas9 in wild-type (WT) and *TP53*$^{-/-}$ RPE-1 cells treated with 2 µg/ml Doxycycline (Dox) and 50 nM trimethoprim (TMP) for LSD1-dCas9 induction. **b,** Quantification of cells with ≥1 large γH2AX foci under indicated conditions. Recruitment of LSD1-dCas9 to the *PDXDC1* promoter or telomeres significantly increases DNA damage marker formation compared to untreated controls. **c,** Immunofluorescence images of γH2AX foci in individual RPE-1 cells targeting the *PDXDC1* promoter or telomeres following Dox/TMP treatment. **d,** Lower magnification immunofluorescence images highlighting γH2AX foci formation at the *PDXDC1* promoter and telomeres after Dox/TMP treatment. Arrowheads indicate foci.
*Note: Data represented in **b**, **c**, **d** was collected and presented by Dr. Mélissa Thomas*

foci only when cells were treated with Dox/TMP to induce LSD1-dCas9 expression (Figure 3.11b, c, d). This was a preliminary result proving that the induction of DNA damage is achievable through this system.

3.6 | Discussion

The work in this chapter represents the beginnings of various strategies to tackle one underlying hypothesis: formaldehyde, created during epigenetic reprogramming, generates crosslinked DNA requiring the FA pathway for its repair.

Use of iPS cells as a model for differentiation The underlying aim of this project was to exacerbate the toxicity of any formaldehyde that is created during the differentiation of iPS cells to i^{3}Ns. Loss of ADH5 or ALDH2 removes a frontline defence against formaldehyde toxicity, hence the aim of knocking these genes out in iPS cells. Clones were obtained for both cell lines, which were sequence-verified to be genetically deleted as expected (Figure 3.3, Figure 3.4). In *ADH5*$^{-/-}$ clones, this deletion results in truncation of ADH5 prior to key catalytic residues at the C-terminus of ADH5. However, one *ADH5*$^{-/-}$ iPS clone, A4, was not sensitive to formaldehyde, as was expected (Figure 3.5). This might reflect redundancy between ADH5 and other formaldehyde detoxifiers, such as ALDH2. Mice showcase significant postnatal lethality only when both *Adh5* and *Aldh2* are deleted[9]. Furthermore, in the *ALDH2*$^{-/-}$ iPS cells, only one clone showed slight sensitivity to formaldehyde that was increased further with the inclusion of ADH5 inhibitor. It is possible, however, that targeting so far into the body of this gene provides sufficient enzymatic activity for protection against formaldehyde. Future work using this clone alongside an ADH5 inhibitor might best resemble a detoxifying defect in these cell lines. If this project were followed up, a first experiment to perform would be to determine if this *ALDH2*$^{-/-}$ D2 clone is less able to effectively differentiate into glutamatergic neurons, in the presence or absence of ADH5i, compared to WT iPS cells. The proposed experiment herein would measure the viability of neurons or the proportion of cells expressing post-mitotic neuronal markers, such as NeuN (RBFOX3), after the usual differentiation process. This would be followed up by the disruption of the FA pathway, namely, disruption of XPF-ERCC1, which is believed to be an essential second line in the protection against differentiation-induced formaldehyde increases.

HT-29 cell differentiation efficiency is lower in XPF-null cells As an alternative differentiation model, HT-29 cells offer a simple system capable of differentiating through a change in growth medium from RPMI to DMEM (Figure 3.7). It was observed that the loss of XPF, especially in the presence of ADH5i, results in the abrogation of the goblet cell marker, MUC2, suggesting that differentiation is inhibited when key formaldehyde detoxifiers are absent (Figure 3.8). This initial evidence could be improved by employing flow cytometry or RT-qPCR to determine quantifiable levels of MUC2 during this process. What remains to be determined is whether complementation of XPF would rescue such a phenotype and, furthermore, if a nuclease-dead mutant of XPF does not. Following the induction of any DNA damage as cells attempt to differentiate could be achieved through immunofluorescence for γH2AX or 53BP1, and is an ongoing experiment.

A tuneable system for the generation of *in situ* formaldehdye This study provides a foundational approach for employing dCas9-fusion systems to enzymatically generate formaldehyde and other reactive aldehydes in proximity to DNA, offering a novel method to study the impact of endogenous aldehydes on DNA ICL repair and cellular stress. By leveraging the precise targeting ability of CRISPR-dCas9, this system aims to create a robust and physiologically relevant model for investigating aldehyde-induced DNA damage and its downstream consequences. The successful generation of LSD1-dCas9 and ADH1B-dCas9 inducible constructs highlights the potential for exploiting enzymatic activity to localise aldehyde generation. LSD1, a histone demethylase, was chosen for its natural role in formaldehyde release during histone modification, while ADH1B was considered for its ability to oxidise methanol to formaldehyde (Figure 3.9). The use of inducible systems minimises adaptive cellular responses to basal aldehyde levels, preserving the integrity of the experimental model. A critical observation was the absence of DNA damage upon untargeted expression of LSD1-dCas9, indicating that the system requires locus-specific sgRNA guidance to induce detectable DNA damage, such as γH2AX foci (Figure 3.11). However, γH2AX induction at the *PDXDC1* promoter and telomeres supports the hypothesis that local aldehyde concentrations can effectively disrupt DNA integrity. Confirmation that the observed DNA damage represents *bona fide* ICLs remains a critical avenue for future work. Immunofluorescence-based methods may provide additional evidence, particularly through colocalisation of γH2AX with markers such as

HMCES, EXO1, RPA, and MRE11, which are enriched at sites of ICL repair (unpublished observations, personal communication with Prof. Jean-Yves Masson). These readouts offer a practical approach for inferring ICL formation at a cellular level. Alternatively, a reporter assay could be developed to provide a more direct and functionally informative measure of ICL-induced repair. One proposed strategy involves adapting existing HR reporter systems[817], by integrating a dCas9-LSD1 target site within a disrupted GFP coding sequence. If an ICL forms within the target site, successful repair by HR using a homologous donor template would restore GFP expression, distinguishing ICL repair events from other forms of damage, such as protein-DNA crosslinks or base lesions, which are unlikely to result in functional gene correction. Such approaches would offer important mechanistic insight into the specificity and repair consequences of aldehyde-induced DNA damage in this system.

While expanding upon these preliminary studies for LSD1-dCas9 fusions warrants immediate follow-up, other avenues for future experiments can enhance and expand the utility of this system. Additional orthologues or mutant variants of ADH orthologues might create systems with higher specificity and efficiency in producing reactive aldehydes and reducing potential off-target effects. As a therapeutic avenue, investigating the ability of dCas9-fused enzymes to activate prodrugs near specific genomic loci could be explored, paving the way for targeted therapeutic interventions. Alternative regulation of dCas9-fusion expression, such as through optogenetics, might allow for higher spatial precision. It is foreseeable that genome-wide screens can be performed with this system to identify loci most sensitive to aldehyde-induced DNA damage, providing insights into vulnerable regions of the genome and their resolution by the FA pathway or ICL repair more generally.

Concluding remarks The study establishes three platforms for studying aldehyde-induced DNA damage and crosslink repair. While preliminary results validate two of these approaches (the HT-29 differentiation model and the inducible LSD1-dCas9 fusion system), expansion of the work on HT-29s and exploration of broader chemistries for LSD1-dCas9 are essential to realise the potential of these studies. The LSD1-dCas9 system not only provides a powerful tool for basic research but also holds promise for developing innovative therapeutic strategies that exploit targeted DNA damage. By bridging enzymatic chemistry with precise genomic targeting, this approach offers a new lens through which

to investigate the interplay between endogenous aldehydes or other DNA-damaging agents and DNA repair mechanisms.

4 | XPF and SLX1 genetic studies

Contents

The exploration of DNA repair mechanisms often begins with targeted genetic modifications to probe the roles of specific proteins and reveal vulnerabilities in their absence. This chapter focuses on efforts to create knockout and knock-in models for two pivotal DNA repair nucleases: XPF, the critical component of NER and ICL repair, and SLX1, a structure-specific endonuclease central to the resolution of DNA recombination intermediates[521]. These studies aim to uncover the consequences of nuclease loss, aiming to identify their genetic interactions and evaluate their potential as targets for therapeutic exploitation. By developing knockout models, this work intends to provide insight into the compensatory mechanisms and vulnerabilities associated with the absence of these nucleases, offering a framework for understanding their diverse roles in maintaining genome stability and addressing unresolved questions about their biological functions. The subsequent sections detail the rationale, methods, and findings surrounding the generation and analysis of XPF and SLX1 knockout cell lines, including their sensitivities to DNA-damaging agents and hypothesis-driven genetic dependencies. The study also establishes cell lines for the later undertaking of CRISPR-KO screens to determine synthetic sickness/lethality with SLX1 or XPF, for which some initial steps have been taken.

4.1 | XPF knockouts and functional studies

XPF-ERCC1 activity is most commonly associated with the repair of UV-induced bulky DNA adducts, such as cyclobutane pyrimidine dimers and 6,4-photoproducts, through its key role in the coordinated action of the nucleotide excision repair (NER) machinery (Section 1.3.3). The incision of DNA 5' to the lesion, mediated by XPF-ERCC1, precedes a complementary 3' incision by the related nuclease XPG, leading to the excision of the damaged DNA segment. This role underpins the alias XPF for the *ERCC4* gene, since, when mutated, it is most notably linked to xeroderma pigmentosum (XP), a condition characterised by heightened sensitivity to UV radiation, as well as increased prevalence of certain cancers.

Mutations in *ERCC4* or the gene encoding its heterodimeric partner *ERCC1* result in a broad and complex array of phenotypes. In addition to its role in NER, XPF-ERCC1 is also implicated in (ICL repair) repair and SSA repair of double-strand breaks[353]. Consequently, the functional loss of this nuclease complex has diverse and severe consequences, with human patients exhibiting XP, Cockayne syndrome, cerebro-oculo-facio-skeletal (COFS) syndrome, and/or Fanconi anaemia (FA), depending on the specific mutation[818]. As noted earlier (Section 3.1), the complete loss of *Ercc1* in mice mirrors the full spectrum of age-related phenotypes, including extreme progeria. The $Ercc1^{-/\Delta}$ genotype, associated with a drastically reduced lifespan of just a few months, manifests with profound neurodegeneration and other hallmark symptoms of premature ageing[776,777]. Using Cre-loxP technology, Niedernhofer and colleagues conditionally deleted *Ercc1* in specific tissues, revealing that localised loss of ERCC1 and XPF consistently induced symptoms of premature ageing across multiple organs.

Patient mutations in XPF are predominantly missense, though frameshift mutations resulting in truncated and destabilised protein variants have also been reported. In XP patients, these mutations often lead to diminished endonuclease activity, resulting in a milder phenotype primarily characterised by skin cancers in later life. However, rarer cases of XPF-null mutations have been linked to severe neurological complications. For example, a patient with XFE progeroid syndrome, CALIF1010, embodies rare reports of CS-like phenotypes without XP caused by biallelic mutations in *ERCC4*. Despite pronounced sensitivity to interstrand crosslinking agents, this patient did not exhibit UV

sensitivity[819]. The phenotype closely resembled that of *Ercc1*$^{-/-}$ or *Ercc4*$^{m/m}$ mouse models of progeria[820]. The identification of XPF as the defective factor in a subset of FA patients explains its designation as FANCQ, one of the 23 FA complementation groups. Similarly, SLX4 is designated as FANCP due to its critical role in recruiting XPF-ERCC1 to sites of interstrand crosslinks (ICLs) (Section 1.4.4). These findings highlight the multifaceted roles of XPF-ERCC1 in maintaining genomic stability and underscore its importance in a wide range of DNA repair pathways.

Homologs of XPF containing the ERCC4 fold have evolutionary origins dating back to early eukaryotes. Prokaryotes and yeast rely on rudimentary forms of NER to repair both bulky adducts and ICLs, indicating that XPF-like functions may have become specialised in metazoans to coordinate NER and FA pathway activities. Further studies are needed to elucidate the evolutionary changes in XPF domains and their impact on pathway regulation. Recent work by Mulderrig and Garaycoechea generated XPF-null HAP1 cells, providing a model to explore XPF functions. siRNA-mediated depletion of XPF has also been effective and has confirmed the heightened sensitivity of XPF-deficient cells to crosslinking agents such as acetaldehyde, mitomycin C, and cisplatin. Notably, this sensitivity exceeds that of cells lacking only NER (e.g., XPA-null), FA pathway components (e.g., FANCL-null), or both. An exception was observed with formaldehyde, where XPF-null cells exhibited sensitivity comparable to *Xpa*$^{-/-}$ *Fancl*$^{-/-}$ cells, suggesting that formaldehyde-induced lesions directly engage both pathways.

The enhanced sensitivity of XPF-null cells to crosslinking agents indicates that XPF-ERCC1 functions beyond its established roles in NER and the FA pathway. One possibility is that XPF-ERCC1 participates in multiple stages of FA-mediated ICL repair. Alternatively, the absence of XPF-ERCC1 may lead to deregulated activity of compensatory nucleases, resulting in lethal DNA damage. Another hypothesis involves a parallel, FA-independent pathway, potentially involving the nuclease SNM1A, which is epistatic with XPF-ERCC1. Emerging evidence also implicates XPF-ERCC1 in transcriptional regulation *via* CTCF recruitment, as XPF and XPG endonuclease activities are required for CTCF localisation to chromatin[821]. The interplay between XPF, NER, and transcription-coupled repair (TCR) pathways adds further complexity. For example, the combination of CSB and XPA deficiencies results in progeroid phenotypes, although XPA single mutants,

intrinsically defective for both global-genome and transcription-coupled NER, suggest the existence of a non-NER TCR pathway. This hypothetical pathway may contribute to resolving the sensitivity disparities observed in XPF-null cells, warranting further investigation. Testing the effects of *Ercc1*$^{-/-}$ and *Csb*$^{m/m}$ double mutants could provide insights into these interactions.

To better understand XPF-ERCC1 function and identify vulnerabilities, the creation of XPF-null cell lines is essential. These cell lines enable the investigation of drug sensitivities and synthetic lethal interactions, as well as the performance of CRISPR-Cas9 genetic screens to uncover pathway dependencies. Two strategies are particularly relevant for such screens: comparing knockout cells to wild-type cells to identify compensatory pathways, and comparing wild-type cells to those expressing enzymatically inactive XPF, which potentially more closely mimics pharmacological inhibition. Nuclease-inactive XPF variants, such as D705A, are valuable tools for these studies. The D705A mutation disrupts the conserved P$\underline{\text{D}}$X$_n$(E/D)XK catalytic motif and has been used extensively in biochemical assays[324]. However, engineering this mutation using CRISPR-Cas9 presents challenges due to the lack of proximal sgRNA targets and limitations of current base editing technologies. An alternative approach involves generating XPF-null cells and complementing them with nuclease-inactive XPF under a weak promoter to approximate physiological expression. This approach is constrained by potential cellular adaptation to XPF loss before complementation. The use of degrons for targeted XPF degradation offers a promising solution, but was beyond the scope of the current study. Despite these challenges, the generation of XPF-null cell lines represents a critical first step toward dissecting XPF-ERCC1 function and evaluating the therapeutic potential of targeting this complex in cancer treatment. A precedent exists for targeting nucleases in DNA repair, as highlighted by the development of MUS81 inhibitors[822], underscoring the rationale for pursuing similar strategies for XPF inhibition.

Efforts to inhibit this structure-specific endonuclease have explored diverse strategies, including direct targeting of its catalytic activity and disruption of essential protein-protein interactions. For example, studies have identified small-molecule inhibitors that impair XPF-ERCC1 endonuclease activity, enhancing the efficacy of DNA-damaging agents such as cisplatin[823,824]. Computational approaches have also facilitated the design of inhibitors

that destabilise the interaction between ERCC1 and XPF, as well as between ERCC1 and XPA, which is critical for the NER pathway[825]. Despite these advancements, challenges remain in targeting ERCC1-XPF selectively due to its dual roles in genome maintenance. Broad inhibition of the complex risks impairing both NER and ICL repair, potentially exacerbating genomic instability and toxicity. Consequently, there is growing interest in strategies that achieve a separation of function, selectively disrupting XPF's role in one repair pathway while preserving others. Such approaches aim to balance therapeutic efficacy with the minimisation of off-target effects, offering a more nuanced means of enhancing the activity of DNA-damaging agents in cancer treatment.

4.1.1 | Knockouts of XPF in multiple cell lines

Previous work in the group had already established U2OS Δ*ERCC4* cells. These were made by sgRNAs targeting exon 1 - one cutting ∼31 bp into the first exon of *ERCC4*, and the other in the first intron to delete a fragment of 168 bp (Figure 4.1a). With these same sgRNAs, *ERCC4* was targeted in HT-29 and A549 cells, and two clones were recovered that were characterised and used in this work (knockout generation was completed by Dr. Ceren Yalaz). These XPF-null clones were verified (and re-verified for U2OS Δ*ERCC4*) by PCR, where the same deletion was observed for the three cell lines (Figure 4.1b). Sanger sequencing also revealed all clones contained the same deletion, of 169 bp, which leads to a frameshift, causing premature termination at codon 20 (Figure 4.1d). XPF deletion was observed in these knockout clones and resulted in the expected instability of ERCC1 (Figure 4.1c).

4.1.2 | Sensitivities of XPF-null cells

The response of XPF-null cells to cisplatin, UV and MMC has been studied extensively, effectively describing the range of XPF function in NER and ICL repair[826]. To further validate these knockouts, and also understand the impact of XPF's contribution to the repair of formaldehyde damage and replication stress, cell viability assays were performed in the presence of increasing concentrations of formaldehyde, olaparib and SJG-136. These agents are suspected to generate damage that engages the FA pathway and ICL repair or replication fork stability. It was considered that HT-29 cells may display more divergent

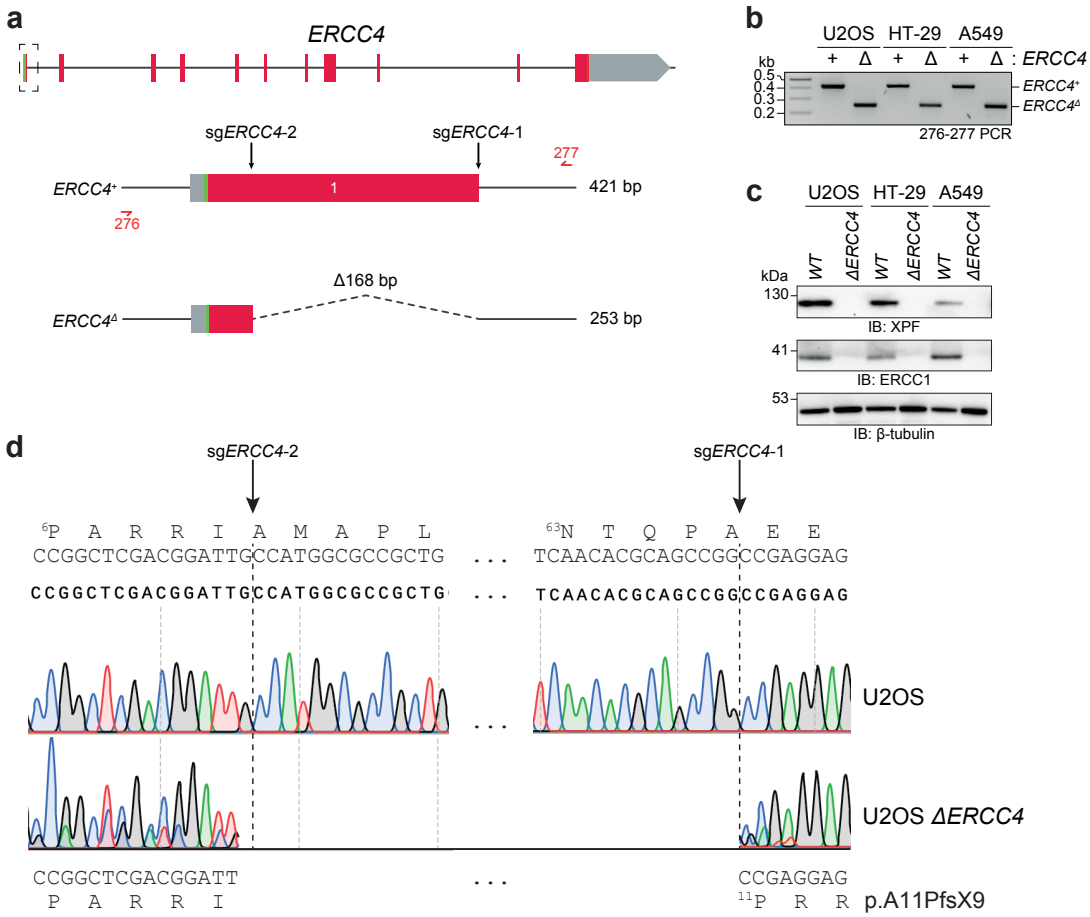


Figure 4.1 | Generation and validation of *ERCC4* knockouts in human cell lines. a, Schematic of the *ERCC4* gene highlighting the targeted region using two sgRNAs (sg*ERCC4*-2 and sg*ERCC4*-1) designed to induce a 168 bp deletion within exon 1. The wild-type (*ERCC4*$^{+}$) allele is depicted with its original sequence, while the edited allele (*ERCC4*$^{\Delta}$) shows the excised region. **b,** PCR validation of the *ERCC4*$^{\Delta}$ allele in U2OS, HT-29, and A549 cells. Products from the 276-277 PCR primer set confirm the presence of the 421 bp fragment in wild-type cells and the 253 bp fragment in knockout cells. **c,** Western blot analysis of XPF and ERCC1 protein levels in wild-type (WT) and Δ*ERCC4* cells. **d,** Representative Sanger sequencing chromatograms of the *ERCC4* target region in U2OS wild-type and Δ*ERCC4* cells. The deletion event in Δ*ERCC4* cells results in a 169 bp deletion that creates a frameshift mutation, leading to a premature stop at codon 20.
Note: U2OS CRISPR-Cas9 knockout cells were generated by Dr. Sanja Brolih. A549 and HT-29 CRISPR-Cas9 knockout cells were generated by Dr. Ceren Yalaz.

phenotypes compared to A549 and U2OS, which are both p53-proficient, while HT-29 cells harbour an inactivating R273H mutation in p53[827]. Expectedly, the Δ*ERCC4* cells were sensitive to formaldehyde, olaparib, and SJG-136 in U2OS and HT-29 (Figure 4.2b-g, Figure 4.3b-g) although they were only sensitive to formaldehyde in A549 (Figure 4.4b-g). However, in the latter's case, the substantial variability observed in the data likely precludes the identification of statistically robust biological interpretations, although the overall pattern of sensitivity can be seen (Figure 4.4b-g). A colony survival assay was

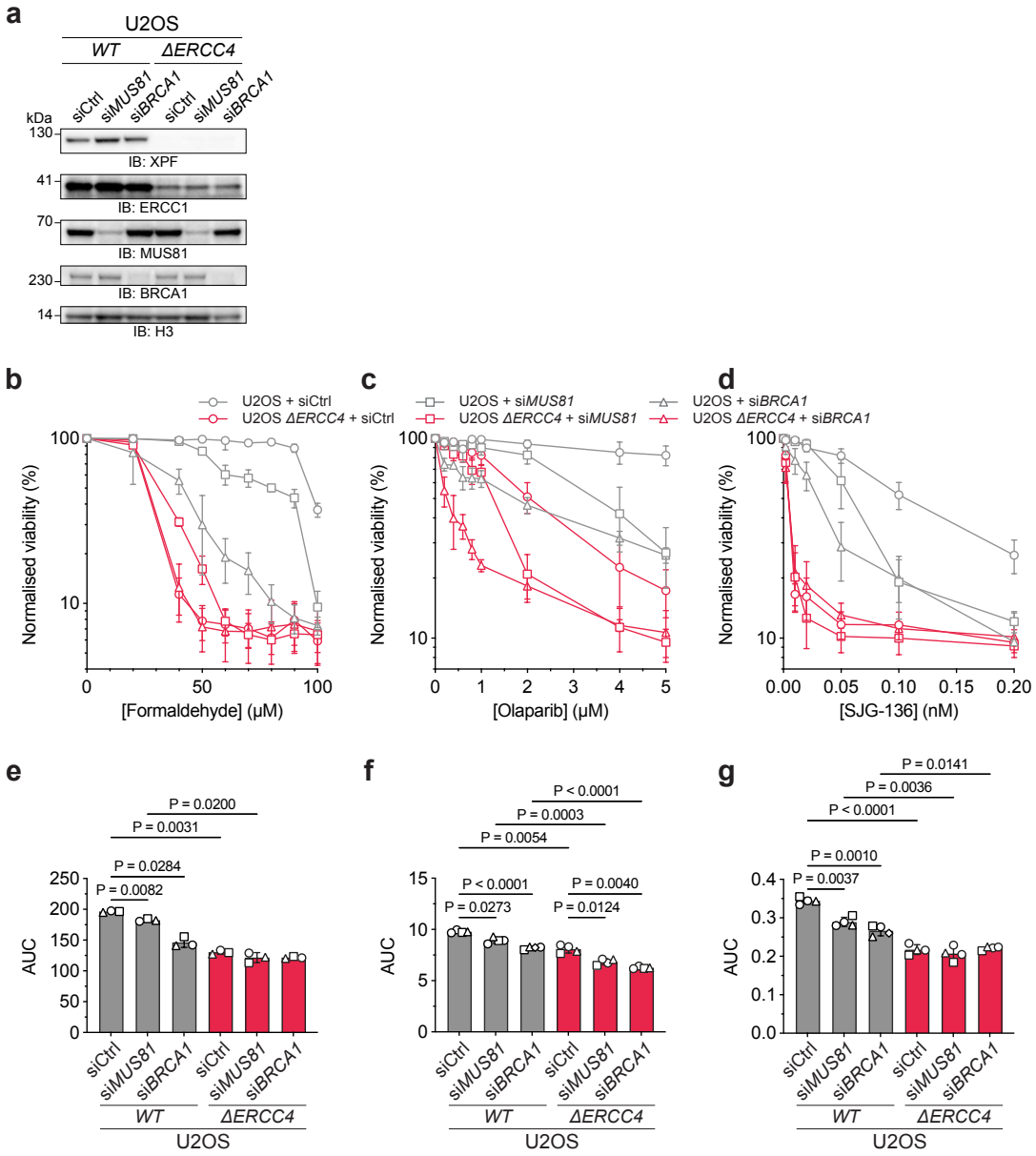


Figure 4.2 | Survival analysis of U2OS Δ*ERCC4* cells following siRNA depletion of MUS81 or BRCA1 and treatment with DNA damaging agents. **a,** Western blot analysis confirming siRNA-mediated depletion of MUS81 and BRCA1 in wild-type (WT) and Δ*ERCC4* U2OS cells. **b, c, d,** Survival of the indicated Δ*ERCC4* U2OS cells grown for 8 days in the presence of the indicated doses of formaldehyde (**b**), olaparib (**c**) and SJG-136 (**d**). Viability by resazurin is shown as the mean ± s.d. percentage relative to untreated cells for that condition, for n ≥ 3 biological experiments of n ≥ 4 technical repeats. **e, f, g,** The area under curve (AUC) for each condition in **b**, **c**, **d** is represented as mean ± s.d. of n ≥ 3 biological experiments (indicated by number of points). Point shape represents matched experiments in order; circle, square, triangle, diamond. Statistical significance was determined by Brown-Forsythe and Welch ANOVA tests, with post-hoc Dunnett's T3 test for multiple comparisons. Comparisons made as follows: For each siRNA treatment, WT vs Δ*ERCC4*; for each WT or Δ*ERCC4* cell line, siCtrl vs siMUS81 and siCtrl vs siBRCA1 and siMUS81 vs siBRCA1.

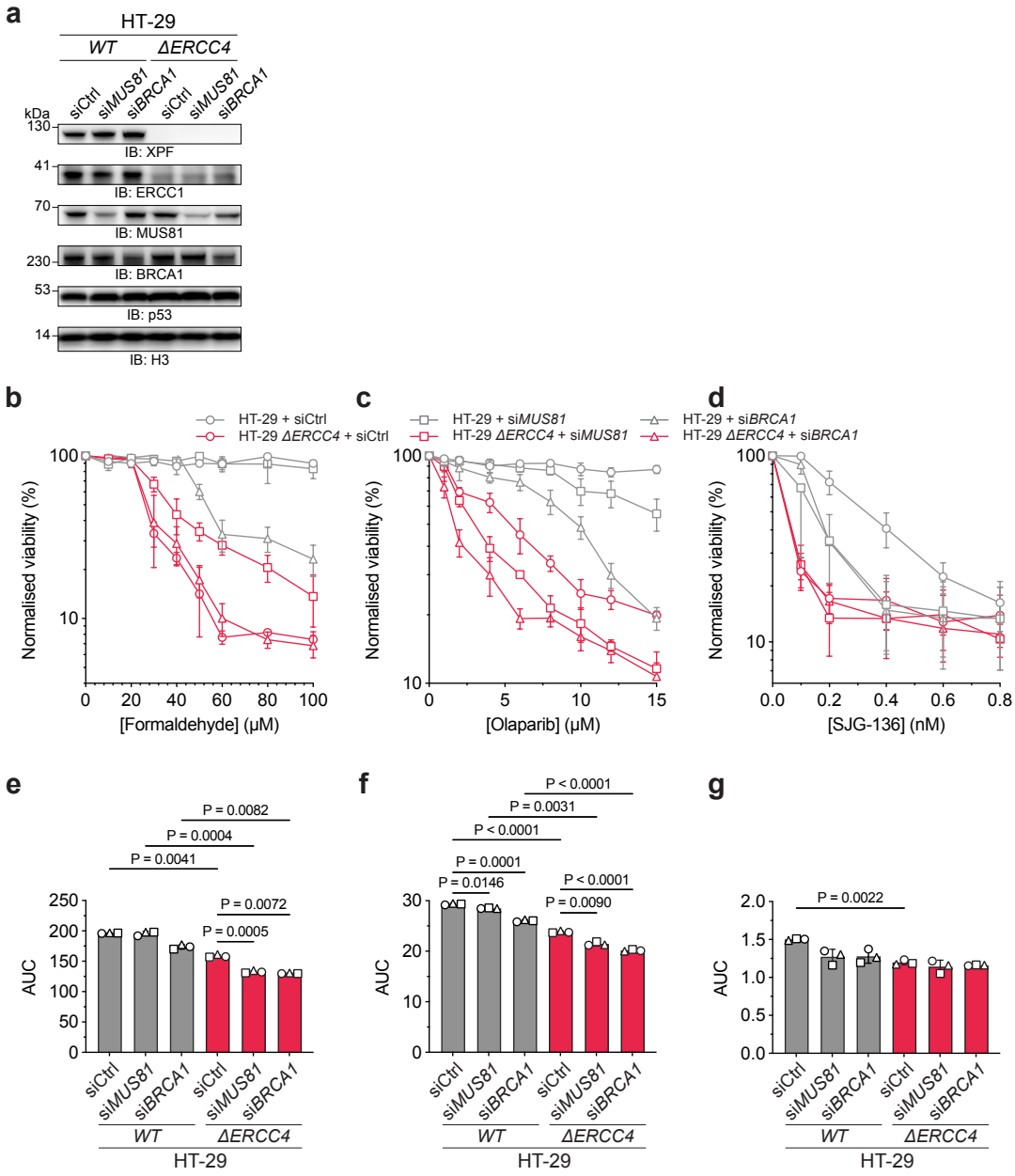


Figure 4.3 | Survival analysis of HT-29 Δ*ERCC4* cells following siRNA depletion of MUS81 or BRCA1 and treatment with DNA damaging agents. **a,** Western blot analysis confirming siRNA-mediated depletion of MUS81 and BRCA1 in wild-type (WT) and HT-29 Δ*ERCC4* cells. **b, c, d,** Survival of the indicated HT-29 Δ*ERCC4* cells grown for 8 days in the presence of the indicated doses of formaldehyde (**b**), olaparib (**c**) and SJG-136 (**d**). Viability by resazurin is shown as the mean ± s.d. percentage relative to untreated cells for that condition, for n ≥ 3 biological experiments of n ≥ 4 technical repeats. **e, f, g,** The area under curve (AUC) for each condition in **b**, **c**, **d** is represented as mean ± s.d. of n ≥ 3 biological experiments (indicated by number of points). Point shape represents matched experiments in order: circle, square, triangle, diamond. Statistical significance was determined by Brown-Forsythe and Welch ANOVA tests, with post-hoc Dunnett's T3 test for multiple comparisons. Comparisons made as follows: For each siRNA treatment, WT vs Δ*ERCC4*; for each WT or Δ*ERCC4* cell line, siCtrl vs siMUS81 and siCtrl vs siBRCA1 and siMUS81 vs siBRCA1.

therefore performed to confirm whether this pattern of sensitivity is due to a particular quirk of XPF biology in A549 cells. It was suspected that genetic interactions between MUS81 or BRCA1 could influence some of the differential sensitivities of Δ*ERCC4* cells, and performed these cell viability assays after depletion of MUS81 or BRCA1 and compared these results to siCtrl after confirmation of depletion by Western blotting (Figure 4.2a, Figure 4.3a, Figure 4.4a). In chicken DT40 cells, XPF is essential, with complete loss resulting in lethality[543]. Notably, DT40 cells lack an orthologue for MUS81, suggesting that the essentiality of XPF in this model may reflect a compensatory role in processing replication-associated DNA damage normally resolved by MUS81. Supporting this, expression of human MUS81 in DT40 cells can rescue XPF loss, indicating overlapping or compensatory functions between these nucleases[543]. This implies that the synthetic relationship between XPF and MUS81 may be relevant in human cells as well, particularly under conditions of genotoxic stress. For this analysis, it is more suitable to look at individual cell lines at a time. In U2OS cells, depletion of MUS81 or BRCA1 sensitises WT cells to formaldehyde and SJG-136 but does not further increase sensitivity in Δ*ERCC4* cells (Figure 4.2b, d, e, g). While XPF loss sensitises cells to olaparib, it is exacerbated by the depletion of MUS81 or BRCA1 (Figure 4.2c, f), which is recapitulated in HT-29 cells (Figure 4.3c, f). Increased sensitivity to SJG-136 upon XPF loss was observed for both U2OS (Figure 4.2d, g) and HT-29 (Figure 4.3d, g) cells, but could not be statistically interpreted for A549 cells (Figure 4.4d, g).

4.1.2.1 | Response of XPF-null cells to trabectedin

Trabectedin is a uniquely promising therapeutic agent that differs from conventional DNA-targeting drugs by exhibiting greater toxicity in cells with functional DNA repair pathways[828,829]. This counterintuitive mechanism presents potential for targeting chemoresistant or repair-proficient tumour populations during chemotherapy or radiotherapy. Derived from the sea squirt *Ecteinascidia turbinata*, trabectedin binds the minor groove of DNA and forms covalent adducts at the N2 position of guanine. Notably, cells deficient in transcription-coupled nucleotide excision repair (TC-NER) exhibit resistance to trabectedin, implying that the lesion becomes more cytotoxic upon recognition and processing by TC-NER[446,830]. Trabectedin induces a shallow bend in the DNA duplex and is preferentially

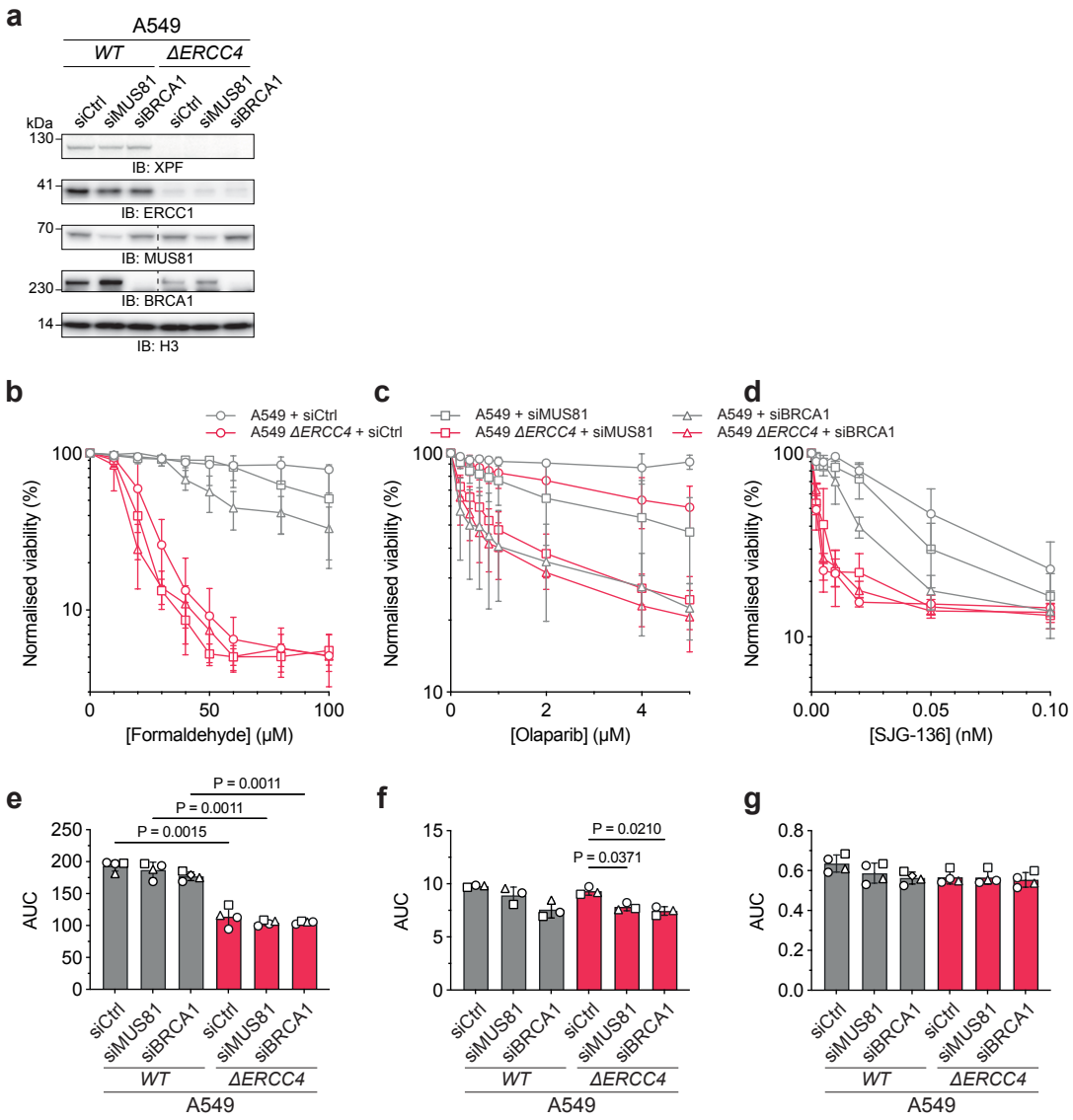


Figure 4.4 | Survival analysis of A549 Δ*ERCC4* cells following siRNA depletion of MUS81 or BRCA1 and treatment with DNA damaging agents. **a,** Western blot analysis confirming siRNA-mediated depletion of MUS81 and BRCA1 in wild-type (WT) and A549 Δ*ERCC4* cells. **b, c, d,** Survival of the indicated A549 Δ*ERCC4* cells grown for 8 days in the presence of the indicated doses of formaldehyde (**b**), olaparib (**c**) and SJG-136 (**d**). Viability by resazurin is shown as the mean ± s.d. percentage relative to untreated cells for that condition, for n ≥ 3 biological experiments of n ≥ 4 technical repeats. **e, f, g,** The area under curve (AUC) for each condition in **b**, **c**, **d** is represented as mean ± s.d. of n ≥ 3 biological experiments (indicated by number of points). Point shape represents matched experiments in order: circle, square, triangle, diamond. Statistical significance was determined by Brown-Forsythe and Welch ANOVA tests, with post-hoc Dunnett's T3 test for multiple comparisons. Comparisons made as follows: For each siRNA treatment, WT vs Δ*ERCC4*; for each WT or Δ*ERCC4* cell line, siCtrl vs siMUS81 and siCtrl vs siBRCA1 and siMUS81 vs siBRCA1.

processed by TC-NER rather than global genome NER (GG-NER), likely due to its ability to stabilise the DNA helix and impede GG-NER recognition[831]. A recent mechanistic study showed that trabectedin lesions are processed by XPF but not XPG during TC-NER, enabling the first mapping of XPF-specific incisions at these sites[832]. These findings suggest XPF may exhibit minor sequence specificity and that single-strand breaks (SSBs) generated by XPF contribute significantly to trabectedin's cytotoxicity. Accumulated SSBs may resemble or be converted into double-strand breaks (DSBs) during replication, especially when DSB repair is impaired.

However, the precise mechanism of trabectedin resistance in TC-NER-deficient cells remains unresolved. It is unclear whether lesions are tolerated or redirected into alternative DNA damage response (DDR) pathways in the absence of XPF. To investigate this, viability assays were performed examining the effect of trabectedin in cells depleted of MUS81 or BRCA1 in the context of XPF loss, using A549 and U2OS cell lines. When TC-NER engages a trabectedin lesion, the resultant cytotoxic intermediate may stall transcription and, during S-phase, lead to replication stress. This, in turn, could necessitate the involvement of homologous recombination (HR) and structure-specific endonucleases such as MUS81 and XPF. Thus, testing whether MUS81 or BRCA1 loss exacerbates trabectedin sensitivity in XPF-deficient cells could clarify whether these factors function in a shared or compensatory repair axis bridging TC-NER and HR. These experiments aimed to elucidate whether alternative repair mechanisms mitigate trabectedin-induced damage in the absence of XPF. Attempts to extend these studies into HT-29 cells were limited by narrow therapeutic windows of trabectedin, which made it difficult to resolve differential sensitivities between wild-type and XPF-deficient cells. As a result, HT-29 data were excluded from the analysis.

Consistent with the transcription-coupled nucleotide excision repair (TC-NER)-specific toxicity of trabectedin, loss of XPF conferred resistance to trabectedin, irrespective of the depletion of MUS81 or BRCA1 (Figure 4.5a-d). It was hypothesised that BRCA1 depletion in wild-type (WT) cells, but not in Δ*ERCC4* cells, would increase sensitivity to trabectedin due to the accumulation of single-strand breaks (SSBs) requiring homologous recombination (HR) in the subsequent S phase. However, this effect was not observed in either A549 or U2OS cell lines (Figure 4.5a-d). Instead, in both cell lines, BRCA1 depletion—and MUS81 depletion in the case of A549 cells—significantly reduced viability

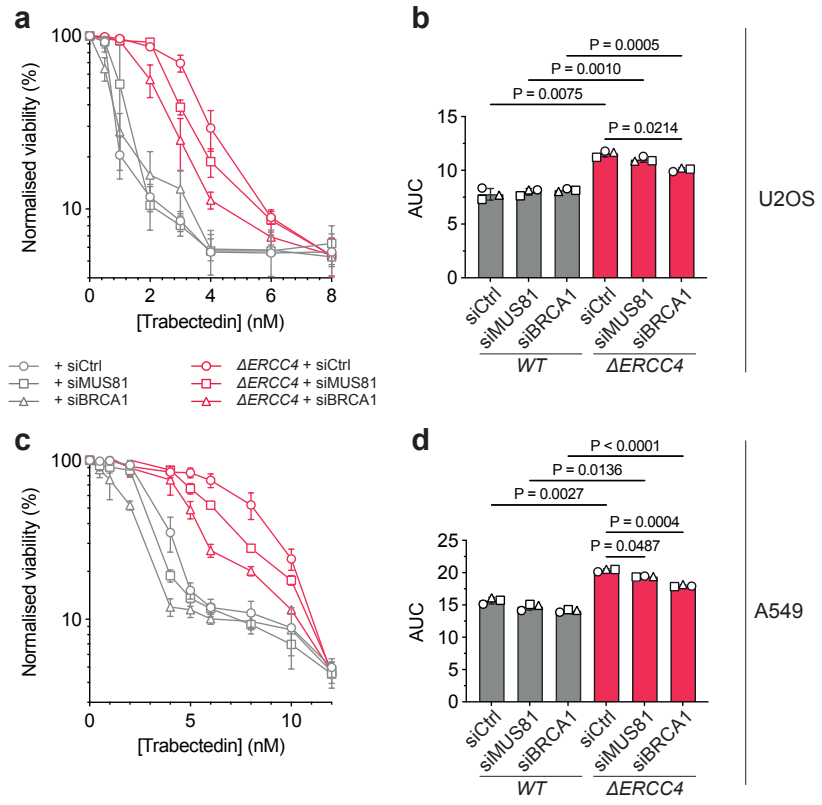


Figure 4.5 | Survival analysis of U2OS and A549 Δ*ERCC4* cells following siRNA depletion of MUS81 or BRCA1 and treatment with trabectedin. a, c, Survival of U2OS (**a**) and A549 (**c**) WT and Δ*ERCC4* cells grown for 8 days in the presence of the indicated doses of trabectedin. Viability by resazurin is shown as the mean ± s.d. percentage relative to untreated cells for that condition, for n = 3 biological experiments of n ≥ 4 technical repeats. **b, d,** The area under the curve (AUC) for each condition in **a** and **c** is represented as mean ± s.d. of n = 3 biological experiments (indicated by number of points). Point shape represents matched experiments in order: circle, square, triangle. Statistical significance was determined by Brown-Forsythe and Welch ANOVA tests, with post-hoc Dunnett's T3 test for multiple comparisons. Comparisons made as follows: For each siRNA treatment, WT vs Δ*ERCC4*; for each WT or Δ*ERCC4* cell line, siCtrl vs siMUS81 and siCtrl vs siBRCA1 and siMUS81 vs siBRCA1.

in Δ*ERCC4* cells in response to trabectedin (Figure 4.5b, d). These results suggest that alternative mechanisms or interactions may underlie trabectedin resistance and toxicity in the context of XPF deficiency.

While these low-throughput genetic interaction studies and observations of toxicity to DNA damage-related agents provide initial insights into XPF activity, further detailed mechanistic studies are required to fully elucidate these pathways. These experiments serve as a precursor to a genome-wide CRISPR-KO screen aimed at identifying genetic interactors of XPF. Although the CRISPR-KO screen was conducted, technical challenges in the high-throughput sequencing of surviving sgRNAs prevented the identification of *ERCC4*-interacting genes. A repeat of library preparation and sequencing is currently

in progress to resolve these issues.

4.2 | Genetic disruption of SLX1

The SLX1 protein is encoded by two genes within humans, *SLX1A* and *SLX1B*. These genes reside on chromosome 16, 762 kb apart (hereafter, reference to the gene *SLX1* includes both genes). SLX1, at least *in vitro*, is a promiscuous nuclease with almost no apparent preference for any defined DNA structure, despite historical characterisation as a 5′-flap endonuclease of the human structure-selective endonucleases[585]. Beyond implications of functions in concert with or redundant with MUS81 and XPF, *via* its interaction with SLX4, only a handful of SLX1-unique responsibilities have been described. Beyond HJ resolution (Section 1.4.3, some studies point to SLX1's involvement in the initiation of MiDAS at CFSs[638], or rare fragile sites such as at the *FRAXA* locus[639]. Disruption of *Slx1* in mice does not result in lethality[540]. Both $Slx1^{+/-}$ and $Slx1^{-/-}$ mice are born at Mendelian frequencies, exhibit no significant developmental defects, and are fertile[540]. In this study, SLX1 protein was undetectable in mouse embryonic fibroblast (MEF) extracts from these knockout mice. The targeted Cre-loxP-dependent deletion of exon 1 (Figure 4.6a), potentially coupled with nonsense-mediated decay (NMD), is expected to ablate SLX1 translation. However, an alternative translation initiation site within exon 2 (methionine at position 62, M62) could theoretically yield a truncated 23.6 kDa form of SLX1 (Figure 4.6a). This hypothetical product could escape detection on Western blots due to overlapping nonspecific bands observed with several SLX1 antibodies tested (Figure 4.6b).

Phenotypically, SLX1-null MEFs exhibited partial sensitivity to mitomycin C (MMC) and nitrogen mustard (HN-2), phenotypes that were rescued by overexpression of wild-type (WT), but not nuclease-inactive, SLX1[540]. If a truncated form of SLX1 were expressed in these cells, it could potentially act as a dysfunctional hypomorph, consistent with the observed phenotypes. To date, the literature indicates the generation of only one SLX1 knockout cell line. In Chan *et al.* (2017[833]), a sgRNA (`GAGCTTGTTCCGAAGCAAGC`) targeted exon 1 of SLX1 in Flp-In T-Rex 293 cells. Sequencing revealed clones with an 11 bp deletion, an 8 bp deletion, or a 73 bp insertion (Figure 4.6a). Given the triplication of chromosome 16 in the parental 293 cells[834], these clones likely contained

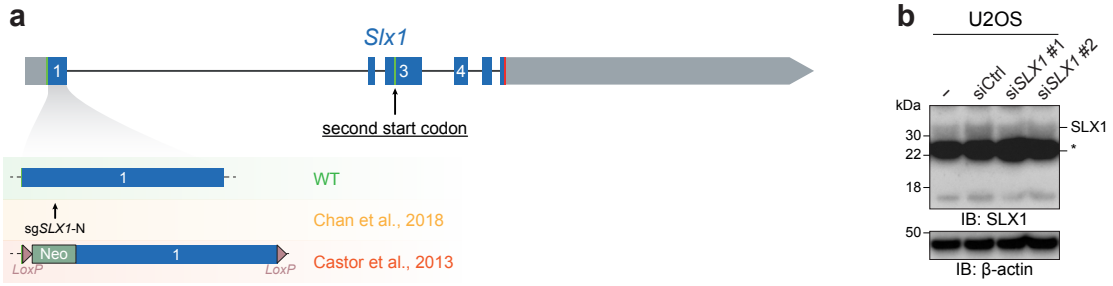


Figure 4.6 | Mouse *Slx1* gene structure and non-specific bands in SLX1 Western blots. a, Schematic representation of the mouse *Slx1* gene showing exons (blue boxes) and a possible second translation start site, beyond exon 1, as targeted by Cre-loxP deletion in Castor *et al.*[540] and sgRNAs in Chan *et al.*[833]. **b,** Western blot analysis of SLX1 protein levels in U2OS cells following siRNA-mediated depletion using two independent siRNAs. The band corresponding to SLX1 is indicated, while a non-specific band present through immunoblotting by most antibodies is also highlighted (*).

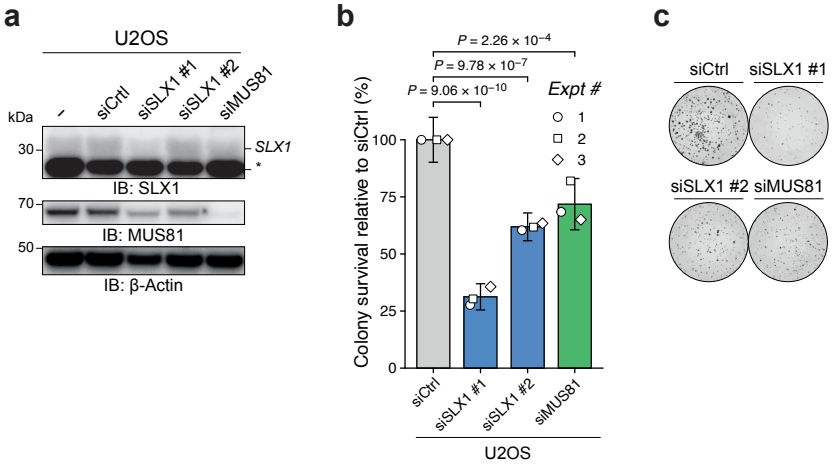


Figure 4.7 | SLX1 knockdown reduces colony survival in U2OS cells. a, Western blot analysis confirming siRNA-mediated depletion of SLX1 (two independent siRNAs) and MUS81 in U2OS cells. **b,** Quantification of colony survival relative to siCtrl-treated cells, as mean ± s.d. from three independent experiments (indicated as circle, square, and diamond for Expt #1, #2, and #3, respectively). Statistical significance determined by one-way ANOVA with post-hoc Dunnett's test. **c,** Representative images of colony formation assays for siCtrl, siSLX1 (#1 and #2), and siMUS81 conditions.

six alleles. *In silico* simulations of these alleles suggested potential dominant-negative effects due to truncated proteins. The 11 bp and 8 bp deletions were predicted to produce truncated products of approximately 8.2 and 8.3 kDa, respectively, while the 73 bp insertion (sequence unspecified) might yield an 11.5 kDa mutant product. Despite these predictions, one clone used for further studies confirmed the absence of SLX1 by Western blot. This clone demonstrated no synthetic lethality upon depletion of GEN1. However, concurrent MUS81 and GEN1 depletion resulted in G2 arrest and a pronounced survival defect[833].

4.2.1 | Depletion of SLX1 associates with sickness where p53 is functional

If the depletion of SLX1 described in previous literature is incomplete due to expression of a hypomorphic gene product, its non-essentiality is worth re-addressing. An initial approach, therefore, involved the depletion of SLX1 through RNA interference (RNAi). During the initial stages of this project, no suitable antibody against SLX1 was available; trials of α-SLX1 antibodies from MRC PPU in Dundee (S778C sheep) and various Abcam α-SLX1 antibodies (ab68807, ab182501) were attempted. However, later, it was found that the Proteintech αSLX1 antibody gave the clearest SLX1 signal by Western blot and was adopted for future experiments. Initial knockdown trials in U2OS using two siRNAs against SLX1 were, at the protein level, inconclusive as to SLX1 knockdown status, especially given the quicker-exposing non-specific bands (Figure 4.7a). Nevertheless, si*SLX1* #1 appeared most effective in SLX1 knockdown (Figure 4.7a). Compared to the addition of a non-targeting siRNA control (siCtrl), SLX1, and to a lesser extent MUS81, knockdown resulted in reduced colony outgrowth (Figure 4.7b, c). When improved Western blotting procedures for SLX1 were later adopted, SLX1 depletion using si*SLX1* #1 was apparent in three cell lines tested, RPE-1, RPE-1 *TP53*$^{-/-}$ and HAP1 (Figure 4.8a). Colony outgrowth was markedly reduced when SLX1 was depleted in RPE-1 (Figure 4.8b, left panel, Figure 4.8c), and to a lesser extent in HAP1 (Figure 4.8b, right panel, Figure 4.8c), but no change was visible for RPE-1 *TP53*$^{-/-}$ cells (Figure 4.8b, middle panel, Figure 4.8c). These relationships link SLX1 depletion to p53-mediated cell death, implicating SLX1 in essential homeostatic processes. To confirm the effective depletion, the colony survival of si*SLX1*-depleted cells with increasing MMC was tested. No notable difference was observed for MMC sensitivity in SLX1-depleted cells (Figure 4.8d).

The observed loss in fitness of SLX1-depleted cells suggests a critical role for SLX1 in the viability of cancer cell lines. Notably, SLX1 is absent from the CRISPR DepMap database[750], likely due to its omission from the TKOv3 lentiviral CRISPR-KO library[755]. One potential explanation for the absence of *SLX1* data in CRISPR-KO screens is that SLX1 is essential enough to result in the retention of viable sgRNAs, thus precluding its detection in knockout experiments. However, another plausible factor is the presence of an ancestral segmental duplication in human chromosome 16, resulting in *SLX1* being

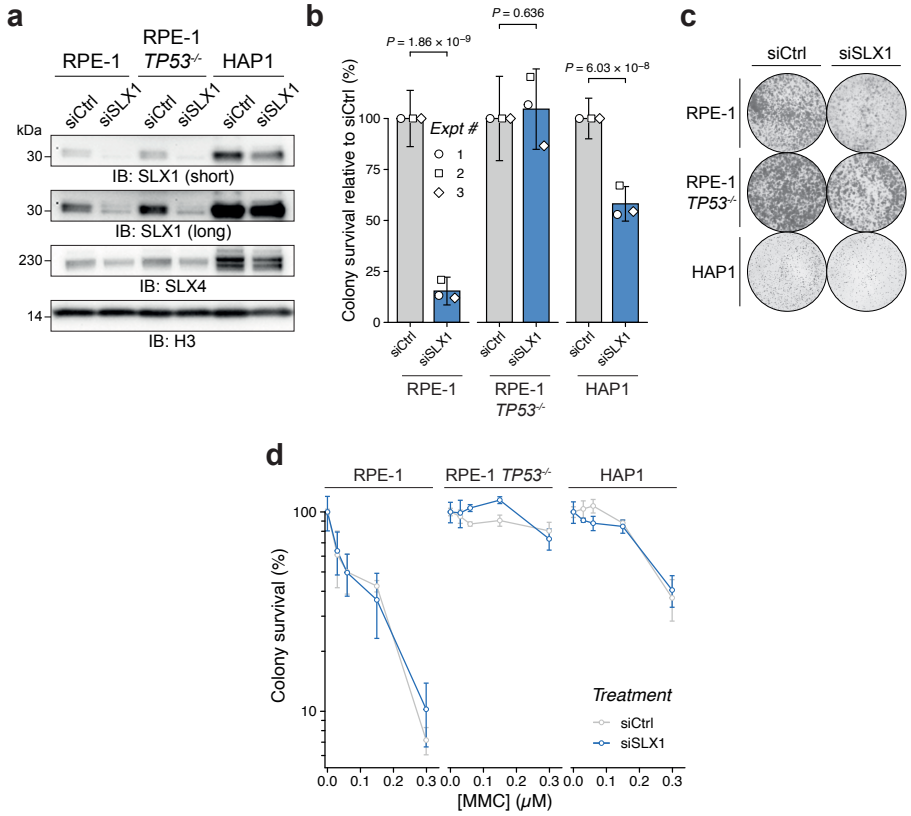


Figure 4.8 | SLX1 knockdown impacts colony survival in RPE-1, RPE-1 *TP53*$^{-/-}$, and HAP1 cells. a, Western blot analysis confirming siRNA-mediated depletion of SLX1 in RPE-1, RPE-1 *TP53*$^{-/-}$, and HAP1 cells. **b,** Quantification of colony survival relative to siCtrl-treated cells, as mean ± s.d. from n = 3 independent experiments (indicated as circle, square, and diamond). Statistical significance was determined by one-way ANOVA with post-hoc Dunnett's test, with comparisons between siCtrl and si*SLX1* for each cell line. **c,** Representative images of colony formation assays for siCtrl and siSLX1-treated conditions across cell lines as in **c**. **d,** Colony survival dose-response curves of RPE-1, RPE-1 *TP53*$^{-/-}$, and HAP1 cells treated with mitomycin C (MMC) following siCtrl or siSLX1 transfection. Data are presented as mean ± s.d. from n = 3 independent experiments.

encoded by two highly homologous genes: *SLX1A* and *SLX1B*. The two paralogues exhibit an extraordinary 99.7% sequence homology, even when analysed using long-read nanopore sequencing approaches (Figure 4.9a). This remarkable similarity highlights a potential selective evolutionary advantage for the preservation of these duplicates. However, the genes are located within a 600 kb cluster known as the BP4–BP5′ region, which includes other near-identical gene pairs such as *BOLA2A/B* and *SULT1A3/4*. This genomic region is characterised by significant sequence homology and exhibits copy-number variations that have been linked to autism spectrum disorder[835]. With near-perfect homology, most algorithms that generate sgRNA libraries for their eventual use in screens would exclude sgRNAs in this region due to the perception that binding to either paralogue locus suggests 'off-target' binding. Indeed, like *SLX1* genes, the paralogues *SULT1A3* and *SULT1A4*, and

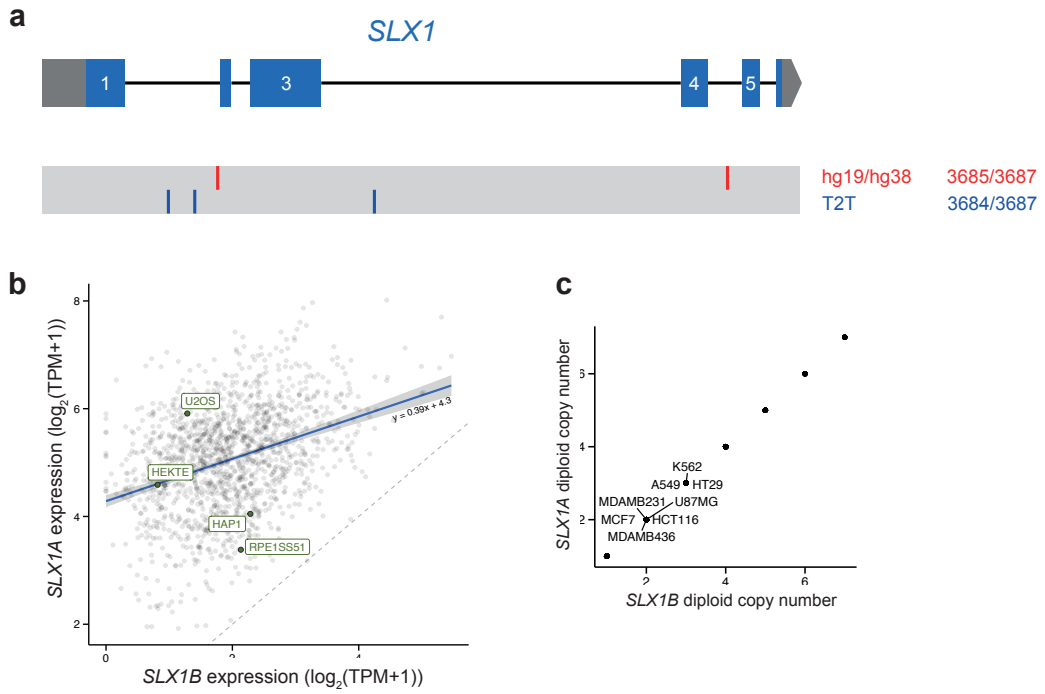


Figure 4.9 | **Analysis of *SLX1* gene structure, expression, and copy number. a,** Schematic representation of the human *SLX1* gene, highlighting exons (blue boxes), intronic regions, and genomic differences in hg19/hg38 and T2T assemblies. Positions of the observed single-nucleotide variants between *SLX1A* and *SLX1B* for each genome assembly are indicated. **b,** Correlation of *SLX1A* and *SLX1B* transcript expression ($\log_2$(TPM+1)) across CCLE cell lines, obtained from the CRISPR DepMap[750]. In-house cell lines are labelled. **c,** Correlation of *SLX1A* and *SLX1B* diploid copy numbers across CCLE cell lines. In-house cell lines are labelled.

NPIPB12 and *NPIPB13* are also totally absent in any of the CRISPR DepMap models. However, of the *BOLA2* paralogues, *BOLA2A*, but not *BOLA2B*, appears in the CRISPR GeCKO model, while SMG1P2 and SMG1P5 appear in the RNAi model within CRISPR DepMap. Nevertheless, chronic loss of SLX1 might confer some toxicity in certain cell lines, yet the mechanisms that govern progress to eventual cell death are largely unexplored.

4.2.2 | Knockout of SLX1

Although potential challenges to cell viability could complicate efforts to generate SLX1-null cells, initial attempts focused on deleting SLX1 in various cell lines before considering a more complex degron-based approach. To identify suitable cell lines for this purpose, CRISPR DepMap data were consulted, evaluating both SLX1 expression levels and absolute copy number. Analysis revealed that the expression of *SLX1A* consistently exceeded that of *SLX1B* across all tested cell lines (Figure 4.9b), with a notable 4.3

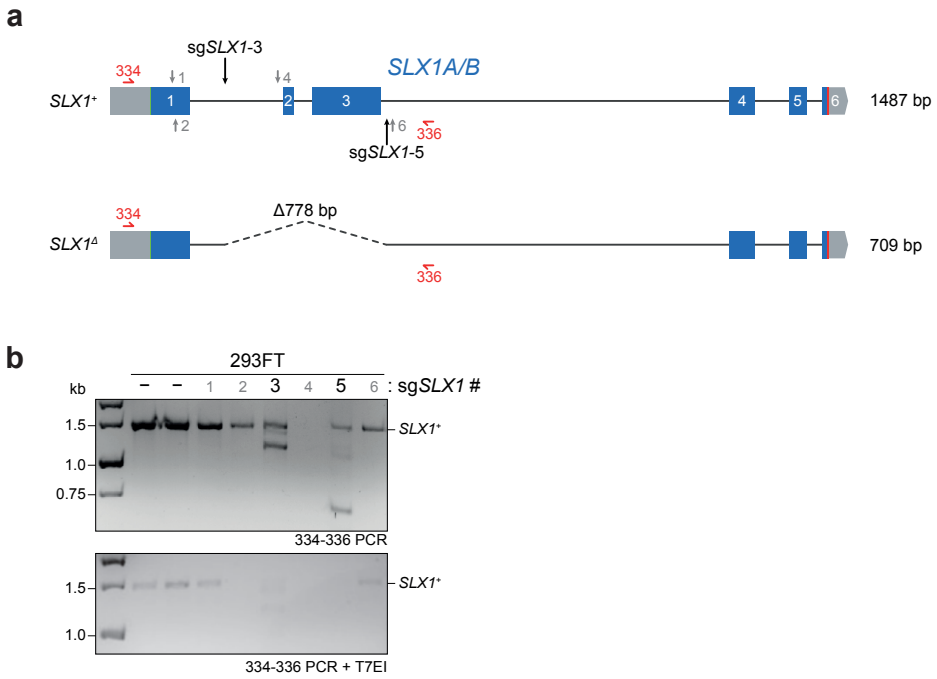


Figure 4.10 | Generation and screening of *SLX1* knockout clones across multiple cell lines. a, Schematic representation of the *SLX1A/B* loci, indicating the positions of sgRNAs (sg*SLX1*-1 - 6) and the expected deletion (*SLX1*Δ). PCR amplicons used for screening (*SLX1** and *SLX1*Δ) are indicated with their expected product sizes. **b,** SURVEYOR assay analysis to determine editing efficiency of sgRNAs targeting *SLX1* in 293FT cells, with the expected 1.5 kb amplicon cleaved by T7 endonuclease I (T7EI) where editing has occurred. Input for the T7EI reaction is shown in the upper panel.

$\log_2$(TPM+1) difference. This disparity is likely attributable to a bias in sequencing read alignment favouring *SLX1A* over *SLX1B*. However, this bias was not observed for other duplicated genes within the same genomic region, such as *NPIPB12 versus NPIPB13* or *SULT1A4 versus SULT1A3* (data not shown). Despite the alignment bias, *SLX1* expression was detectable in multiple in-house cell lines (Figure 4.9b). However, copy number data were available for only a subset of these lines. Cell lines A549 and HT29 were initially considered but ultimately excluded due to the presence of three copies of each *SLX1* gene, requiring a total of six editing events to achieve complete knockout (Figure 4.9c). To address these challenges, RPE-1, U2OS, 293FT, and HAP1 cell lines were selected for subsequent experiments. RPE-1, a non-cancerous model predicted to be diploid, was chosen to ensure a manageable number of editing targets. 293FT was selected for its high transfection efficiency, facilitating genome editing. Finally, HAP1, predicted to have a single copy of each *SLX1A* and *SLX1B*, was prioritised to maximise the likelihood of achieving complete SLX1 knockout.

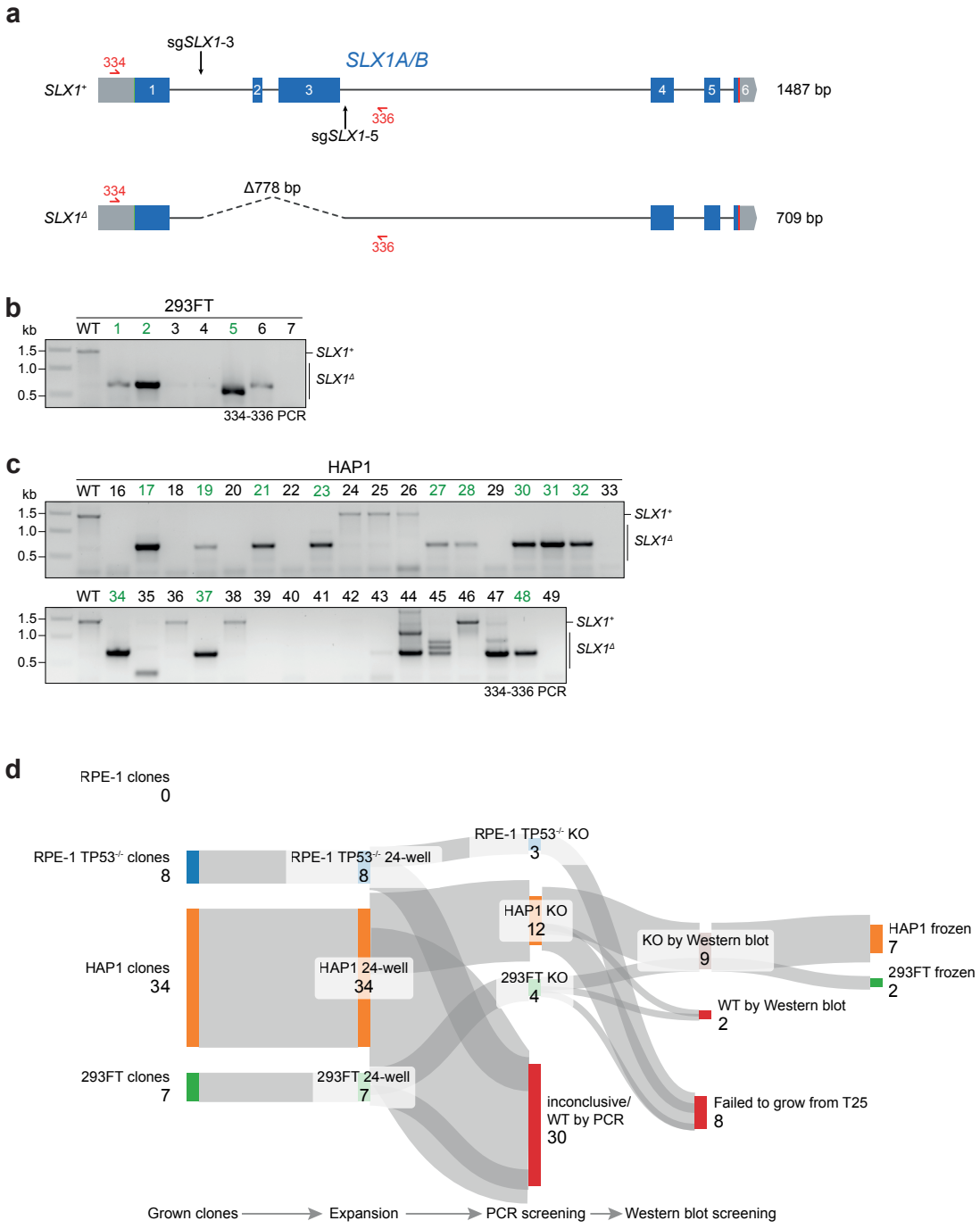


Figure 4.11 | Generation and screening of *SLX1* knockout clones across multiple cell lines. a, Schematic representation of the *SLX1A/B* loci, indicating the positions of sgRNAs (sgSLX1-3, sgSLX1-5) and the expected deletion (*SLX1*Δ). PCR amplicons used for screening (*SLX1** and *SLX1*Δ) are indicated with their expected product sizes. **b,** PCR screening of single-cell-derived 293FT clones for *SLX1*Δ. Wild-type (*WT*) and Δ*SLX1* alleles are shown for clones 1–7. **c,** PCR screening of HAP1 clones for *SLX1*Δ. Products corresponding to *WT* and Δ*SLX1* alleles are shown. Clones that progressed for further validation are highlighted in green. **d,** Sankey diagram summarising the generation, expansion, PCR screening, and Western blot validation of *SLX1* knockout clones in RPE-1, RPE-1 *TP53*$^{-/-}$, HAP1, and 293FT cells. The diagram illustrates the outcomes of each stage, including the number of clones processed, expanded, and validated as knockouts.

As discussed earlier, targeting exon 1 of the SLX1 genes does not offer a reliable knockout strategy and risks a haplosufficient SLX1 gene product. Therefore, we focused on deleting downstream exons. Exon 2 deletion alone circumvents any risk of a downstream start codon rescuing SLX1 fragment expression, but the https://crispor.gi.ucsc.edu/ tool did not predict any high-efficiency sgRNAs in its downstream intron (intron 2). Therefore, sgRNAs targeting intron 3, alongside complementary sgRNAs targeting intron 1, were sought. This combination of sgRNAs leads to the excision of both exons 2 and 3, a deletion which is not divisible by three and, when skipped, results in a premature stop codon in exon 4. The predicted product of this editing is SLX1 frameshifted from W63, with a premature stop codon introduced 14 amino acids later (p.E64MfsX14). Four sgRNAs upstream of exon 2 (two in exon 1 and two in intron 1) were selected and cloned into pX458, and two sgRNAs (in intron 3) were cloned into pX458-mRuby2 (Section 2.2.7.5) (Figure 4.10a). A SURVEYOR assay was conducted to validate the efficacy of each sgRNA, identifying sgRNA-3 (intron 1) and sgRNA-5 (intron 3) as the most efficient pair of guides (Figure 4.10b), and these were subsequently used for knockout attempts. We introduced this sgRNA combination into all cell lines and sorted double-positive GFP^+/$mRuby^+$ cells (Appendix B.2, Figure B.2). Double-positive cells could be sorted for both RPE-1and RPE-1 *TP53*$^{-/-}$ cells, although the RPE-1 cells did not yield any colonies, and the RPE-1 *TP53*$^{-/-}$ colonies rapidly deteriorated upon expansion beyond a 96-well plate and could not be further propagated. For 293FT, 7 colonies were recovered from a total of 288 seeded single cells (Figure 4.11d). In HAP1, 34 colonies were recovered from 288 single cells (Figure 4.11d). Some colonies failed to expand at various stages in the protocol of splitting from a 96-well plate to a 24-well plate to a T25 flask to a T175 flask (for freezing); the number of clones expanded is summarised in Figure 4.11d. Clones from 293FT and HAP1s were confirmed to contain *SLX1* edited alleles by PCR, some with no obvious presence of the WT allele (Figure 4.11b, c). Clones 1 and 5 in 293FTs, and clones 17, 21, 32, 37 and 48 in HAP1s were confirmed to contain a deletion of *SLX1* exons 2 and 3 by Sanger sequencing (data for HAP1 WT vs Δ*SLX1* 37 shown in Figure 4.12a, b).

As with knockdown trials of SLX1, Western blotting for SLX1 has resulted in varying degrees of success and provided inconsistent results even for repeats of identical blots. Nevertheless, proving that the obtained clones were knockouts at the protein level was

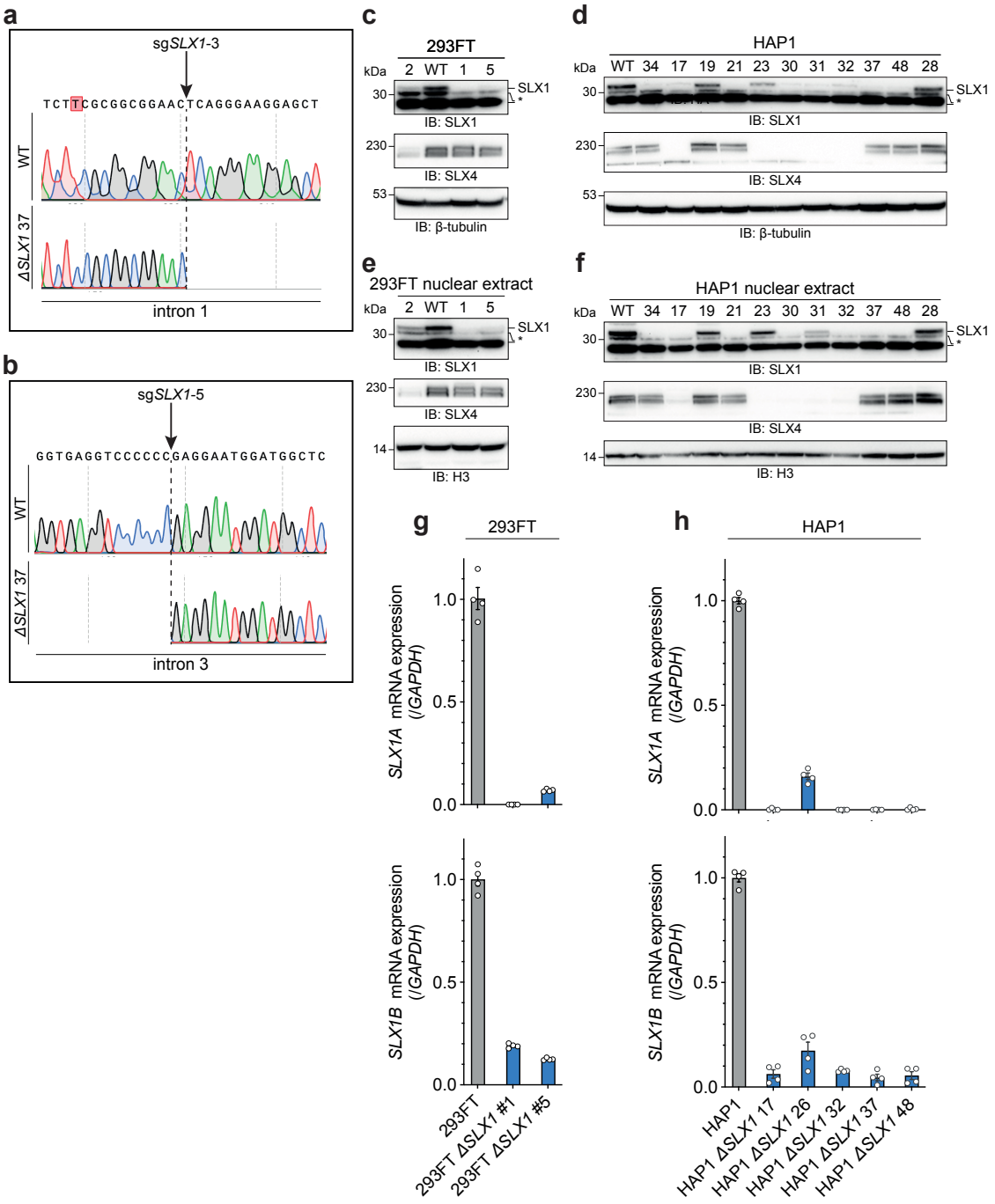


Figure 4.12 | Validation of *SLX1* knockout clones through Sanger sequencing, Western blot and RT-qPCR analysis. a, b, Sanger sequencing of HAP1 wild-type (WT) and a representative Δ*SLX1* clone (37), deletion of intervening sequence between cut sites of sg*SLX1*-3 and sg*SLX1*-5 at intron 1 (**a**) and intron 3 (**b**). **c, d,** Whole cell extract Western blot analysis of Δ*SLX1* and wild-type (WT) 293FT (**c**) and HAP1 (**d**) clones. **e, f,** Nuclear extract Western blot analysis of Δ*SLX1* and wild-type (WT) 293FT (**e**) and HAP1 (**f**) clones. **g, h,** RT-qPCR analysis of *SLX1A* and *SLX1B* mRNA expression in 293FT (**g**) and HAP1 (**h**) clones, normalised to GAPDH expression. Error bars represent mean ± s.d. of three technical replicates per condition.

attempted by Western blot. A first pass resulted in no apparent bands fully disappearing in the clones (data not shown) - this was concerning, but given no WT band was present in the PCRs, adaptations were made to Western blotting protocols. Firstly, optimisation of antibody dilutions and blocking conditions was tried using cells with knockdown of SLX1 as in Figure 4.8. Unfortunately, this could also not be reproducibly repeated. Instead, nuclear extracts of cells more definitively identified clones without SLX1 for 293FT (Figure 4.12e) and HAP1 (Figure 4.12f). In these clones, SLX4 expression was not uniformly affected consistently, with some knockout clones expressing no SLX4 (HAP1 Δ*SLX1* 17, 32), while some others expressed less SLX4 (293FT Δ*SLX1* 1, 5) or were mostly unaffected (HAP1 Δ*SLX1* 37, 48 (Figure 4.12f). Clones 1 and 5 for 293FT and clones 17, 32, 37 and 48 for HAP1 were expanded for further analysis as likely being null for SLX1. HAP1 clone 26 was also maintained as a likely mixed allele clone (Figure 4.11c). Of these clones, RT-qPCRs were performed and allowed further validation of knockout status. All Western blot-confirmed clones had near-undetectable levels of *SLX1* mRNA (except for 293FT Δ*SLX1* clone 5) by a probe set for *'SLX1A'*, but only showing greatly reduced *SLX1* mRNA using a second probe set designated for *'SLX1B'* (Figure 4.12g, h). This discrepancy is likely due to the loss of exon 3 by design, which would allow detection by the TaqMan probes against *'SLX1B'* (detects exons 4 and 5), but not by the probes against *'SLX1A'* (detects exons 3-4; 2.5.2). Some minimal detection of *SLX1* mRNA by both probe sets in HAP1 clone 26 confirmed that this clone likely includes an intact SLX1 allele (Figure 4.12g, h).

4.2.2.1 | Viability assays of SLX1 knockout clones

Viability assays were used to map the sensitivity profiles of the 293FT and HAP1 Δ*SLX1* clones compared to WT, however, instead of the resazurin assay, end-point viability was measured by ATP anabolism *via* CellTiter-Glo. Agreeing with previous work[540], the loss of SLX1 sensitises cells to crosslinking chemotherapeutics like cisplatin, SJG-136 and MMC in all HAP1 Δ*SLX1* clones, except for the haploinsufficient 26 clone (Figure 4.13i, j, l, m, n, p) and 293FT Δ*SLX1* clone 1, but not 5 (Figure 4.13a, b, d, e, f, h). The lack of MMC-sensitivity in 293FT Δ*SLX1* clone 5 might reflect some minimal SLX1 expression observed by RT-qPCR (Figure 4.12g). SLX1 loss also sensitised

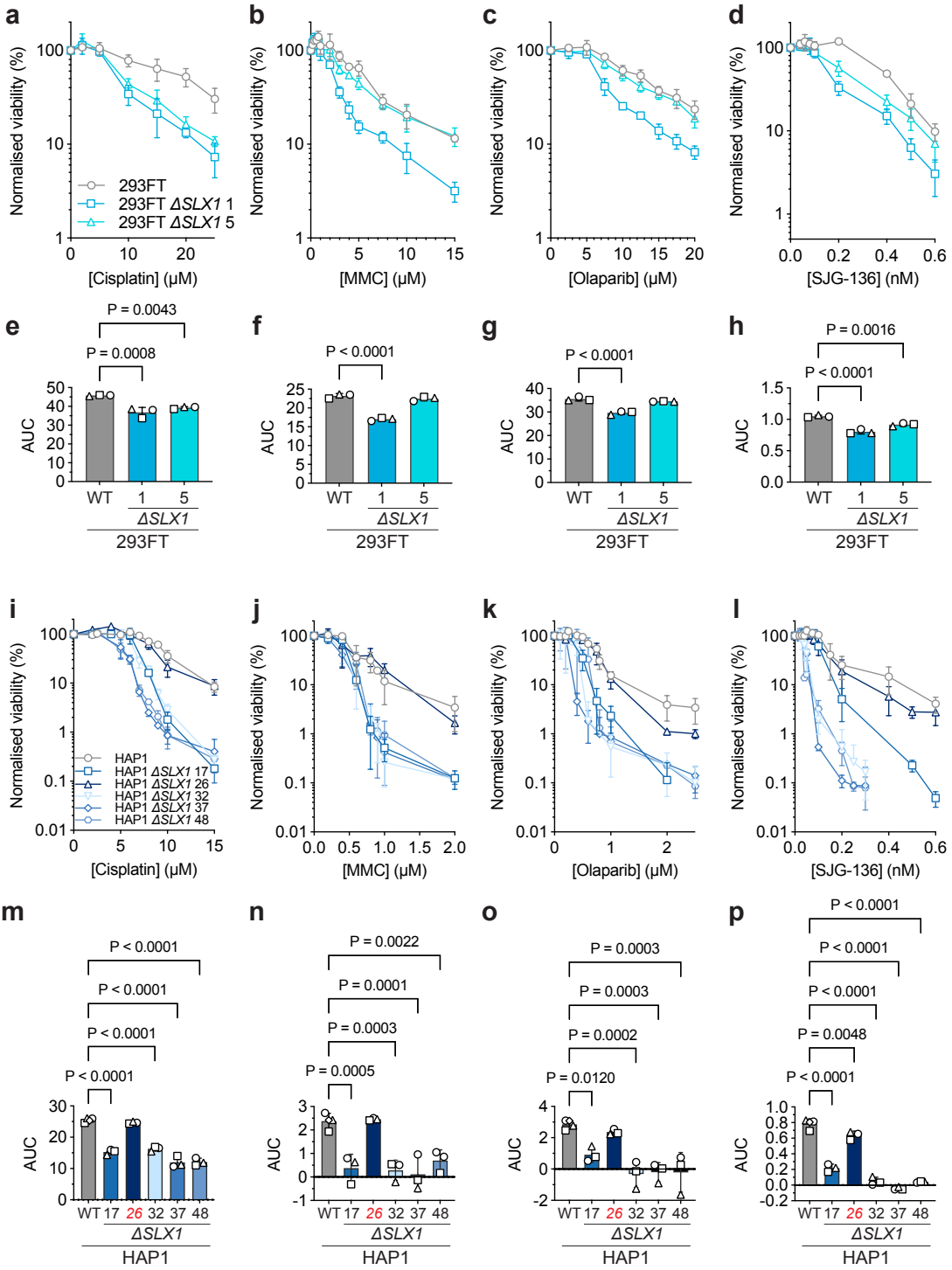

a
b
c
d
Normalised viability (%)
293FT
293FT ΔSLX1 1
293FT ΔSLX1 5
[Cisplatin] (μM)
[MMC] (μM)
[Olaparib] (μM)
[SJG-136] (nM)
e
f
g
h
AUC
P = 0.0043
P = 0.0008
P < 0.0001
P < 0.0001
P = 0.0016
P < 0.0001
WT
1
5
ΔSLX1
293FT
i
j
k
l
HAP1
HAP1 ΔSLX1 17
HAP1 ΔSLX1 26
HAP1 ΔSLX1 32
HAP1 ΔSLX1 37
HAP1 ΔSLX1 48
m
n
o
p
P < 0.0001
P < 0.0001
P < 0.0001
P < 0.0001
P = 0.0022
P = 0.0001
P = 0.0003
P = 0.0005
P = 0.0003
P = 0.0003
P = 0.0002
P = 0.0120
P < 0.0001
P < 0.0001
P < 0.0001
P = 0.0048
P < 0.0001
WT 17 26 32 37 48
ΔSLX1
HAP1

Figure 4.13 | Survival analysis of Δ*SLX1* clones following treatment with DNA-damaging agents. a-d, Quantification of cell viability assays of wild-type (WT) and Δ*SLX1* 293FT clones (clones 1 and 5) treated with cisplatin (**a**), mitomycin C (MMC, **b**), olaparib (**c**), and SJG-136 (**d**). Viability was measured using resazurin fluorescence after 8 days and is presented as mean ± s.d. relative to untreated cells for n ≥ 3 biological replicates with n ≥ 4 technical repeats. **e-h,** Area under curve (AUC) for the survival curves in **a-d**, represented as mean ± s.d. of n = 3 biological replicates. Statistical significance was determined using one-way ANOVA with Dunnett's T3 multiple comparisons test, with comparisons between WT and Δ*SLX1* clones for each condition. **i-l,** Quantification of cell viability assays of WT and Δ*SLX1* HAP1 clones (clones 17, 26, 32, 37, and 48) treated with cisplatin (**i**), MMC (**j**), olaparib (**k**), and SJG-136 (**l**). Viability was measured as in **a-d**. **m-p,** AUC for the survival curves in **i-l**, represented as mean ± s.d. of n = 3 biological replicates. Statistical significance was determined using one-way ANOVA with Dunnett's T3 multiple comparisons test, with comparisons between WT and Δ*SLX1* clones for each condition. Individual clones are indicated by their clone ID.

HAP1 Δ*SLX1* clones, but not 26, to olaparib (Figure 4.13j, n). This was recapitulated in 293FT Δ*SLX1* clone 1, but not clone 5 (Figure 4.13c, g). To verify that the sensitivities observed for the representative HAP1 Δ*SLX1* 37 and 293FT Δ*SLX1* 1 clones was due to SLX1 loss, Dox-inducible GFP-SLX1 was complemented back into these cells by generation of a stable heterogeneous population after sorting Dox-induced GFP^+ cells (Appendix B.3, Figure B.3; Appendix B.4, Figure B.4). Dox-induced expression of GFP-SLX1 in established complemented populations was confirmed by Western blot for HAP1 (Figure 4.14b), and 293FT (Figure 4.14b). SLX4 expression appeared reduced in HAP1 Δ*SLX1* 37, which was restored under Dox-treatment of iGFP-SLX1-transduced cells but not iGFP-EV-transduced cells, although overall protein loading in HAP1 Δ*SLX1* 37 transduced with iGFP-EV cells was also reduced (Figure 4.14b).

In 293FT Δ*SLX1* 1, re-expression of SLX1 partially reversed the sensitivities observed against olaparib (Figure 4.15c, g) and SJG-136 (Figure 4.15d, h). A similar pattern was observed for cisplatin and MMC complementation but did not meet statistical significance (Figure 4.15a, b, e, f). In contrast, HAP1 Δ*SLX1* 37 sensitivity to all drugs was not rescued by expression of GFP-SLX1 (Figure 4.15i-p).

4.2.3 | SLX1 degron knock-in

Given the challenges associated with obtaining RPE-1 *SLX1* knockout clones (Section 4.2.2, as well as the inconsistencies in restoring viability to genotoxic agents when attempting complementation of the HAP1 and 293FT *SLX1* knockout clones, an alternative strategy was therefore attempted to establish a CRISPR-KO screening model in RPE-1 cells. This is

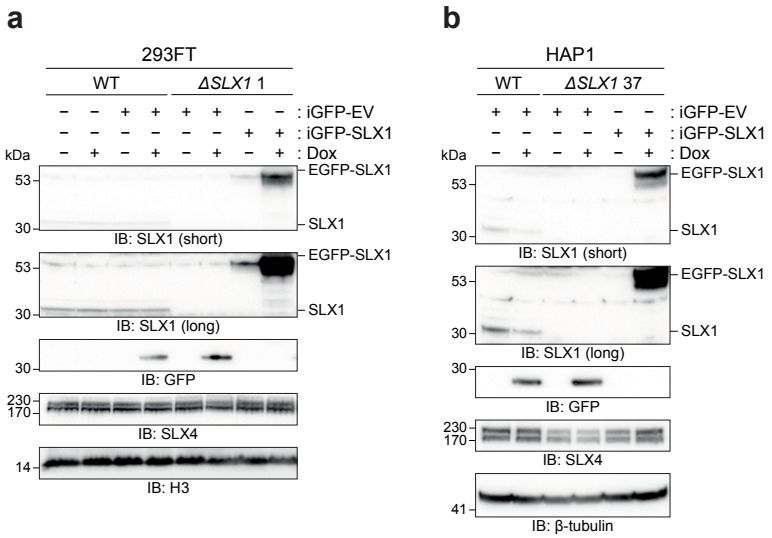


Figure 4.14 | Complementation of Δ*SLX1* clones with inducible GFP-SLX1. **a,** Nuclear extract Western blot analysis of 293FT *WT* and Δ*SLX1* clone 1 transduced with inducible GFP-SLX1 (iGFP-SLX1) or GFP-empty vector (iGFP EV). Cells were either untreated or treated with doxycycline (Dox) to induce GFP-SLX1 expression overnight. **b,** Western blot analysis of HAP1 *WT* and Δ*SLX1* clone 37 transduced and treated as in **a**.

comparable to those utilised in the DNA damage response (DDR) field[446,500], and involved employing a degron-based knock-in approach.

4.2.3.1 | Design considerations for an *SLX1* knock-in strategy

Within the 762 kb separating the two *SLX1* paralogues, approximately 31 protein-coding genes and several miRNAs are encoded. Following the duplication of the *BOLA2-SLX1-SULT1A* gene segment in humans, positive selection bias has prevented neutral drift between these homologous regions (Figure 4.9a). Although this duplication is rare, it poses significant challenges to achieving targeted knockout of *SLX1*. In pursuit of an inducible method to abrogate *SLX1*, the Cre-loxP system was initially considered for controlled excision of a critical exon, such as exons 2 and 3 as identified in 4.2.2. Deletion of these exons would introduce a frameshift, generating a premature stop codon, with no downstream start codon capable of rescuing hypomorphic *SLX1* expression. However, this approach was ultimately abandoned due to the risk of removing the 0.7 Mb genomic region intervening between the loxP sites of *SLX1A* and *SLX1B* during Cre-mediated recombination. Such an event would excise all intervening genes, thereby confounding studies specific to *SLX1*. To circumvent these challenges, attention shifted to the acute

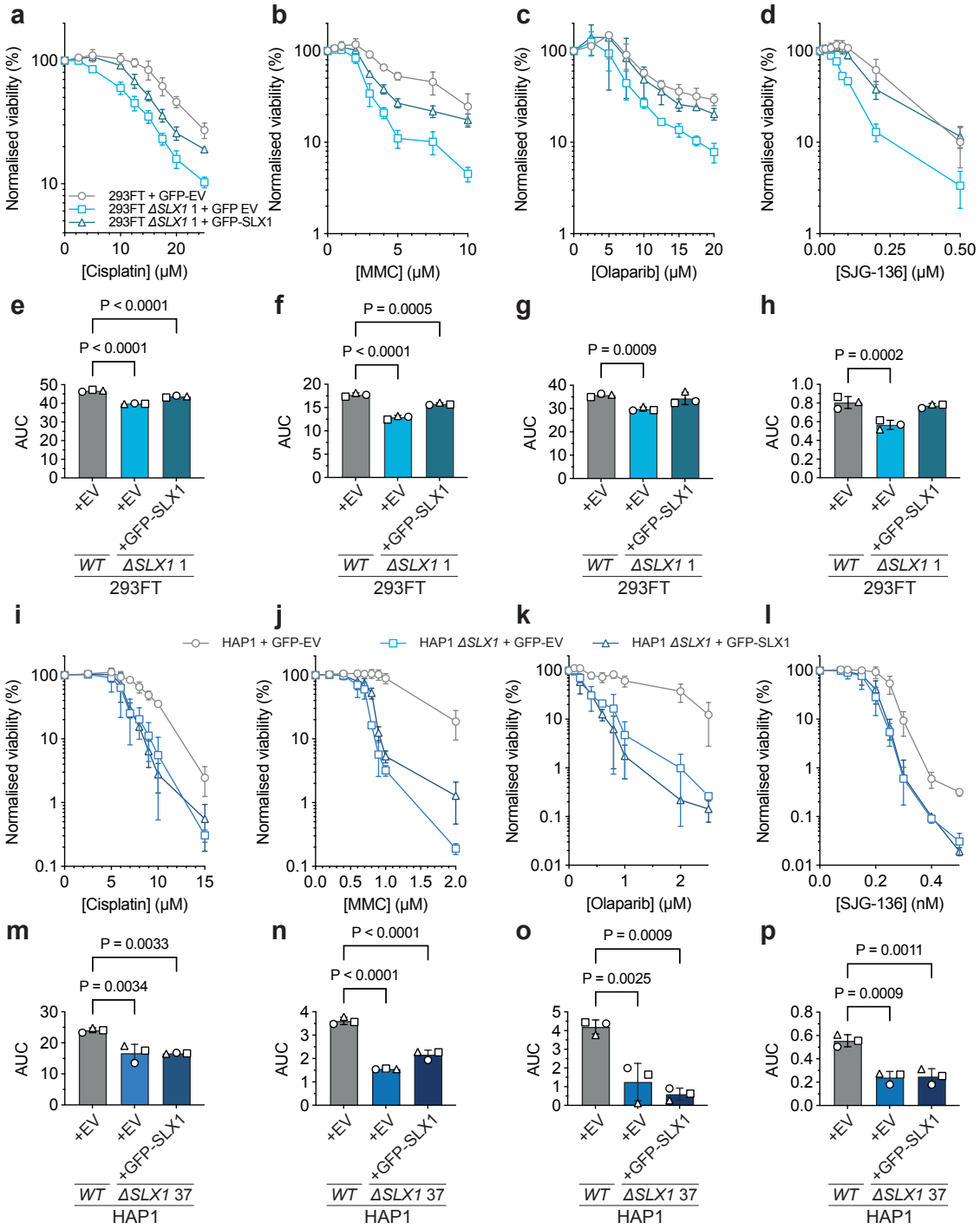

a
b
c
d
Normalised viability (%)
293FT + GFP-EV
293FT ΔSLX1 1 + GFP EV
293FT ΔSLX1 1 + GFP-SLX1
[Cisplatin] (μM)
[MMC] (μM)
[Olaparib] (μM)
[SJG-136] (μM)
e
f
g
h
AUC
P < 0.0001
P < 0.0001
P = 0.0005
P < 0.0001
P = 0.0009
P = 0.0002
+EV
+EV
+GFP-SLX1
WT
ΔSLX1 1
293FT
i
j
k
l
HAP1 + GFP-EV
HAP1 ΔSLX1 + GFP-EV
HAP1 ΔSLX1 + GFP-SLX1
[Cisplatin] (μM)
[MMC] (μM)
[Olaparib] (μM)
[SJG-136] (nM)
m
n
o
p
P = 0.0033
P = 0.0034
P < 0.0001
P < 0.0001
P = 0.0009
P = 0.0025
P = 0.0011
P = 0.0009
ΔSLX1 37
HAP1

Figure 4.15 | Survival analysis of complemented Δ*SLX1* clones following treatment with DNA-damaging agents. a-d, Quantification of cell viability assays of wild-type (WT), Δ*SLX1* 293FT clone 1, and Δ*SLX1* clone 1 complemented with GFP-SLX1 (+GFP-SLX1) treated with cisplatin (**a**), mitomycin C (MMC, **b**), olaparib (**c**), and SJG-136 (**d**). Viability was measured using resazurin fluorescence after 8 days and is presented as mean ± s.d. relative to untreated cells for n ≥ 3 biological replicates with n ≥ 4 technical repeats. **e-h,** Area under curve (AUC) for the survival curves in **a-d**, represented as mean ± s.d. of biological replicates. Statistical significance was determined using one-way ANOVA with Dunnett's T3 multiple comparisons test, with comparisons between WT, Δ*SLX1*, and Δ*SLX1*+GFP-SLX1 for each condition. **i-l,** Quantification of cell viability assays of WT, Δ*SLX1* HAP1 clone 37, and Δ*SLX1* clone 37 complemented with GFP-SLX1 (+GFP-SLX1) treated with cisplatin (**i**), MMC (**j**), olaparib (**k**), and SJG-136 (**l**). Viability was measured as in **a-d**. **m-p,** AUC for the survival curves in **i-l**, represented as mean ± s.d. of biological replicates. Statistical significance was determined using one-way ANOVA with Dunnett's T3 multiple comparisons test, with comparisons between WT, Δ*SLX1*, and Δ*SLX1*+GFP-SLX1 for each condition.

degradation of SLX1, which offers an opportunity to study its essentiality without allowing cells to adapt to a permanent genetic change, as might occur with CRISPR-Cas9 knockout approaches. A degron-based system enables rapid and controlled modulation of SLX1 protein abundance, facilitating the investigation of both acute and chronic loss of SLX1. The FKBP/dTag system was selected due to its minimal genetic footprint—requiring a single transgene knock-in—and its relatively small size (12 kDa), which minimises disruption to the already small SLX1 protein product[836]. This proteolysis targeting chimaera (PROTAC) system was designed to integrate the $FKBP12^{F36V}$ degron domain in-frame with either the N-terminus or C-terminus of SLX1 at its endogenous loci using CRISPR-Cas9-mediated knock-in. However, modifications to the sequence of either *SLX1* paralogue immediately upstream of the start codon risk disrupting the promoter regions of the adjacent *BOLA2* and *BOLA2B* genes, which are encoded on the antisense strand and have promoters located within the first few exons of their respective adjacent *SLX1* genes (Figure 4.16a). Similarly, incorporating a dTag construct at the C-terminus of SLX1 could interfere with the promoter regions of the downstream *SULT1A3/4* genes (Figure 4.16a). Tagging the N-terminus was settled on after modelling the putative fusion constructs in AlphaFold, which showed that the specific N-terminal tags were moderately more flexible/free (data not shown).

Thus, since SLX1 knock-out models could not be generated in RPE-1 cells (Figure 4.11d), a degron knock-in was instead designed to tag the N-terminus of SLX1 in this cell line. Initially, a strategy involving the co-transfection of a synthetic sgRNA/Cas9 ribonucleoprotein (RNP) complex with a plasmid donor template was employed (see

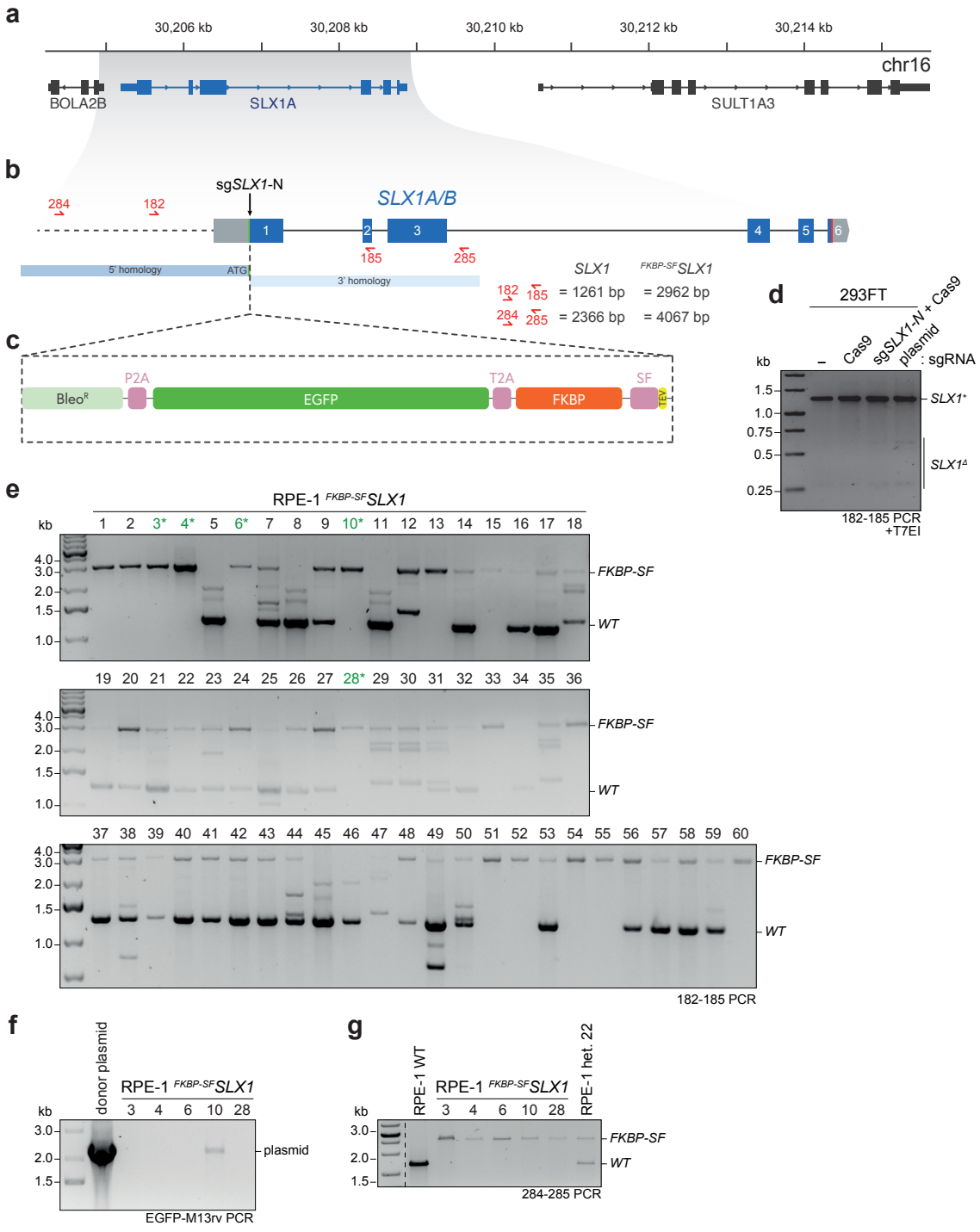

a
30,206 kb
30,208 kb
30,210 kb
30,212 kb
30,214 kb
chr16
BOLA2B
SLX1A
SULT1A3
b
sgSLX1-N
SLX1A/B
284
182
185
285
5' homology
ATG
3' homology
SLX1
FKBP-SFSLX1
182 185 = 1261 bp
= 2962 bp
284 285 = 2366 bp
= 4067 bp
c
P2A
T2A
SF
Bleo^R
EGFP
FKBP
TEV
d
293FT
– Cas9
sgSLX1-N + Cas9
plasmid
: sgRNA
kb
1.5
1.0
0.75
0.5
0.25
SLX1+
SLX1Δ
182-185 PCR
+T7EI
e
RPE-1 FKBP-SFSLX1
kb
1 2 3* 4* 5 6* 7 8 9 10* 11 12 13 14 15 16 17 18
4.0
3.0
2.0
1.5
1.0
FKBP-SF
WT
19 20 21 22 23 24 25 26 27 28* 29 30 31 32 33 34 35 36
37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
182-185 PCR
f
donor plasmid
RPE-1 FKBP-SFSLX1
3 4 6 10 28
kb
3.0
2.0
1.5
plasmid
EGFP-M13rv PCR
g
RPE-1 WT
RPE-1 FKBP-SFSLX1
RPE-1 het. 22
3 4 6 10 28
kb
3.0
2.0
1.5
FKBP-SF
WT
284-285 PCR

Figure 4.16 | Targeting strategy and screening for *FKBP-SF SLX1* in RPE-1 cells. a, Genomic location of *SLX1A* on chromosome 16, flanked by the neighbouring *BOLA2B* and *SULT1A3* genes. The region is near-perfectly identical to its homologous region upstream in chromosome 16, consisting of *SLX1B*, *BOLA2* and *SULT1A4*. **b,** Schematic representation of the *SLX1A/B* loci and the sgRNA (sg*SLX1*-N) target sites used for homology-directed repair (HDR). Expected amplicons for wild-type (*SLX1*$^{+}$) and targeted ($^{\text{FKBP}-\text{SF}}$*SLX1*) alleles are shown. **c,** Design of the donor construct for HDR, containing Bleo$^{\text{R}}$-P2A-EGFP-T2A-FKBP, and a TwinStrep-FLAG (SF) tag flanked by homology arms for integration at the *SLX1A/B* locus. **d,** PCR validation of HDR events in 293FT cells transfected with sg*SLX1*-N and the donor plasmid. Products corresponding to *SLX1*$^{+}$ and $^{\text{FKBP}-\text{SF}}$*SLX1* alleles are indicated. **e,** PCR screening of single-cell-derived RPE-1 clones for $^{\text{FKBP}-\text{SF}}$*SLX1* integration using the 182-185 amplicon. Clones progressing for further validation are highlighted in green. **f,** PCR to ensure donor plasmid integration did not occur into selected RPE-1 $^{\text{FKBP}-\text{SF}}$*SLX1* clones using primers targeting EGFP and the M13 reverse sequence. The 'donor plasmid' lane refers to PCR amplification of genomic DNA purified from RPE-1 cells transiently transfected with only the donor plasmid. **g,** PCR validation of RPE-1 $^{\text{FKBP}-\text{SF}}$*SLX1* knock-in in clones using the longer 284-285 amplicon. WT and $^{\text{FKBP}-\text{SF}}$*SLX1* alleles are shown for comparison.

Methods). This template included a bleomycin (zeocin)-resistance gene in-frame with SLX1, enabling selection through zeocin treatment after editing events occurred. The remaining insert sequence consists of a P2A-separated EGFP, then a T2A-separated FKBP12^{F36V} degron fused to an SF tag (Figure 4.16b, c). However, no surviving cells were observed following zeocin selection (data not shown), and a SURVEYOR assay indicated that transfection of the sgRNA/Cas9 RNP failed to induce cutting at the SLX1 start codon (Figure 4.16d). However, cutting at the *SLX1* gene could be observed when RPE-1 cells were transfected with a plasmid expressing the same sgRNA (Figure 4.16d). This plasmid also enabled selection *via* puromycin (a Puro$^{\text{R}}$ gene is located downstream of Cas9). This revised strategy incorporated puromycin selection for the sgRNA-Cas9 plasmid prior to zeocin treatment and sorting for EGFP^{+} cells, which indicated successful editing of at least one allele. Sorting of cells that were EGFP^{+} required identification of a population with signal only slightly above background, likely due to the minimal expression of SLX1 (Appendix B.5, Figure B.5). Over 60 clones were recovered, and PCR analysis around the start codon revealed that most clones carried the desired knock-in edit, with some displaying only the knock-in *SLX1* PCR product and no WT alleles (Figure 4.16e). Sanger sequencing confirmed that five clones (3, 4, 6, 10, 28) possessed the correct knock-in edit. Primers designed for colony screening by PCR targeted regions within the left and right homology arms of the donor plasmid (Figure 4.16b), which raises the concerns that knock-in alleles observed by PCR might instead represent random integration events. PCR using primers annealing to the plasmid itself, including M13rv

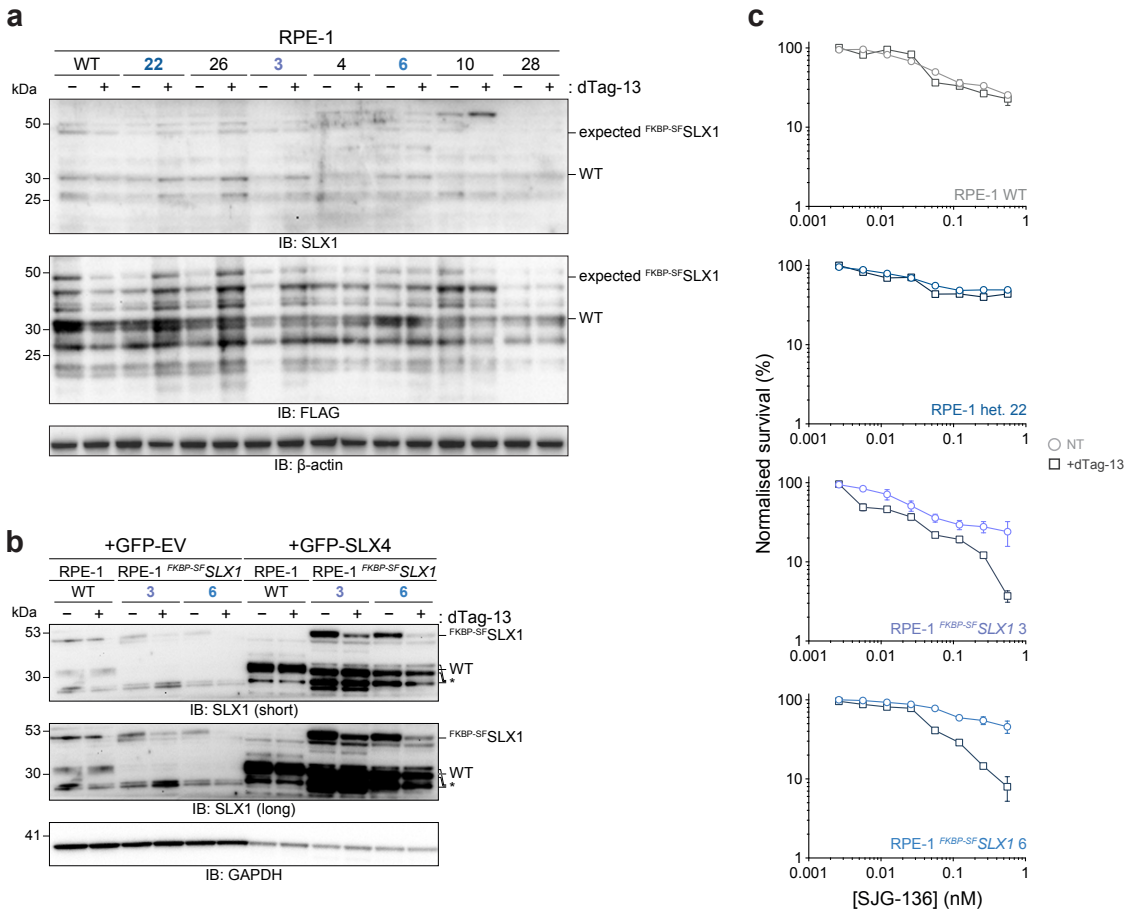


Figure 4.17 | Validation of *FKBP-SF SLX1* expression in RPE-1 cells and functional impact of dTag-13-induced degradation. a, Western blot analysis of whole-cell lysates from RPE-1 $^{\mathrm{FKBP-SF}}$*SLX1* clones treated with or without 1 1 µM dTag-13. **b,** Western blot analysis of RPE-1 $^{\mathrm{FKBP-SF}}$*SLX1* clones transiently transfected with GFP-SLX4 or GFP-EV ± 1 µM dTag-13. **c,** Quantification of cell viability assay of RPE-1 WT, heterozygous, and $^{\mathrm{FKBP-SF}}$*SLX1* clones 3 and 6 treated with increasing doses of the DNA crosslinking agent SJG-136 ± 1 μM dTag-13, representing mean ± s.d. of n = 2 biological repeats, each consisting of three technical repeats.

(which would not be part of a successful knock-in event), showed that clones 3, 4, 6, and 28 had no such random integration PCR product, while DNA from RPE-1 cells transiently transfected with the donor plasmid and clone 10 showed a strong PCR product (Figure 4.16f). Further verification with primers annealing upstream of the 5′ homology region, and within the promoter of *BOLA2B*, showed that clones 3, 4, 6, 10, and 28 all produced a single knock-in amplicon, while WT RPE-1 and mixed allele (het.) clone 22 displayed the WT amplicon (Figure 4.16g).

The poor specificity of commercial antibodies for SLX1 significantly hindered the detection of knock-in events at the protein level (Figure 4.17a). Western blotting attempts to detect FLAG, Strep, or FKBP-tagged SLX1 in clones were largely unsuccessful

(Figure 4.17a). FLAG expression in presumed knock-in clones was undetectable, necessitating further validation to confirm the functionality of dTag-inducible SLX1 degradation. This detection challenge was attributed to the relatively low expression levels of SLX1, particularly in RPE-1 cells (Figure 4.9b). Insights from prior studies (detailed in later chapters 5, 5.4.2; Figure 5.20d) suggested that overexpression of GFP-SLX4 could enhance SLX1 expression, thereby improving the detection of SLX1 wild-type (WT) and knock-in bands. SLX4 WT has been shown to increase SLX1 expression, while a mutant version incapable of interacting with SLX1, MUS81-EME1, and XPF-ERCC1 (SLX4$^{529FLW>A,1577EL>AA,C1805R}$), does not[661]. This overexpression effect was observed only with GFP-SLX4 and not GFP alone (Figure 4.17b).

In clones 3 and 6, GFP-SLX4 overexpression revealed an $^{FKBP-SF}$SLX1 band in the absence of the WT SLX1 band (Figure 4.17b). Longer blot exposures under GFP empty vector (GFP-EV) overexpression allowed better resolution of the WT and $^{FKBP-SF}$SLX1 bands (Figure 4.16b). Treatment with dTag-13 overnight appeared to reduce, though not completely eliminate, $^{FKBP-SF}$SLX1 expression, particularly in clone 3, while WT SLX1 remained unaffected (Figure 4.17b). To validate these clones further, cell viability was assessed against increasing concentrations of SJG-136, which can distinguish SLX1 loss in HAP1 and 293FT cells (Figure 4.17c). With dTag-13 treatment, clone 6, and to a lesser extent clone 3, displayed increased sensitivity to SJG-136 (Figure 4.17c), correlating with the degree of $^{FKBP-SF}$SLX1 degradation (Figure 4.17b). Although these experiments provided preliminary validation, they did not conclusively demonstrate the complete degradation of SLX1 by dTag-13. Consequently, experiments using these cell lines were placed on hold due to the inability to confirm definitive SLX1 degradation.

4.3 Discussion

This study aimed to elucidate the roles of XPF and SLX1 in maintaining genomic stability, with a focus on their individual contributions to DNA repair pathways. Through the generation of knockout and knock-in cell lines, the findings highlight the complexities of targeting these nucleases for therapeutic purposes and reveal hints into their biological functions.

XPF knockouts: implications for DNA repair and synthetic lethality XPF-ERCC1 is pivotal for the repair of a diverse array of DNA lesions, spanning NER, ICL repair, and single-strand annealing (SSA). The extreme sensitivity of XPF-null cells to crosslinking agents like SJG-136 underscores its critical roles in the FA pathway, while also agreeing with previous work using MMC or cisplatin[826]. Interestingly, the sensitivity profile of XPF-null cells varied across cell lines, with U2OS cells displaying greater sensitivity to ICL-inducing agents than HT-29 cells, likely reflecting intrinsic differences in p53 status and DNA damage response (DDR) mechanisms (Figure 4.2, Figure 4.3). Formaldehyde sensitivity was also marked in these cells, in agreement with previous work of HAP1 XPF^{-} cells[837]. In this work, the loss of NER and FA (XPA^{-}) explains the cytotoxicity of XPF^{-} cells towards formaldehyde. However, only in HT-29 cells, depletion of MUS81 and BRCA1 compounded formaldehyde sensitivity in Δ*ERCC4* cells (Figure 4.3b, e). The precise nature of this formaldehyde defect in HT-29 cells specifically requires further work, though it may reflect the relative stemness of HT-29 cells observed in Chapter 3. The comparison of viability between A549 wild-type (WT) and Δ*ERCC4* cells revealed much greater variability in all measurements of viability, suggesting increased heterogeneity in metabolic output. This observation warrants further investigation to determine the underlying causes of such variability and its implications for cellular function. The work here also corroborated the resistance observed for XPF-null cells towards the TC-NER poison, trabectedin (Figure 4.5). Loss of XPF resulted in resistance to trabectedin, while the depletion of BRCA1 (in U2OS) and either BRCA1 or MUS81 (in A549) exacerbated the toxicity of this agent only in the context of Δ*ERCC4*. One explanation for this is that the abortive NER DNA break generated by XPF at trabectedin lesions is eventually repaired mostly by a non-HR pathway, but that an unengaged (at least by NER) trabectedin lesion requires HR when cells eventually reach S phase. An extension of this work, including an NHEJ inhibitor (DNA-PKcs inhibitor) or depletion of NHEJ factors, might shed light on whether this is the case.

SLX1 Knockouts: Independent Roles and Genetic Dependencies SLX1, traditionally characterised as a structure-selective endonuclease, was investigated in isolation from its interactions with other SSEs, such as MUS81 and XPF. The study demonstrated that SLX1-null cells exhibit partial sensitivity to ICL-inducing agents, consistent with their established

role in resolving recombination intermediates. Additionally, knockdown and knockout experiments revealed synthetic sickness in p53-proficient cells (Figure 4.8), implicating SLX1 in some essential homeostatic processes. This observation aligns with earlier studies suggesting a role for SLX1 in maintaining genome stability under replicative stress, albeit in fission yeast[511]. Efforts to generate SLX1-null clones highlighted the challenges posed by its genomic duplication, with *SLX1A* and *SLX1B* exhibiting near-identical sequences. The use of CRISPR-Cas9 to target these paralogues proved effective in generating HAP1 and 293FT clones, though incomplete editing or retention of functional alleles was observed in some instances (Figure 4.11, Figure 4.12). Complementation with inducible GFP-SLX1 partially restored sensitivity to DNA-damaging agents only in 293FT Δ*SLX1* cells and not their counterpart in HAP1 (Figure 4.14). While it is challenging to reconcile the contrasting results in these two cell lines, it perhaps suggests that achieving the correct dosage of SLX1 is key, with the HAP1 cells more sensitive to the higher expression levels of SLX1 obtained on complementation. Further validation of SLX1 knockout status could be achieved by functional assays. HJ resolution defects after IR are phenotypically described by increased SCEs and synthetic sickness with MUS81/GEN1([511,833].

In light of retrieving no knockout clones of SLX1 in RPE-1 cells, either WT or p53-deficient, endogenous tagging of SLX1 with a degron was designed and initiated. After some careful development, the knock-in strategy resulted in the construction of RPE-1 cells with an N-terminal FKBP-SF fusion, allowing protein degradation with dTag-13[836]. While some minimal SLX1 degradation was eventually visible by Western blotting (Figure 4.17), this partial loss does not capture the knockout status sought for further study. The primary goal of creating these lines is to enable their use in CRISPR-KO screens that identify the precise functional role of SLX1 through its genetic interactions. Given the relatively limited knowledge of SLX1 cellular biology, especially in human cells and especially compared to its associated MUS81 and XPF nucleases, a screen like that proposed here is likely the most helpful way to glean insights as to homeostatic SLX1 functions. Performing a similar screen was also the fundamental reason for generating XPF-null cells, though in this case, the experiments are nearer to completion (Sections 2.3.2.4, 2.5.4).

Challenges in Functional Validation and Future Directions The poor detectability of SLX1 protein due to its low expression levels posed a significant challenge for validation. While Western blotting and RT-qPCR provided partial confirmation of knockouts, future efforts could benefit from the use of more specific tags detectable even with minimal expression. A second knock-in construct (SLX1 knock-in v2) was designed to address the limitations of the single SF tag, which proved inadequate for detection by Western blotting. This version incorporated a $Bleo^R$-EGFP fusion upstream of a P2A sequence to facilitate both antibiotic selection and cell sorting for knock-in events. Additionally, it included a double HA tag fused to FKBP-V for inducible degradation. However, attempts to use this construct in RPE-1, RPE-1 *TP53*$^{-/-}$, and HAP1 cells were unsuccessful, as zeocin selection caused unexpected cell death, and surviving cells were not GFP^+ (data not shown). A third construct was subsequently designed, incorporating a multi-functional Halo-tag. The Halo-tag enables fluorescent cell sorting through conjugation with a Janelia fluor chloroalkane and facilitates SLX1 degradation *via* the HaloPROTAC system. HaloPROTAC3 covalently links the Halo-tagged protein to the von Hippel-Lindau (VHL) E3 ubiquitin ligase complex, promoting proteasomal degradation[838]. This approach, representing ongoing work beyond the scope of this project, offers a promising alternative for future studies aiming to achieve precise and controlled degradation of SLX1. Additionally, the variability in the sensitivity of SLX1-null clones to DNA-damaging agents suggests a need for more robust experimental systems, such as degron-based approaches for conditional protein depletion. The development of a Halo-tag-based SLX1 degron construct represents a promising avenue for achieving precise and controlled SLX1 degradation, facilitating the study of both acute and chronic loss of function.

Concluding Remarks This study provides insight into the individual roles of XPF and SLX1 in DNA repair and highlights the complexities of targeting these nucleases for therapeutic purposes. The generation of knockout and knock-in cell lines, combined with functional and genetic interaction studies, lays the groundwork for future investigations into the biological functions and therapeutic potential of these nucleases. Further studies are needed to refine strategies for selective targeting, such as achieving separation-of-function inhibition, and to explore the broader implications of nuclease loss in genome stability and cancer therapy.

5 | Purifications of SMX

Contents

Achieving high-quality purifications of the SMX complex has been a long-standing aim of the McHugh Lab; as such, a short summary of purification trials performed by colleagues before and during my DPhil and their shortcomings are presented here. Firstly, single-gene expression constructs of individual human SMX components (full-length) were co-infected as baculoviruses into insect cells, but the preparations therein suffered from limited protein yield and reproducibility. Secondly, a MultiBac approach[839] was adopted for subcomplexes SLX1-SLX4, MUS81-EME1, and XPF-ERCC1 as three separate baculoviruses. In this case, all components were stably expressed, but mixing components did not yield tractable combined complexes, neither suitable for structural biology nor proven to form a single

complex by size exclusion chromatography. When this was done, some subcomplexes showed non-aggregate co-elution on size exclusion chromatography (SEC), suggesting stronger interactions: SLX4C (comprising SLX4-SLX1-MUS81-EME1) was obtained in overlapping fractions of SEC and electron microscopy showed unique and homogeneous particles. Third, using the MultiBac system, all six proteins were contained in a single plasmid and, hence, a single baculovirus. With this system, expression of all components was uneven, and affinity purification resulted in elutions dominated by the tagged SLX4.

Thus, the main aim of the work presented here was to establish an expression and purification system for full-length SMX that ensures equal stoichiometry of all components. The polyprotein approach, adopted by many viruses, was chosen here to promote equal expression and immediate co-translational folding of the complex's constituent proteins before release of the subunits through cleavage by an upstream-encoded TEV protease[840]. Initially, a human-frog (Hs-Xl, denoted with h and x before the gene name, hereafter) chimaeric (χ) version of the complex was expressed and purified. The mouse homologues (denoted with 'm' before the gene) were then also expressed due to the observed success of mouse N-terminal SLX4-XPF-ERCC1 purifications by colleagues and the shorter length of the central disordered region compared to other vertebrate homologues (Figure 1.7, which was truncated yet further to aid future studies. Lastly, attention was turned to the purification of human SMX, in this case expressed in mammalian 293F-derivative cells. The constructs used throughout this work and how they were cloned are summarised in Methods (Section 2.2.6).

5.1 | Purification of a chimaeric SMX complex

The first strategy attempted to purify the SMX complex was that of a human-frog chimaeric complex. The rationale behind the use of *X. laevis* homologues of these proteins is due to their use in pioneering functional studies of interstrand crosslink (ICL) repair performed by the Knipscheer group. Their work has found that the behaviour of these proteins is radically different from that of their human homologues: they are far less prone to aggregation and are mostly produced in a soluble single-state form, making them well-suited for structural applications. The *H. sapiens* or *X. laevis* only versions of XPF-ERCC1

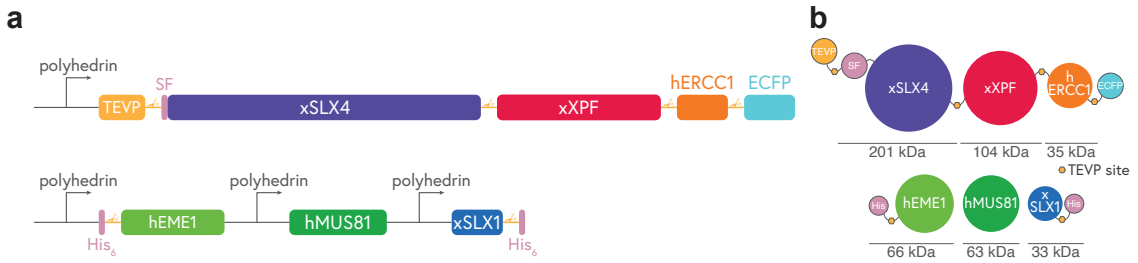


Figure 5.1 | Schematic representation of χSMX constructs and subunits. a, Diagram of the expression constructs used for the co-expression of χSMX complex subunits. The constructs include xSLX4 fused to an N-terminal TEV protease cleavage site and a TwinStrep-FLAG tag (SF), xXPF linked to hERCC1 and an ECFP tag, and individual constructs for hEME1, hMUS81, and xSLX1. Each subunit is under the control of the polyhedrin promoter, with His-tags included for purification where indicated. **b,** Theoretical molecular weights and domain organisation of each SMX subunit. The subunits are colour-coded to match their respective constructs in **a**, with TEV protease cleavage sites and His-tags highlighted. The full SMX complex includes components ranging from 33 kDa (xSLX1) to 201 kDa (xSLX4).

could not be purified well due to expression and solubility issues, but the chimaeric xXPF-hERCC1 formed a stable complex and could be readily purified[435]. Thus, the chimaeric full-length complex was also settled on based on the insolubility and poor expression of *H. sapiens*-only or *X. laevis*-only complexes.

A workflow was designed to purify the SMX complex and assess its composition, solubility, enzymatic activity, and structural integrity to facilitate structural and functional studies. Constructs were engineered to enable co-expression of the individual subunits and to be compatible with affinity purification. Variations in buffer composition were tested across three separate trials to optimise purification. The quality of preparations was evaluated using Coomassie-stained SDS-PAGE, mass spectrometry, solubility and size exclusion chromatography, nuclease assays, and negative-stain electron microscopy where possible. The aim was to balance protein yield, solubility, and enzymatic functionality while minimising contaminants and aggregation.

5.1.1 | Purification 1 of χSMX

Human and *Xenopus* SLX4 share 31% sequence identity across the full-length protein and are predicted to retain all domains required for interaction with XPF-ERCC1 (MLR), MUS81-EME1 (SAP), and SLX1 (CCD) (Appendix B.6, Figure B.6). *Xenopus* SLX4 also contains conserved BTB and UBZ4 motifs but lacks identifiable SIMs or a TRF2-binding

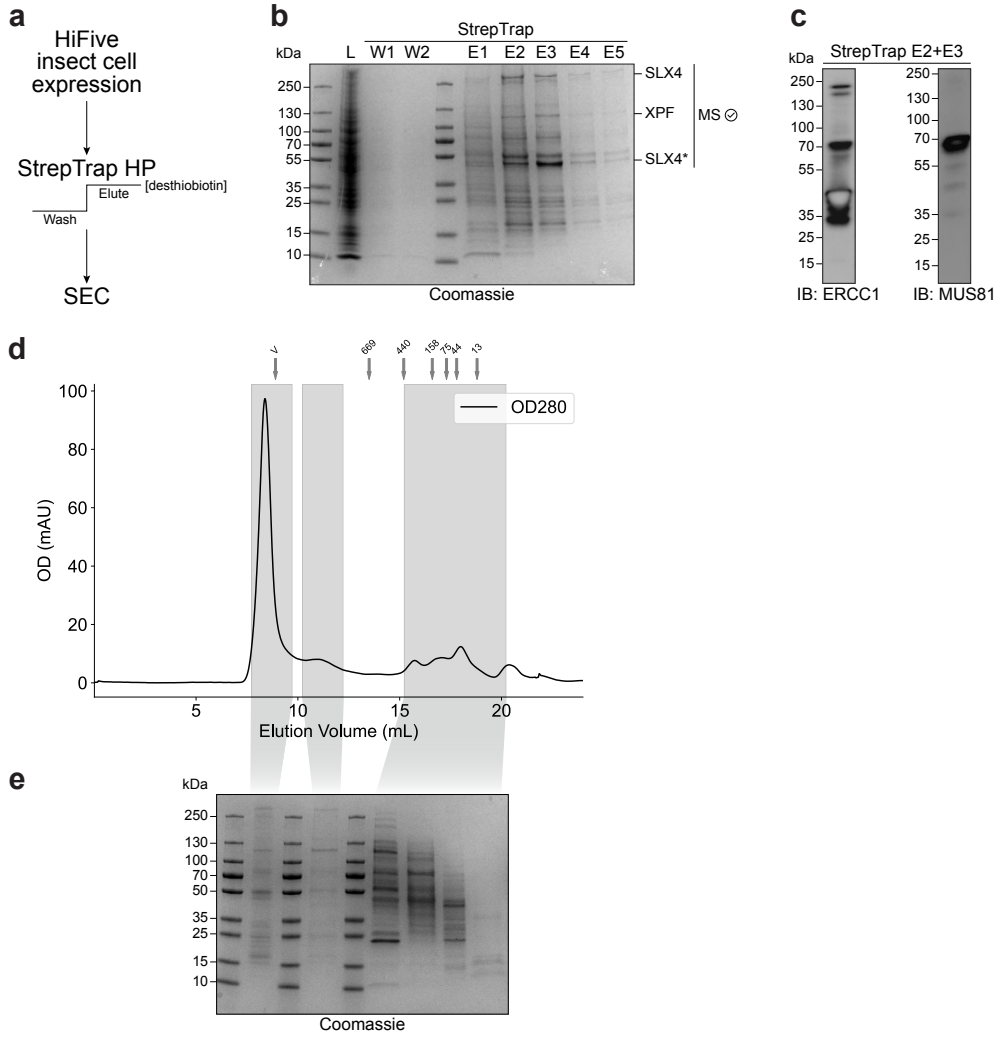


Figure 5.2 | First expression and purification of the χSMX complex from insect cells. a, Workflow for the expression and purification of the χSMX complex. The complex was expressed in HiFive insect cells and purified using StrepTrap HP affinity chromatography followed by size-exclusion chromatography (SEC). **b,** SDS-PAGE analysis of StrepTrap purification fractions. Lane L: cell lysate; W1 and W2: wash fractions; E1-E5: elution fractions. SLX4, XPF, and SLX4 degradation products (*SLX4) are indicated were verified by in-gel tryptic digest mass spectrometry (MS). **c,** Western blotting of pooled elution fractions (E2 + E3) to probe for the presence of MUS81 and ERCC1, confirming the presence of these subunits in the eluted complex. **d,** Size exclusion chromatogram (Superose 6 Increase 10/300 GL) showing OD_{280} absorbance. Arrows indicate the peaks of protein standards used to define molecular weight ranges, with fractions collected for analysis by SDS-PAGE in *e.* **e,** SDS-PAGE analysis of SEC fractions as indicated by grey shading in the chromatogram.

motif (TBM) (Figure 1.7); however, these features are not expected to compromise its ability to associate with a semi-human SMX complex.

The constructs for expression of χSMX were split such that one baculovirus would express SLX4-XPF-ERCC1 as a single ORF, fragmented into the constituent chains through cleavage of internal sites by TEV protease, expressed upstream within the same ORF (Figure 5.1a, upper panel). Human MUS81 and EME1, and *Xenopus* SLX1 were

expressed from a second baculovirus under the MultiBac approach. After confirmation of individual baculovirus transduction of insect cells during the viral amplification stage (Section 2.3.3.1), expressions were carried out in HiFive insect cells by co-transduction of both viruses. The first preparation involved recirculating clarified lysate over a StrepTrap HP column and eluting with desthiobiotin (Figure 5.2a). Elutions from StrepTrap contained mass-spectrometry-verified SLX4 and XPF (Figure 5.2b), while the presence of ERCC1 and MUS81 could be confirmed by Western blotting (Figure 5.2c), however, the presence of components SLX1 and EME1 could not be confirmed. Nevertheless, the SEC profile showed that the majority of this elution was aggregated (Figure 5.2d), and no protein signal was observed in the range expected for the size of the SMX complex ($\sim$0.5 MDa for a monomer). SDS-PAGE did not appear to show that any of the smaller-sized peak fractions contained SLX4 or other components (Figure 5.2e).

5.1.2 | Purification 2 of χSMX

To improve the aggregation issues from the first purification, $MgCl_2$, KCl and CHAPS were included in the base buffer and used from lysis onwards. The protocol was followed in an otherwise similar manner (Figure 5.3a). Elutions from StrepTrap were similar to the first purification, with only SLX4 and XPF distinguishable by in-gel tryptic digest mass spectrometry, but with the ERCC1 band visible (Figure 5.3b). SEC profile of the elution showed a faint peak close to the expected complex size, though most of the complex was still aggregated (Figure 5.3c). A faint band could be observed for SLX4 only, despite concentrating these fractions ten-fold (Figure 5.3d). Screening of this preparation on negative stain (uranyl acetate) cryoelectron microscopy showed very heterogeneous particles of varying sizes (Figure 5.3e).

5.1.3 | Purification 3 of χSMX

The third purification was also similar to purification 2 but with the change involving performing 'scouting' runs for SEC, which avoided the requirement to concentrate the StrepTrap elutions (Figure 5.4a). Once again, StrepTrap elutions contained SLX4, XPF and ERCC1 and mass spectrometry confirmed the presence of MUS81, EME1 and SLX1. Tubulin, as per previous preparations of SMX, appeared as a contaminating band[324]

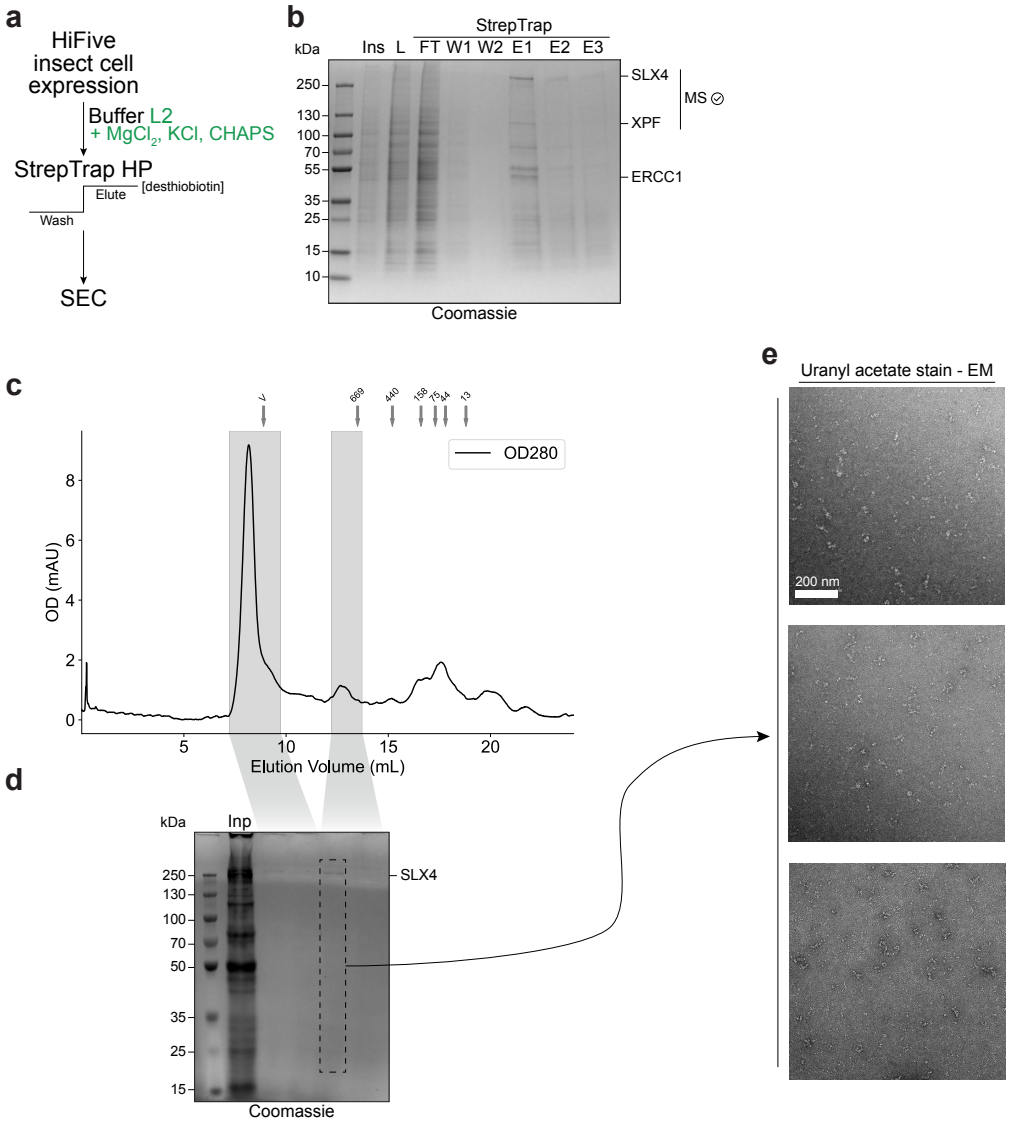


Figure 5.3 | Second expression and purification of the χSMX complex from insect cells. a, Workflow for the expression and purification of the χSMX complex. The complex was expressed in HiFive insect cells and purified using StrepTrap HP affinity chromatography followed by size-exclusion chromatography (SEC). Purification included Buffer L2 supplemented with $MgCl_2$, KCl, and CHAPS. **b,** SDS-PAGE analysis of StrepTrap purification fractions. Lane Ins: insoluble fraction of insect cell lysis; L: soluble lysate; FT: flow-through after loading on StrepTrap; W1 and W2: wash fractions; E1-E3: elution fractions. SLX4 and XPF bands were verified by mass spectrometry (MS). **c,** Size-exclusion chromatogram (Superose 6 Increase 10/300 GL) showing OD_{280} absorbance. Arrows indicate molecular weight standards, and grey-shaded regions denote collected fractions. **d,** SDS-PAGE analysis of SEC fractions from the grey-shaded regions in **c**, showing the weak presence of SLX4 after elution. **e,** Electron microscopy (EM) visualisation of SLX4-XPF complexes stained with uranyl acetate. Scale bar: 200 nm.

(Figure 5.4b). Under SEC, each run consisted of a smaller aggregation peak compared to the expected complex size peak (Figure 5.4c), and concentrating this fraction showed that all components were present, though faintly (Figure 5.4d, e).

From this purification, ∼13.5 µg of χSMX complex was recovered. Enzymatic assays were employed for each purification, but only this third preparation displayed any notable activity. In corroboration with previous findings[541], XPF-ERCC1 activity on a simple fork and 5′-flap increases greatly when co-purified with an N-terminal fragment of SLX4 (encompassing 173-947 of SLX4) - complex referred to as "mini-SLX4" (Figure 5.5a). Purified χSMX was more active than XPF-ERCC1 alone but less active than mini-SLX4 (Figure 5.5a, b). However, possible combined cleavage activities of both MUS81 and XPF prevent definitive conclusions about whether the full-length SMX complex specifically enhances XPF activity alone. On a 3′ flap substrate, the co-purification of SLX4^{C}-SLX1 with MUS81-EME1, termed SLX4-C, increases cleavage activity compared to MUS81-EME1 alone (Figure 5.4c). Again, χSMX was more active than MUS81-EME1 alone but less active than SLX4-C (Figure 5.4c), only showing cleavage at 40 nM. This activity profile was recapitulated for an intact Holliday junction (HJ) substrate, where χSMX was again only active at 40 nM (Figure 5.4d).

Micrographs of negative-stained χSMX showed improved homogeneity of particles compared to the second preparation (Figure 5.6a); however, the modal particle size was estimated to be 3 MDa, though a minor projection of ∼0.6 MDa was closer to the estimated 0.5 MDa mass for χSMX (Figure 5.6b). Two-dimensional class averages indicated an extended SMX conformation (Figure 5.6c), possibly reflecting the absence of a DNA substrate that could compact the complex. A 3D model derived from these data revealed an elongated, cylindrical structure with curved edges (Figure 5.6d). The estimated volume (∼1.5 MDa) of this model suggested the presence of a trimeric SMX complex. It is, however, unlikely that this processing represents a uniquely identified oligomerisation state of SMX. In this case, the molecule, though somewhat as expected, appeared very flexible in most class averages. Classes are also very blurry and most likely average out different conformations or oligomerisation states. Refinement of this structure with further negative stain EM data, and improvement of resolution with cryo-EM data would benefit deciphering the real organisation of SSEs in this volume. For now, the volume serves as

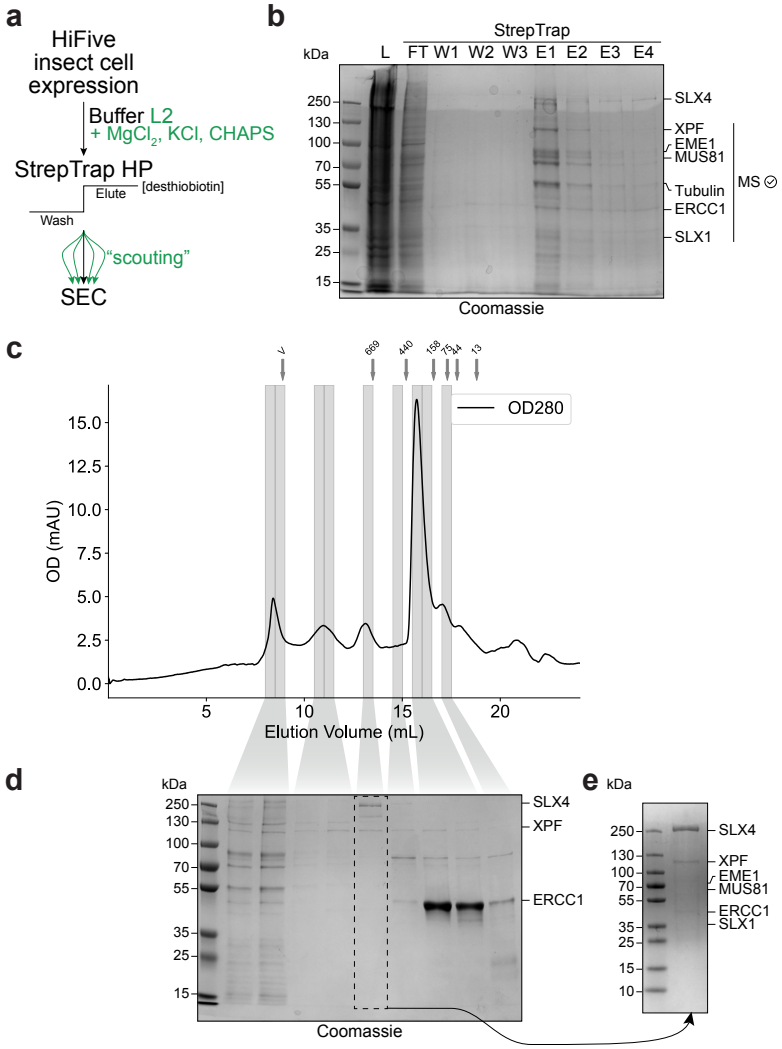


Figure 5.4 | Third expression and purification of the χSMX complex from insect cells. a, Workflow for the expression and purification of the χSMX complex. The complex was expressed in HiFive insect cells, purified using StrepTrap HP affinity chromatography, and further processed with a "scouting" size-exclusion chromatography (SEC) step. "Scouting" refers to the sequential injection of a large-volume sample in multiple individual size exclusion chromatography (SEC) runs. Elution fractions from each run are pooled, allowing collection of the entire sample, and therefore avoiding extensive concentration of the input sample beforehand. Purification used Buffer L2 supplemented with $MgCl_2$, KCl, and CHAPS as per purification 2. **b,** SDS-PAGE analysis of StrepTrap purification fractions. Lane L: lysate loaded onto the StrepTrap column; FT: flow-through after loading; W1–W3: wash fractions; E1–E4: elution fractions. SLX4, and additional associated subunits (XPF, EME1, MUS81, ERCC1, SLX1) were identified by mass spectrometry (MS). The contaminant Tubulin was also identified by MS. **c,** Size-exclusion chromatogram (Superose 6 Increase 10/300 GL) showing OD_{280} absorbance with peaks corresponding to molecular weight ranges as before. Grey-shaded regions indicate fractions collected for analysis in **d**. **d,** SDS-PAGE analysis of SEC fractions from the indicated grey-shaded regions in **c**, highlighting the elution profile of the χSMX complex. **e,** Detailed SDS-PAGE of a pooled fraction from SEC, concentrated 20 fold, demonstrating the co-purification of SLX4, XPF, MUS81, EME1, ERCC1, and SLX1.

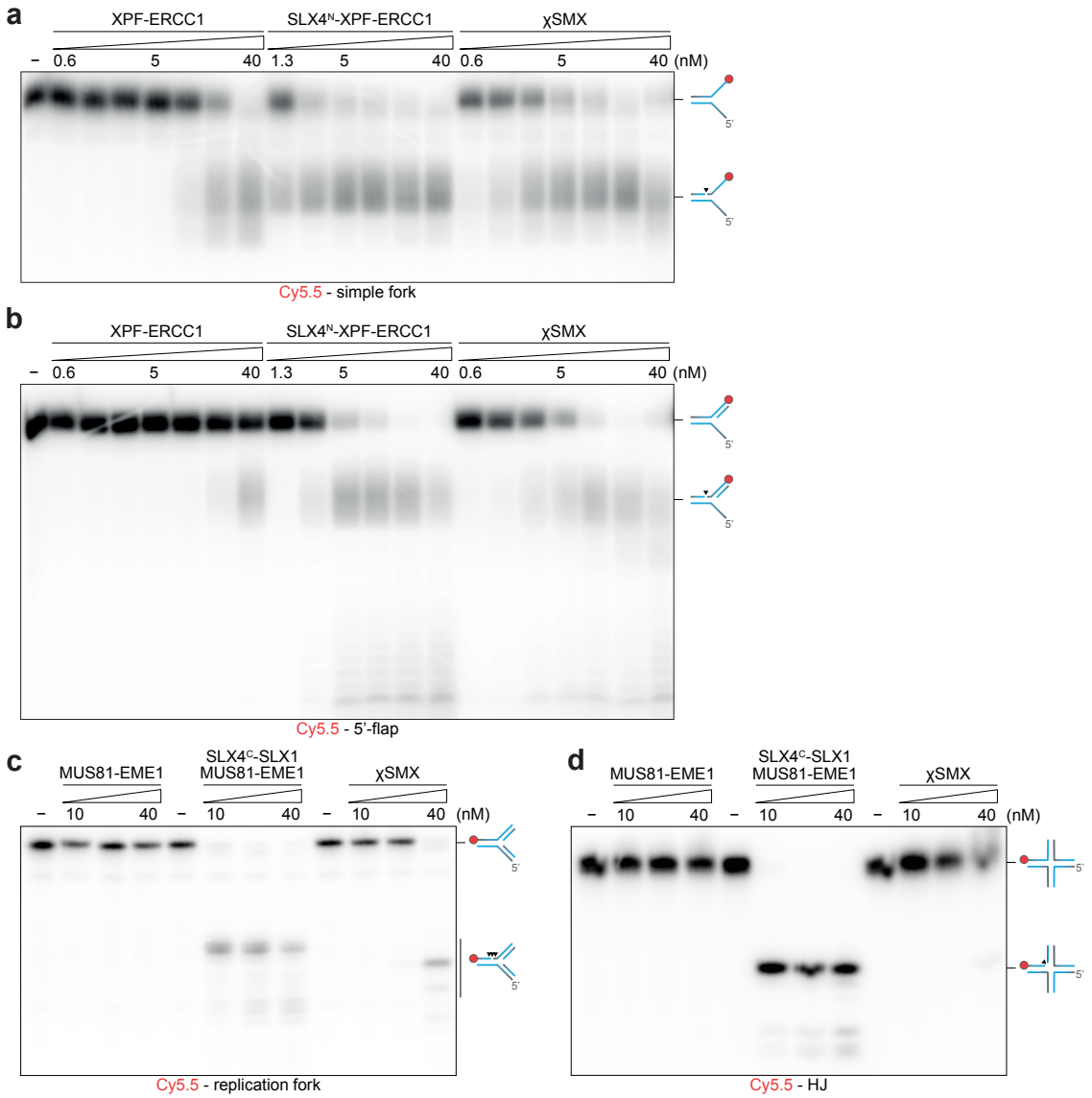


Figure 5.5 | Biochemical activity assays of the χSMX complex and other sub-complexes. **a,** Endonuclease cleavage of a Cy5.5-labelled simple fork was verified *via* a gel-based assay. Substrate processing was assessed with the incubation of two-fold increasing concentrations (0.6125 to 40 nM) of human XPF-ERCC1, human "mini-SLX4" or SLX4^{N}-XPF-ERCC1 (SLX4$^{173-947}$), and χSMX, with 10 nM DNA substrate for 1 h at 37 °C. **b,** Endonuclease cleavage of a Cy5.5-labelled 5′-flap was verified *via* a gel-based assay. Substrate processing was assessed with the incubation of two-fold increasing concentrations (0.6125 to 40 nM) of human XPF-ERCC1, human "mini-SLX4" or SLX4^{N}-XPF-ERCC1 (SLX4$^{173-947}$), and χSMX, with 10 nM DNA substrate for 1 h at 37 °C. **c,** Endonuclease cleavage of a Cy5.5-labelled replication fork was verified *via* a gel-based assay. Substrate processing was assessed with the incubation of two-fold increasing concentrations (10 to 40 nM) of human MUS81-EME1, human "SLX4-C" or SLX4^{C}-SLX1-MUS81-EME1 (SLX4$^{1535-1834}$), and χSMX, with 10 nM DNA substrate for 1 h at 37 °C. **d,** Endonuclease cleavage of a Cy5.5-labelled intact Holliday junction (HJ) was verified *via* a gel-based assay. Substrate processing was assessed with the incubation of two-fold increasing concentrations (10 to 40 nM) of human MUS81-EME1, human "SLX4-C" or SLX4^{C}-SLX1-MUS81-EME1 (SLX4$^{1535-1834}$), and χSMX, with 10 nM DNA substrate for 1 h at 37 °C. In all cases, χSMX refers to that purified from Figure 5.4e and reaction products were resolved on a 20% polyacrylamide denaturing gel, and fluorescence visualised.

Note: the purification of SLX4-C, XPF-ERCC1, MUS81-EME1, and mini-SLX4 was performed previously by Dr. Denis Ptchelkine and Dr. Abimael Cruz-Migoni.

an encouraging initial discovery that can direct future EM data analysis, especially as a 3D reference for model generation. Attempts to refine these models through cryoelectron microscopy screening of purified χSMX were unsuccessful.

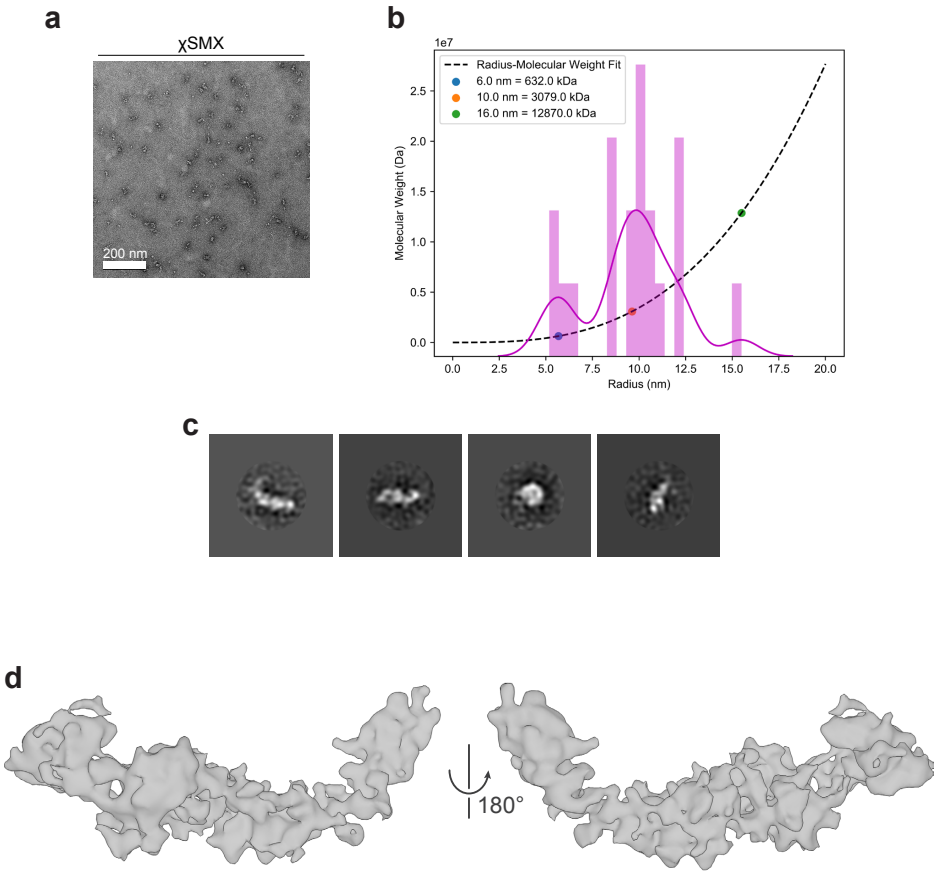


Figure 5.6 | Electron microscopy of purified χSMX complex. a, Representative negative-stain cryoelectron microscopy (EM) micrograph of the χSMX complex stained with uranyl acetate. Individual particles are visible, with a scale bar indicating 200 nm. **b,** Radius-molecular weight plot derived from the average size of particles from **a** and other micrographs obtained for χSMX. Molecular weight estimates are plotted against the radius, with observed radii and corresponding molecular weights shown for 6 nm (632.0 kDa), 10 nm (3079.0 kDa), and 16 nm (12870.0 kDa), modal particle sizes from the histogram. **c,** Representative two-dimensional class averages obtained from EM images. **d,** Three-dimensional electron density map of the 2D class averages in **c**, also shown after a 180° rotation in the right panel.

To address the challenges associated with co-infection or the combination of lysates from separate baculovirus transductions, a single baculoviral construct was pursued to deliver the chimaeric SMX components. Attempts to achieve this using the MultiBac system were unsuccessful due to difficulties in obtaining the correct clones. As an alternative, the constructs were designed and cloned into a dual-promoter baculoviral vector, pFastBac-Dual.

In the new system, the constructs included 'T-SXE' (TEV protease, xSLX4, xXPF, hERCC1) under the polyhedrin promoter and 'EMS' (hEME1, hMUS81, xSLX1) under the p10 promoter, transcribed in opposite directions from a central point in the plasmid. SLX4 was tagged with a twin-Strep-FLAG tag to enable purification using Strep-Tactin resin. Additionally, ECFP and mCherry2 tags were incorporated in-frame at the end of each open reading frame (ORF), linked by a TEV protease site. This facilitates monitoring of transfection, transduction, and expression efficiency during baculovirus generation, amplification, and initial expression trials. The difficulties encountered in achieving proper cloning and the complexities of working with a multi-species chimaeric construct ultimately led to a reassessment of the approach. Given the non-physiological nature of combining subunits from different species, further pursuit of this strategy was deemed impractical. Consequently, the focus shifted to alternative approaches that could better address these technical and biological limitations.

5.2 | Expression and purification of mouse SMX

Though the chimaeric SMX complex could be purified to adequate homogeneity, mouse homologues of these proteins present a more human-relevant model to decipher the mechanisms of action and circumvent any concerns regarding coalescing several species subunits. Concurrently, colleagues found that the purification of a subcomplex of mouse SMX (termed miniSX, containing SLX4's N-terminus (1-764), XPF and ERCC1) yielded more homogeneous, non-aggregated protein than its human or human/frog chimaeric counterpart. Furthermore, the central disordered region of mouse SLX4 and overall SLX4 is shortest in mice, 25 kDa smaller than human (Figure 1.7, upper panel). Following this, *M. musculus* SLX4, XPF, ERCC1 was cloned as a single ORF yielding one TEV-cleavable polyprotein (Figure 5.7a, upper panel; 2.2.7.2). EME1, MUS81, and SLX1 were cloned into a separate vector in a similar fashion, producing a separate single polyprotein (Figure 5.7a, lower panel). Baculoviruses expressing these two separate polyproteins were created through transfection of Sf9 insect cells, and used to transduce further populations of either Sf9 or HiFive insect cells (Section 2.3.3.1) to increase viral titres through amplification and for final expression, where transduced cells were harvested.

5.2.1 | Test purification of Strep-tagged mSMX from insect cells

An initial test expression was conducted to establish the presence of mSMX in insect cells infected with combinations of two viruses for each polyprotein. SLX4 was expressed and eluted after loading onto Strep-Tactin XT 4Flow resin, which implied at least the presence of Strep-tag-bearing ERCC1 and its obligate heterodimer XPF, though XPF-ERCC1 could not be visualised by SDS-PAGE (Figure 5.7c). By SEC analysis, SLX4 appeared to be present in a separate complex to XPF and TwinStep-FLAG (SF)-ERCC1 (Figure 5.7d, e, f), despite co-eluting from the Strep-Tactin resin. The size of the peak containing SLX4 here also corresponds to a size far greater than a monomer of SLX4, and more likely to be a tetramer of SLX4.

5.2.2 | Expression and purification of mS$^{\Delta\text{central}}$MX from insect cells

The expression of full-length mSLX4 suggested potential issues with co-folding alongside XPF and ERCC1. It was hypothesised that deleting the central region of mSLX4 could enhance expression and eliminate the need for cofactors that might complicate SLX4 purification (Figure 5.7g, h). This central deletion, spanning residues 759–1252 (Section 2.2.7.2), was designed to address this issue. Deletion of this extended, largely disordered region may enhance expression by promoting a more stable and compact SLX4 fold. Additionally, deletion of the intravening SUMO-interacting motifs (SIMs) was considered beneficial, as these motifs may contribute to protein condensate formation[604], potentially impeding overexpression or promoting insolubility. In these plasmids, nuclease-inactive mutations were introduced in XPF (D706A), MUS81 (D307A), and SLX1 (R38A, E79A) to facilitate the expression of later complexes with various combinations of inactive nucleases (Section 2.2.7.2).

To generate the truncated SMX complex, baculoviruses were produced as previously described. During the amplification of these baculoviruses, the correct expression of the SMX complex was monitored *via* FLAG (tracking ERCC1 of the "SXE" component) and His-tag (tracking SLX1 of the "EMS" component) (Figure 5.8a) abundance by Western blot. For purification, the SXE and EMS components were expressed separately in insect cells and combined prior to cell lysis (Figure 5.8a). Initial affinity purification using

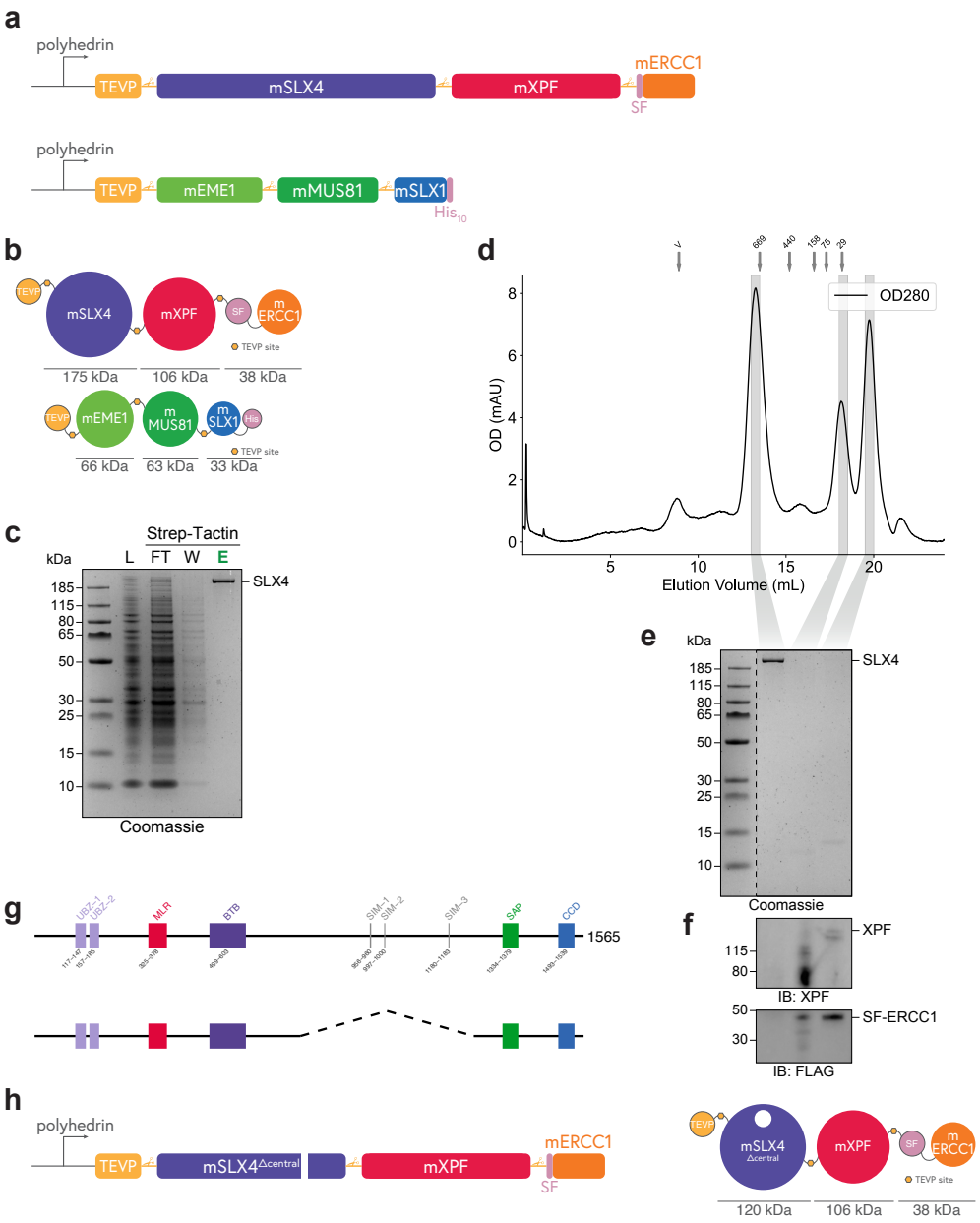


Figure 5.7 | Design and expression trial of mouse SMX constructs. a, Schematic of the mouse SMX construct including mSLX4, mXPF, and mERCC1. Domains are tagged with TEV protease sites (TEVP), a TwinStrep-FLAG Tag (SF), and a His_{10} tag for purification. **b,** The corresponding molecular weights are shown for each component (mSLX4: 174 kDa, mXPF: 106 kDa, mERCC1: 38 kDa). **c,** SDS-PAGE analysis of StrepTrap purification of mSLX4-containing constructs as in **a, b**. Lane L: lysate; FT: flow-through after loading onto the StrepTrap column; W: wash fraction; E: elution fraction. SLX4 is observed in the elution fraction. **d,** Size exclusion chromatography (SEC) profile (Superose 6 Increase 10/300 GL) of StrepTrap elution. Arrows indicate molecular weight standards, and grey-shaded regions denote collected fractions. **e,f,** SDS-PAGE analysis of SEC fractions from the grey-shaded regions in **d**, with Western blotting for XPF and ERCC1 in **f**. **g,** Schematic of an alternative construct, $mSLX4^{\Delta central}$-mXPF-mERCC1, featuring a deletion of the central region of mSLX4. Domains are similarly tagged, and molecular weights are provided on the right panel.

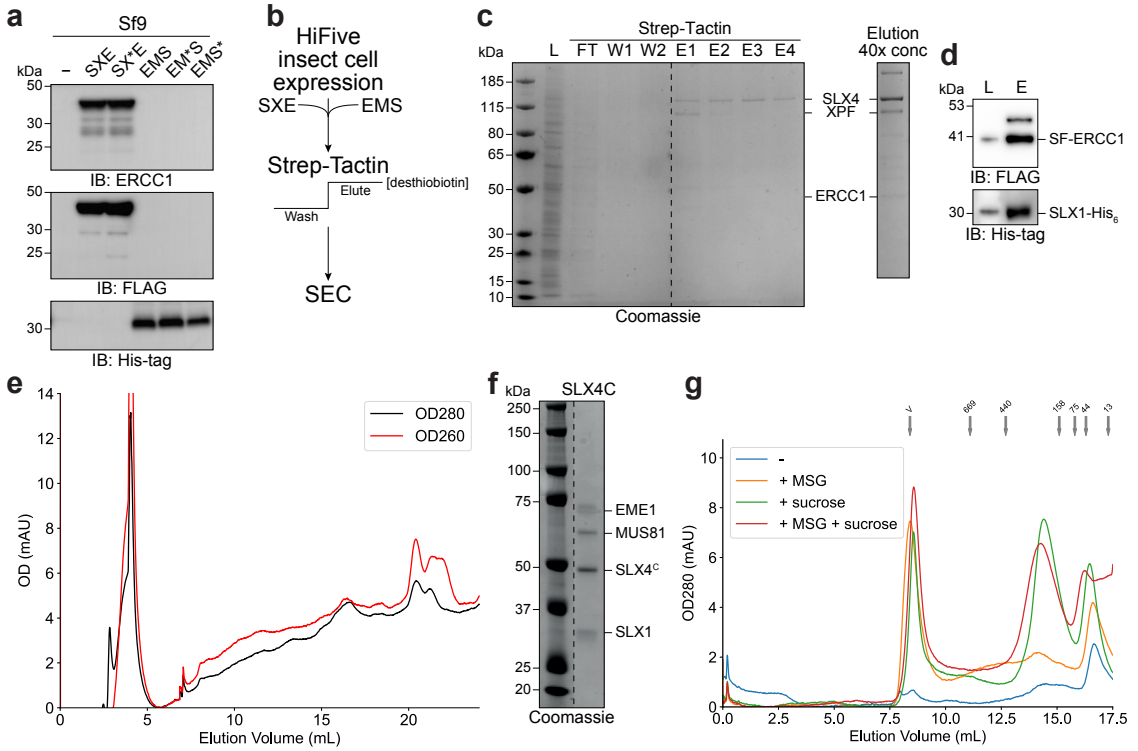


Figure 5.8 | Expression and purification of mS$^{\Delta central}$MX components and assembly analysis. **a,** Western blot analysis of mS$^{\Delta central}$MX components expressed in Sf9 cells during the final baculovirus amplification to confirm expression of mSLX4-XPF-ERCC1 (SXE) by anti-FLAG and anti-ERCC1 and EME1-MUS81-SLX1 (EMS) by anti-His-tag. Nuclease-inactive containing variants of the baculovirus are also shown; SX*E - (XPF^{D706A}); EM*S - MUS81^{D307A}); EMS* - SLX1^{R38A,E79A}. **b,** Workflow for the expression and purification of mS$^{\Delta central}$MX components. The complex was expressed in HiFive insect cells, purified using Strep-Tactin XT resin, and subjected to size-exclusion chromatography (SEC). **c,** SDS-PAGE analysis of Strep-Tactin purification fractions, stained with Coomassie. Lane L: lysate; FT: flow-through; W1-W2: wash fractions; E1-E4: elution fractions. The final elution fraction (E4), concentrated 40-fold, highlights the enrichment of SLX4, XPF, and ERCC1. **d,** Western blotting of Strep-Tactin elutions using anti-FLAG and anti-His, confirming the presence of SF-ERCC1 and SLX1-His in the Strep-Tactin elutions. **e,** SEC profile (Superose 6 Increase 10/300 GL) showing OD_{280} (black) and OD_{260} (red) absorbance profiles of mS$^{\Delta central}$MX purification. Peaks correspond to distinct molecular weight populations. **f,** Coomassie-stained SDS-PAGE analysis of purified SLX4-C (SLX1, SLX4^{C}, MUS81, and EME1) complex, which was run on the same size exclusion chromatography method as **e**. **g,** SEC profile (Superose 6 Increase 10/300 GL) showing OD_{280} of SLX4-C under various buffer conditions, including the addition of monosodium glutamate (MSG) or sucrose, or a combination of both. Grey arrows denote molecular weight standards.

SF-tagged ERCC1 with Strep-Tactin resin yielded a small amount of the SMX components detectable by SDS-PAGE, with clear bands for SLX4 and XPF (Figure 5.8c). However, EMS components were only identified or their presence inferred through Western blotting against the His-tag (Figure 5.8d).

Size-exclusion chromatography (SEC) of 40× concentrated Strep-Tactin eluates did not reveal any significant peaks (Figure 5.8e). A subsequent purification trial produced similar results (data not shown), but a noticeable peak was observed during a wash step

using 6 M guanidinium hydrochloride, suggesting that some protein material had attached to the column. Given that the nucleases of SMX bind nucleic acids, it is possible that the complex might interact with the sugar-based Superose columns, mimicking DNA-binding interactions with ribose sugars. To address this issue, it was suggested that including additives that either neutralise the basic charge of DNA-binding proteins or quench the resin might aid the progression of the protein complex through the column (personal communication with Prof. Alex Vecchio, University of Nebraska–Lincoln). With limited mSMX samples available, testing whether these additives could alleviate issues at the SEC stage was achieved by purifying the C-terminal SLX4 complex (SLX4-C), which consisted of human SLX4$^{1535-1834}$, SLX1, MUS81 and EME1, previously purified by Dr. Abimael Cruz-Migoni (Figure 5.8f). When the C-terminal SLX4 complex was run independently on a Superose 6 column, no peak corresponding to the expected molecular size ($\sim$190 kDa) was observed. The addition of 200 mM monosodium glutamate (MSG) slightly improved peak detection ($\sim$14.5 mL), while 10% sucrose significantly increased the predicted band intensity, both with and without 200 mM MSG (Figure 5.8g). However, these additives pose challenges for downstream cryo-EM applications, where simple buffers are preferred, as small molecules such as sucrose and MSG can increase background scattering.

Despite these limitations, the subsequent purification adopted these additives after Strep-Tactin elution (Figure 5.9a). Once again, mSLX4, mXPF, and mERCC1 were observed in the elutions from Strep-Tactin (Figure 5.9b, c). This time, 10% sucrose was included during SEC and resulted in two small peaks at 13 mL (peak 1) and 15.5 mL (peak 2) (Figure 5.9d). Concentrating each peak approximately 20-fold and analysing them *via* SDS-PAGE revealed that only the second peak contained SMX components, comprising solely the SXE constituents (Figure 5.9e, f). Western blotting indicated that His-tagged SLX1 was absent, suggesting that the EMS subunits did not co-purify with SXE (Figure 5.9f, lower panel). The purified SXE complex demonstrated endonuclease activity on a simple fork substrate (Figure 5.9g). The lack of EMS subunits may reflect issues stemming from the deletion of the SLX4 central region or, more plausibly, that split expressions in separate insect cell populations hinder proper co-folding and assembly of the complete SMX complex.

To address co-folding challenges, a co-infection strategy was implemented, where SXE and EMS viruses were simultaneously transduced into the same cell population. This

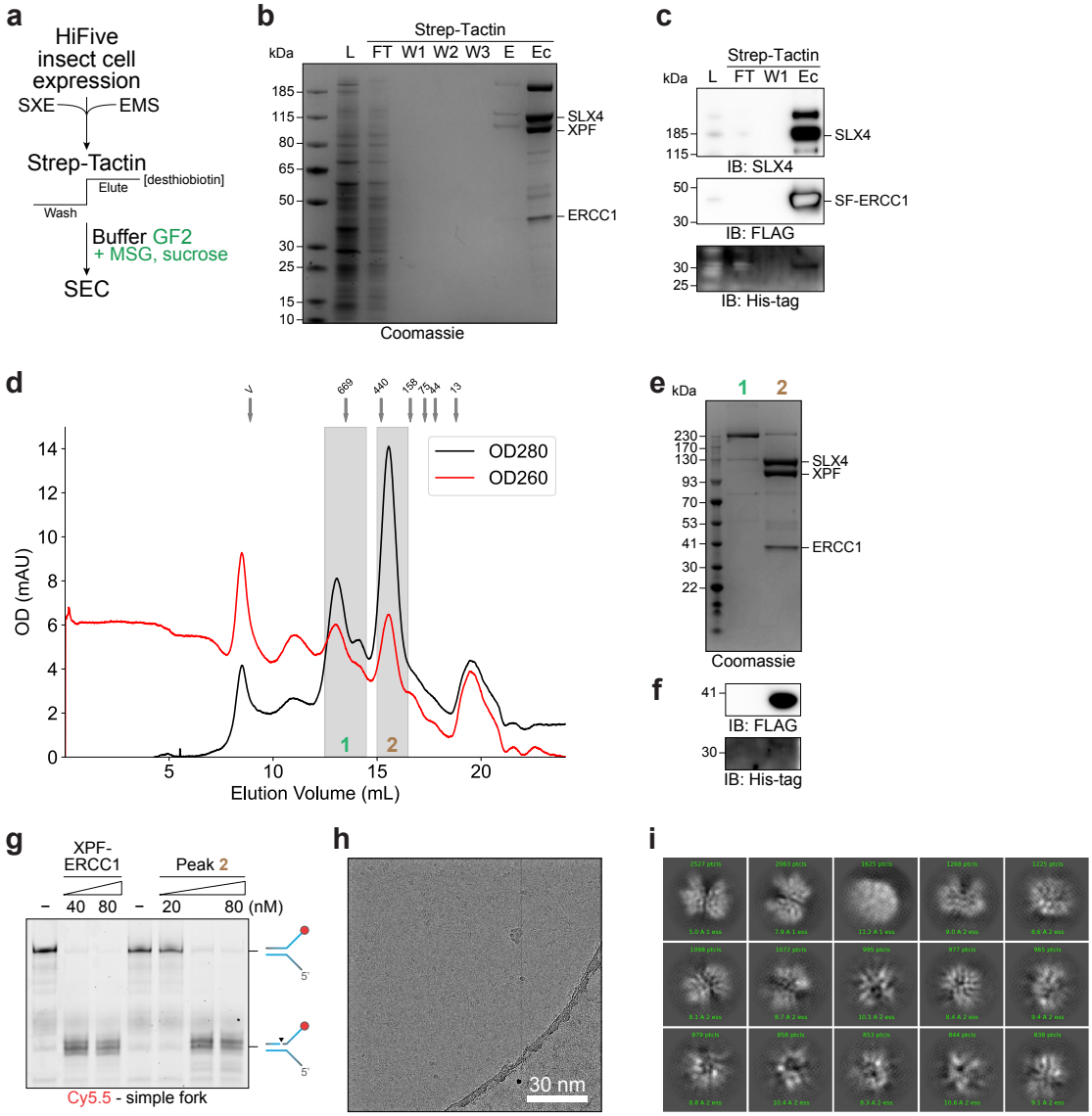


Figure 5.9 | Expression and purification of mS$^{\Delta central}$MX complexes with buffer optimisation and structural analysis. a, Workflow for the expression and purification of mS$^{\Delta central}$MX complexes. SXE mSLX4$^{\Delta central}$-mXPF-mERCC1) and EMS (mEME1-mMUS81-mSLX1) were expressed in HiFive insect cells, purified using Strep-Tactin resin, and subjected to SEC in Buffer GF2 supplemented with MSG and sucrose. **b,** SDS-PAGE analysis of Strep-Tactin fractions. Lane L: lysate; FT: flow-through; W1-W3: wash fractions; E: elution; Ec: concentrated elution fraction. **c,** Western blot analysis of Strep-Tactin purification fractions using anti-SLX4, anti-FLAG, and anti-His-tag, confirming the presence of SF-mERCC1 and mSLX1-His in elution fractions. **d,** SEC chromatogram (Superose 6 Increase 10/300 GL) showing OD_{280} (black) and OD_{260} (red) absorbance profiles for mS$^{\Delta central}$MX. Peaks 1 and 2, shaded grey, correspond to distinct protein fractions collected for analysis. Grey arrows denote molecular weight standards. **e,** SDS-PAGE of Peak 1 and Peak 2 fractions from SEC, showing enrichment of mSLX4, mXPF, and mERCC1. **f,** Western blot of Peak 2 from SEC in **e**, probing for the presence of FLAG (mERCC1) and His-tag (SLX1). **g,** Endonuclease cleavage of a Cy5.5-labelled simple fork was verified *via* a gel-based assay. Substrate processing was assessed with the incubation of two-fold increasing concentrations (20 to 80 nM) of human XPF-ERCC1, and peak 1 elution from **e**, with 10 nM DNA substrate for 1 h at 37 °C. Reaction products were then resolved on a 20% polyacrylamide denaturing gel, and fluorescence visualised. **h,** Cryo-EM micrograph of purified mS$^{\Delta central}$MX complexes. Scale bar: 30 nm. **i,** Representative 2D class averages obtained from cryo-EM, showing distinct structural features of the mS$^{\Delta central}$MX complex in various orientations, with corresponding class resolutions.

approach aimed to facilitate proper SMX assembly through multiple gene expressions within the same cellular environment. However, as in previous trials, Strep affinity eluates contained only SXE components (data not shown). While isolating the SXE complex was not the primary goal of the project, the arrangement and orientation of the XPF-ERCC1 heterodimer by SLX4 remain of significant interest. Preliminary cryo-EM investigations were conducted using vitrified mSXE particles (Figure 5.9h). Reference-free 2D classification yielded particle averages suggesting partial two-fold symmetry, though heterogeneity was evident in the micrographs and projections (Figure 5.9h, i). Some degree of particle denaturation or aggregation was also observed, raising concerns about sample stability, especially given that vitrification required a fast buffer exchange to remove MSG and sucrose (Section 2.5.6).

5.2.3 | Expression of mS$^{\Delta central}$MX in insect cells from a single baculovirus

Since co-infection did not achieve whole SMX assembly, it was reasoned that placing all genes on a single baculovirus plasmid might be more effective. To this end, the SXE open reading frame (ORF), including an upstream TEV protease and a Halo-tag upstream of XPF, and EGFP upstream of ERCC1, was cloned into pFastBac-Dual — a plasmid containing a p10 promoter adjacent to an oppositely oriented polyhedrin promoter[841]. The EMS ORF sequence was then cloned downstream of the p10 promoter, enabling single-baculovirus expression of both complexes (Figure 5.10a,b).

The inclusion of an EGFP tag for ERCC1 provided a two-fold benefit: it enabled monitoring of the real-time fluorescence levels of cells, which ought to correlate with expression levels, and it facilitated purification. This modification allowed for more efficient validation and amplification of baculoviruses, replacing the need for Western blotting at each stage with fluorescence-based confirmation of transduction and viral particle amplification. A test expression was conducted to compare the effects of 48-hour and 72-hour transduction periods on complex assembly and subsequent large-scale purification (Figure 5.10c). No significant differences in expression levels were observed when assessing EGFP fluorescence after 48 or 72 hours (Figure 5.10d). However, after incubation

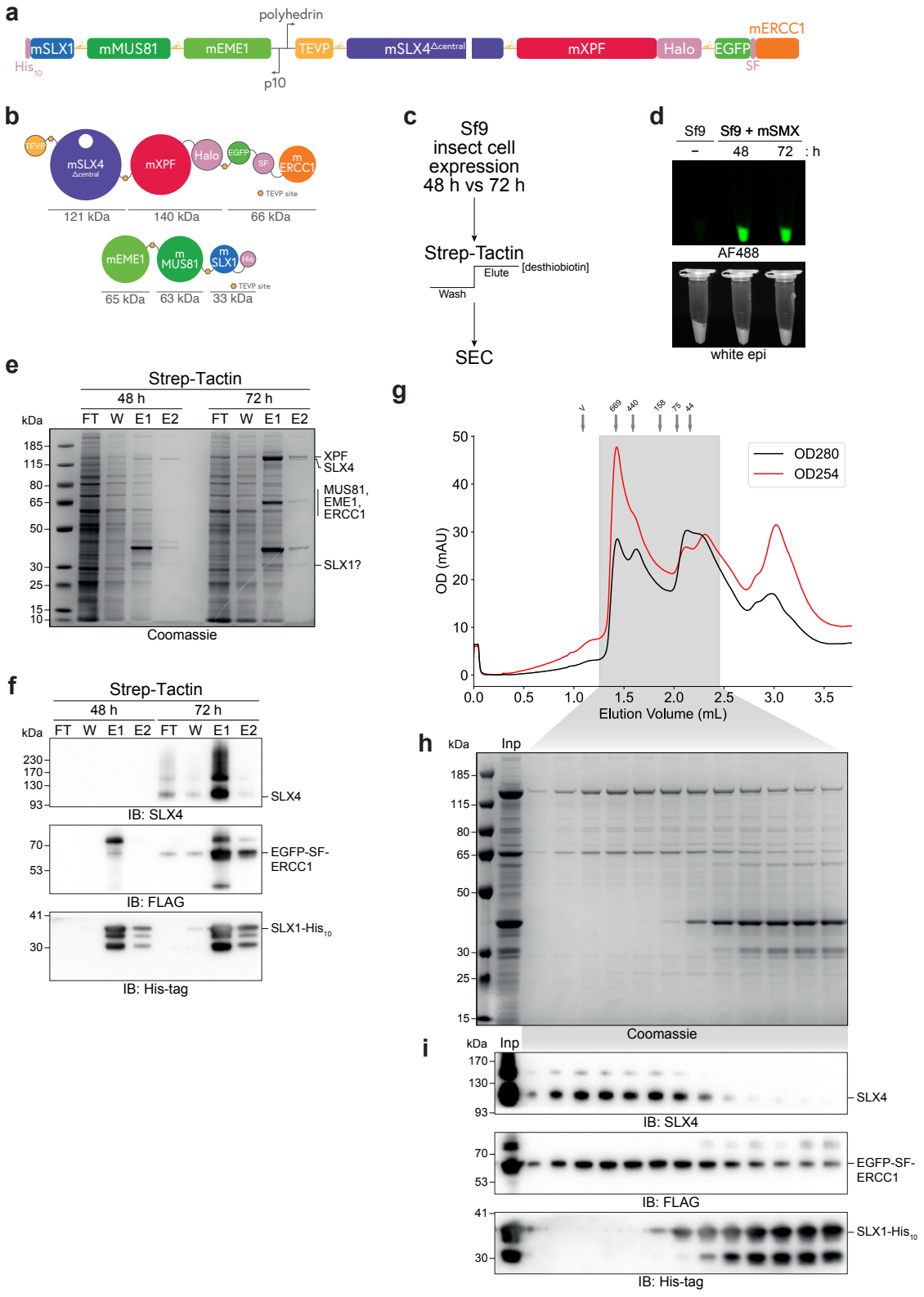

a
polyhedrin
mSLX1
mMUS81
mEME1
TEVP
mSLX4Δcentral
mXPF
Halo
EGFP
mERCC1
His10
SF
p10
b
mSLX4 Δcentral
mXPF
Halo
EGFP
SF
mERCC1
TEVP site
121 kDa
140 kDa
66 kDa
mEME1
mMUS81
mSLX1
65 kDa
63 kDa
33 kDa
c
Sf9 insect cell expression 48 h vs 72 h
Strep-Tactin
Elute
[desthiobiotin]
Wash
SEC
d
Sf9
Sf9 + mSMX
–
48
72
: h
AF488
white epi
e
Strep-Tactin
48 h
72 h
kDa
FT W E1 E2 FT W E1 E2
185
115
80
65
50
30
25
15
10
XPF
SLX4
MUS81,
EME1,
ERCC1
SLX1?
Coomassie
f
Strep-Tactin
48 h
72 h
kDa
FT W E1 E2 FT W E1 E2
230
170
130
93
SLX4
IB: SLX4
70
53
EGFP-SF-
ERCC1
IB: FLAG
41
30
SLX1-His10
IB: His-tag
g
V
669
440
158
75
44
OD280
OD254
OD (mAU)
Elution Volume (mL)
h
kDa
Inp
185
115
80
65
50
30
25
15
Coomassie
i
kDa
Inp
170
130
93
SLX4
IB: SLX4
70
53
EGFP-SF-
ERCC1
IB: FLAG
41
30
SLX1-His10
IB: His-tag

Figure 5.10 | Expression and first purification of mS$^{\Delta central}$MX under a single baculovirus. a, Schematic representation of the mS$^{\Delta central}$MX construct used for single-baculovirus transduction of Sf9 insect cells. Components include mSLX1-His$_{10}$, mMUS81, mEME1, mSLX4$^{\Delta central}$, mXPF-Halo, and mERCC1-EGFP-SF. **b,** Expected molecular weights of individual components and their fluorescent or affinity tags. SF - TwinStrep-FLAG; His - His$_{10}$-tag; TEVP - TEV protease; orange hexagons represent TEV protease cleavage sites. **c,** Workflow for the expression and purification of mS$^{\Delta central}$MX in Sf9 cells, comparing 48 h and 72 h post-infection expression times. Purification included affinity purification using Strep-Tactin resin and size exclusion chromatography (SEC). **d,** Fluorescent imaging of Sf9 cell pellets expressing mS$^{\Delta central}$MX visualising EGFP (mERCC1) expression through the Alexa Fluor 488 channel (AF488) and under white epi-illumination. **e,** SDS-PAGE analysis of Strep-Tactin purification fractions. FT: flow-through; W: wash fractions; E1-E2: elution fractions. **f,** Western blot analysis of Strep-Tactin purification fractions probing for SLX4, FLAG (mERCC1) and His-tag (mSLX1). **g,** SEC (Superdex 200 Increase 5/150 GL) chromatogram showing OD$_{280}$ (black) and OD$_{254}$ (red) absorbance profiles of Strep-Tactin eluted mS$^{\Delta central}$MX. Shaded regions correspond to the elution fractions analysed in subsequent SDS-PAGE and Western blot experiments. **h,** SDS-PAGE analysis of SEC fractions from the shaded region in **g**. **i,** Western blot analysis of SEC fractions using antibodies against SLX4, FLAG (mERCC1), and His-tag (mSLX1), confirming the assembly of mS$^{\Delta central}$MX components in the eluted fractions.
Note: The protein purification presented here was carried out by Dr. William Foster.

with Strep-Tactin resin, the 72-hour transduction sample yielded significantly higher protein content, with mS$^{\Delta central}$MX components discernible by SDS-PAGE alone (Figure 5.10e).

Western blotting confirmed that mSLX4$^{\Delta central}$, ERCC1, and SLX1 were present at higher concentrations in the 72-hour transduction eluates (Figure 5.10f). Analytical size-exclusion chromatography (SEC) of the Strep-Tactin eluate from the 72-hour transduction revealed polydisperse protein content (Figure 5.10g, h). Surprisingly, Western blotting demonstrated that larger complexes contained mSLX4$^{\Delta central}$, while mSLX1 was predominantly confined to smaller complexes (Figure 5.10i, upper and lower panels). ERCC1, and by association XPF, were present throughout the peaks but accumulated more significantly in the larger complexes (Figure 5.10i), middle panel). MUS81 and EME1 were not probed in this preparation due to insufficient material. Despite being a preliminary purification, the high expression levels of SLX1 and ERCC1 were promising for the scope of this system, however understanding why mSLX1 and mSLX4 were not present in similar-sized species requires further study. Complexes from the central region of the SEC profile (Figure 5.10g, h, i), where SLX1, ERCC1, and SLX4 overlapped ($\sim$1.7 mL), were further concentrated and frozen for subsequent enzymatic activity assays and cryo-EM screening.

5.3 Expression and purification of human SLX4-scaffolded complexes from mammalian cells

The expression of human protein complexes in insect cells, while widely used due to the scalability and ease of baculovirus-mediated transduction, often fails to replicate the exact nature of human native folding, specific post-translational modifications (PTMs), and interactions found in mammalian systems. This has been particularly evident in the case of complex multi-protein assemblies such as the SMX complex, where attempts to express human components in insect cells have been met with limited success. Insect cells can accommodate the expression of large protein complexes, and are capable of depositing PTMs, such as phosphorylation, ubiquitylation, and glycosylation, which is a key rationale for their use over bacterial expression systems. However, these PTMs, which are often essential for the proper folding and functionality of these proteins[842] can be deposited inaccurately in ectopically expressed proteins. This limitation presents a strong justification for exploring human mammalian cell expression systems as a viable alternative. Another benefit to mammalian cell expression is the shorter experimental timeline between plasmid and protein expression - while requiring, at minimum, 10 days for baculovirus amplification, transient transfection of mammalian cells can be performed immediately once sufficient plasmid is available. However, transient transfection for large-scale protein expression using commercial transfection reagents such as Lipofectamine is not financially feasible in the long term. Therefore, after establishing an Expi293F culture system, the transfection efficiency of Lipofectamine was compared to that of polyethyleneimine (PEI). PEI-transfection resulted in a comparable if not more significant proportion of GFP+ cells, allowing the use of this system for larger-scale mammalian SMX expressions (Appendix B.7, Figure B.7). Therefore, human SMX genes were cloned into plasmids with expression under the cytomegalovirus (CMV) promoter and further detailed in 2.2.7.3; these plasmids were sub-cloned from pTLCV2 plasmids generated in 5.3.4. In this case, the SLX1 and SLX4 genes were brought together into one plasmid (‘SS’), since insights from mouse SMX purifications suggested that the stability/expression of SLX1 was dependent on simultaneous co-expression of SLX4 (Figure 5.11a). Naturally, XPF and ERCC1 were grouped on a second plasmid (‘XE’), while MUS81 and EME1 (‘ME’) were cloned into a third plasmid (Figure 5.11a). Different affinity tags were placed on almost all genes

to expand options for purification: SLX4 was fused to an SF tag, SLX1 was fused to EGFP, ERCC1 was fused to an Avi-His_10 that can be biotinylated by the presence of a downstream BirA gene, EME1 was fused to HA-Halo (Figure 5.11a). The SF tag allows purification by Strep-Tactin resin as adopted for mouse SMX 5.2.1. However, there are further options made available by the remaining tags: biotinylated ERCC1-Avi-His can be purified either through XPF or *via* the His-tag. This also enables future purifications of individual subunits or sub-complexes. For example, deletion of SLX4's N-terminus in the SS plasmid could be co-transfected with the 'ME' plasmid to express human SLX4-C in Expi293F cells. Alternatively, a form of 'mini-SLX4' can be expressed by co-transfection of a C-terminal SLX4/SLX1 gene-deleted 'SS' plasmid alongside 'XE'.

5.3.1 | Exploring an advantage of Expi293F cells as an expression system

Another benefit from using human cell lines for expression is the option of treating cells before harvest with agents known to influence the regulation of human SMX subunits. G2/M confined CDK1-cyclin B activity adds phosphate groups at SLX4's MUS81-binding region within the SAP domain, specifically at T1544, T1561, and T1571[577,578]. Phosphorylation of these residues increases the affinity of this short SLX4 peptide towards MUS81-EME1. This highlights the benefit of using a system that recapitulates PTMs perfectly. In Expi293F cells, human CDK1 can phosphorylate SLX4, increasing the strength of its association with MUS81-EME1 and likely improving protein preparation yields, especially if cells are synchronised in G2/M. Human cells can be robustly arrested in G2/M through overnight treatment with nocodazole, and to confirm that this was possible with suspension Expi293F cells, varying concentrations of nocodazole were tested to monitor arrest. Cell cycle analysis showed that G2/M synchronisation was achieved only at a concentration of 250 ng/ml, five times the concentration suggested for adherent human cancer cell lines[843] (Figure 5.11b, c). Treatment with higher concentrations of nocodazole did not further increase the fraction of cells in G2/M (Figure 5.11b, c). It is worth noting that insect cells and other expression systems can also be synchronised in G2/M but require a far greater drug concentration, and this process is far less efficient (see section later).

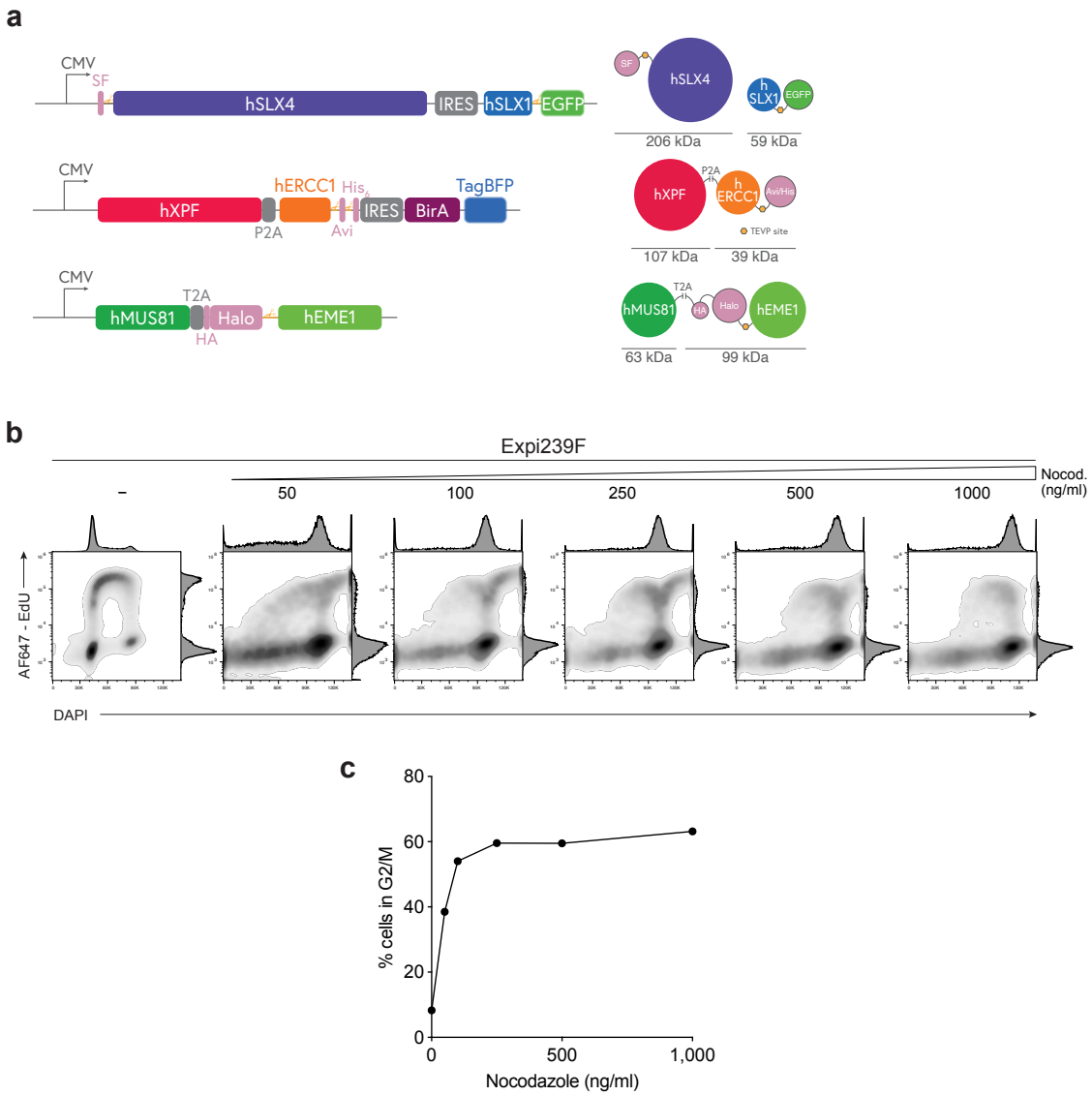


Figure 5.11 | Design of human SMX constructs and flow cytometry analysis of nocodazole-induced G2/M arrest. a, Schematic representation of hSMX components designed for mammalian expression under CMV promoters. Constructs include SF-hSLX4 with hSLX1-EGFP co-expressed after an internal ribosome entry site (IRES), hXPF and hERCC1-Avi-His_{10} (co-expressed with a P2A peptide linker) with BirA-TagBFP co-expressed after an IRES, and hMUS81-hEME1 (co-expressed *via* T2A). Tags include SF - TwinStrep-FLAG; His - His_{10}; Avi - biotinylation tag; Halo - HaloTag. The molecular weights of individual components and their tags are indicated in kilodaltons (kDa) on the right panel. **b,** Cell cycle analysis of Expi293F cells treated with increasing concentrations of nocodazole (0–1000 ng/ml) for 16 h to induce G2/M arrest. Cells were stained with DAPI for DNA content and EdU for S-phase detection. Representative density plots of AF647-EdU *versus* DAPI signal are shown for each condition. **c,** Quantification of the percentage of cells in G2/M phase based on the flow cytometry profiles shown in **d**

5.3.2 | Transient transfections of SMX in Expi293F cells

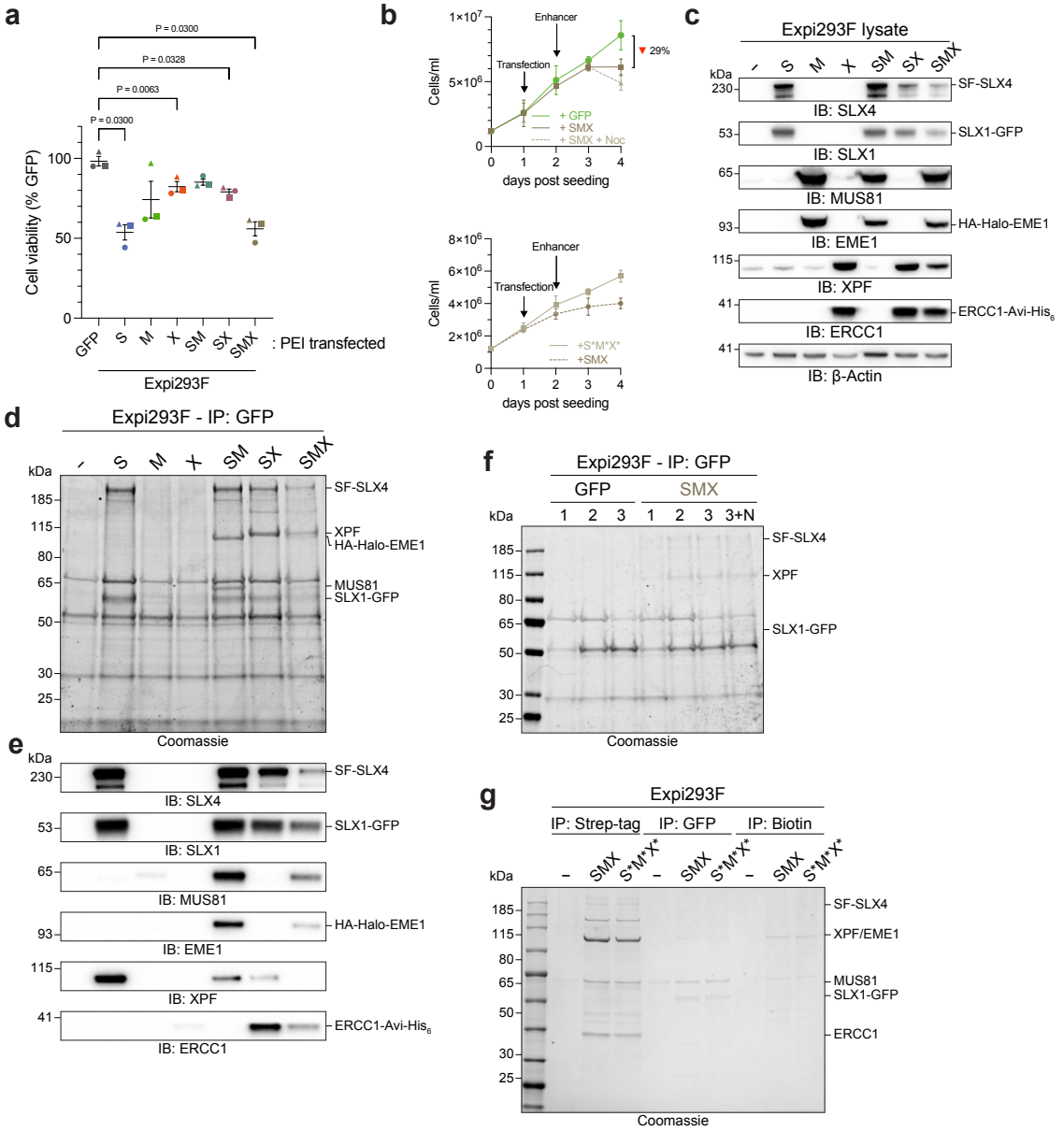
The expression and purification of the SMX complex in Expi293F mammalian cells highlighted several challenges associated with toxicity and protein assembly. Observations during initial experiments suggested that overexpression of SMX components could significantly reduce cell viability post-transfection. Therefore cell counts were quantified through the transfection protocol and showed that equal transfection of plasmids encoding SMX components in Expi293F cells resulted in reduced viable cell numbers compared to an empty EGFP vector control, particularly when overexpressing the entire SMX complex or the SLX4-SLX1 ('S') subcomplex (Figure 5.12a). Viability dropped by approximately 50% under these conditions, underscoring the cytotoxic impact of SMX overexpression. Monitoring cell growth over time further revealed that SMX overexpression had minimal impact within the first 48 hours, but by 72 hours, growth was hampered by approximately 30% relative to EGFP control transfections (Figure 5.12b, upper panel). In contrast, the transfection of SMX subunits with nuclease-inactive variants: SLX1^{E41A,R82A}, XPF^{D705A} MUS81^{D307A} (S*M*X*) maintained viability during expression compared to enzymatically active SMX (Figure 5.12b, lower panel).

These findings guided subsequent approaches in two key ways. First, expressions were limited to 48 hours post-transfection to minimise cytotoxic effects while still allowing sufficient protein production. Second, the potential for controlled, inducible expression systems was considered as a means to mitigate toxicity, balancing cell viability with adequate SMX expression. However, with transient transfection of all combinations of SMX subunits, Western blot analysis confirmed the expression of all SMX components in Expi293F cells as transfected, although a slight reduction in SLX4 and SLX1 levels was observed when expressing the entire SMX complex (Figure 5.12c). With a GFP immunoprecipitate (using home-made GFP nanobody-conjugated agarose beads) of these expressions, SLX4 and SLX1-EGFP could be observed where 'S' was expressed, missing in untransfected Expi293F cells (Figure 5.12d), confirmed by Western blotting (Figure 5.12e). XPF and ERCC1 were observed in SX and SMX expressions by SDS-PAGE (Figure 5.12d), though Western blotting only confirmed ERCC1 presence by SMX immunoprecipitate (Figure 5.12e). MUS81-EME1 was seen as expected in SM and SMX, confirmed by Western blotting (Figure 5.12d, e). Purification trials using various affinity tags and bead systems

were performed to identify optimal methods for isolating SMX components. In expressions of SMX compared to GFP alone, no significantly increased expression/immunoprecipitate of SLX4 nor XPF was observed beyond 2 days post-transfection, nor with the treatment of nocodazole (Figure 5.12f). In all these immunoprecipitates, however, many contaminating proteins were present, likely attributed to the poorer quality of nanobodies from our preparations compared to commercial GFP nanobody-conjugated beads (ChromoTek; used hereafter). To identify the best strategy for affinity purification of SMX, three different tags were immunoprecipitated (Strep-tag on SLX4, GFP on SLX1 and biotinylated ERCC1) using their counterpart beads and boiled to release bound proteins. SMX and S*M*X* appeared to best bind to the Strep-tag as the greatest yield of SMX components was obtained from this affinity purification, with this likely representing the most appropriate strategy for future purifications.

5.3.3 | Purification of human MUS81-EME1 by transient transfection in Expi293F cells

To assess the versatility of the Halo-tag system for purification, the SLX4-C complex (comprising SLX4$^{1535-1834}$, SLX1, MUS81, and EME1) was expressed in Expi293F cells and applied to magnetic Halo-tag capture beads (Section 2.4.2.3). Elution from Halo-tag beads requires proteolytic cleavage, as the Halo-tag forms a covalent bond to the beads. Halo-tagged TEV protease was used for cleavage, releasing SLX4-C components while remaining bound to the beads themselves, thereby eliminating the need for downstream TEV protease removal. The efficiency of TEV protease cleavage was evaluated by splitting the beads into three fractions: pre-cleavage boiled beads, TEV protease elution, and post-cleavage boiled beads following washing. TEV protease was also run on SDS-PAGE at the same concentration used in the reaction, to determine the fraction of TEV protease that remains bound to the beads. Boiling the beads before TEV cleavage released MUS81, while TEV cleavage successfully eluted EME1 (Figure 5.13c, d) and MUS81 (Figure 5.13c, e). The identities of MUS81 and cleaved EME1 were confirmed by mass spectrometry (Figure 5.13c). However, some MUS81 and EME1 (though separated from the Halo-tag) remained attached to the beads, as boiling the post-cleavage beads released additional quantities of these proteins (Figure 5.13c, e). TEV protease largely remained bound to



the beads due to its own Halo-tag, as indicated by SDS-PAGE (Figure 5.13c). Despite the successful expression of SLX1 and SLX4 (Figure 5.13d, e), these components did not co-elute with MUS81 and EME1, suggesting incomplete complex formation. While the absence of SLX4 co-elution was unexpected, the successful purification of MUS81-EME1 through this method was encouraging, particularly as previous attempts using insect cell expression systems had proven challenging.

To determine whether immunoprecipitation using alternative tags could enable co-purification of SLX4 components, additional preparations were conducted using both SLX4-C and the C-terminal complex with full-length SLX4 (SLX4-SLX1-MUS81-EME1)

Figure 5.12 | Transient expression, immunoprecipitation, and cell viability assays for human SMX components in Expi293F cells. a, Cell viability of Expi293F cells transfected with 10 µg of GFP alone, individual SMX components (S, M, X), or their combinations (SM, SX, SMX) to a total of 10 µg of all plasmids. Cells were transfected using PEI, and viability was measured as a percentage of GFP-positive cells after 2 days post-transfection. Error bars denote SD from n = 3 biological repeats, each a mean of three technical measurements. A repeated-measures (RM) one-way ANOVA with Geisser-Greenhouse correction followed by a Holm-Sidak posthoc multiple comparison test was used to calculate significance and only $p < 0.05$ is shown. **b,** Growth curves of Expi293F cells transfected with 10 µg GFP alone or SMX components to a total of 10 µg of all plasmids, with or without nocodazole (Noc). Top: GFP and SMX with and without nocodazole; Bottom: SMX *versus* nuclease-dead SLX1, MUS81 and XPF mutants (S*M*X*). Arrows indicate transfection and enhancer addition. Cell growth was monitored over four days post-seeding. Error bars denote SD from n = 3 biological repeats. **c,** Western blot analysis of lysates from Expi293F cells expressing individual (S: SLX4-SLX1, M: MUS81-EME1, X: XPF-ERCC1) or combined SMX components, probed for SLX4, SLX1, MUS81, EME1, XPF, ERCC1, and β-actin as a loading control. **d,** Immunoprecipitation (IP) of GFP-tagged proteins from Expi293F cells expressing SMX components. SDS-PAGE confirms the presence of SMX subunits in immunoprecipitated complexes. **e,** Western blot analysis of immunoprecipitated complexes from **d**, probing for SLX4, SLX1, MUS81, EME1, XPF, and ERCC1 to confirm SMX assembly. **f,** GFP IP of SMX components from Expi293F cells transfected with SMX or GFP alone comparing immunoprecipitates from expressions harvested 1-3 days post-transfection, and including 250 ng/ml nocodazole for 24 h in the last lane (3+N). SDS-PAGE confirms the co-presence of SMX subunits (SF-SLX4, XPF, SLX1-GFP) in the immunoprecipitated complexes. **g,** Strep-tag, GFP, and biotin IP of SMX components from Expi293F cells expressing SMX components. SDS-PAGE confirms the presence of SMX subunits (SF-SLX4, MUS81, SLX1-GFP, EME1, XPF, and ERCC1) across the different IP methods. In the case of biotin IP, expression of SMX or S*M*X* included 10 µM biotin for the final 24 h to induce biotinylation of ERCC1-Avi.

(Figure 5.13a, b). Expression of all components was confirmed, except for SLX4-C, which was inferred by its presence in immunoprecipitates (Figure 5.13f). The experiment was performed with and without nocodazole treatment to induce G2/M arrest, which had no discernible effect on the expression of any subunit (Figure 5.13f), in agreement with constant steady-state expression of SMX subunits[571]. SLX4 was pulled down using its Strep-tag, SLX1 through its GFP tag, and EME1 by biotinylating its Halo-tag with a HaloTag PEG-biotin ligand, followed by capture with streptavidin Dynabeads. Strep-tag pull-down of SLX4 enriched SLX4-C subunits, though EME1 was largely absent as determined by Western blotting (Figure 5.13g). In contrast, immunoprecipitation *via* SLX1-GFP resulted in co-purification of all components, as shown by Western blotting (Figure 5.13h). Halo-tag-based pull-down of EME1 did not appreciably co-elute SLX1 or SLX4 (Figure 5.13i). This may be attributed to insufficient elution of EME1, as no SLX4-C or full SLX4 complex subunits were visible by SDS-PAGE in this preparation (Figure 5.13g, h, lower panel), unlike the GFP or Strep pull-downs (Figure 5.13i, lower panel). These experiments also sought to evaluate whether nocodazole-induced G2/M arrest altered the phosphorylation patterns of SLX4-C or full-length SLX4 complexes.

No significant band shifts indicative of altered phosphorylation were observed for either SLX4-C or SLX4-SLX1-MUS81-EME1 under nocodazole treatment (Figures 5.13g–i). These results suggested that nocodazole treatment does not affect the post-translational modification or assembly of SLX4 complexes.

5.3.4 | Generating stable cell lines inducibly expressing SMX

A strategy was developed to generate Expi293F cells capable of inducibly expressing SMX components, either as individual subunits or as an entire stable complex in one cell line. The pTLCV2 plasmid served as the backbone for this inducible system, initially designed for Cas9 expression[844] but recently adapted for the expression of multi-subunit protein complexes[845]. Each subunit was cloned upstream of an IRES2 sequence, which facilitated the expression of a second open reading frame (ORF) encoding one of four fluorescent markers (Figure 5.14a). Additionally, the system incorporated puromycin resistance coupled to a Tet-On element, enabling selection of efficiently transduced cells and inducible control over SMX complex expression. This framework was employed to construct plasmids for human SMX subunits, which were codon-optimised, synthesised, and assembled into various pTLCV2 plasmids (Figure 5.14a, see 2.2.7.3). Backbone constructs were generously provided by Professor Yin Dong (University of Oxford), with pTLCV2_mOrange2 generated in this study (Section 2.2.7.3). Due to concerns arising from mouse SMX preparations (Section 5.2.1), where splitting SLX4 and SLX1 adversely affected SLX1 expression, these genes were co-expressed on a single plasmid. Specifically, a construct was designed with N-terminally SF-tagged SLX4 and C-terminally EGFP-tagged SLX1, separated by an IRES2 sequence, referred to as 'pTLCV2_SS-EGFP'. Similarly, MUS81 and HA-Halo-EME1 were separated by a P2A sequence and co-expressed with mCherry downstream of an IRES2 ('pTLCV2_ME-mCherry2'). XPF and ERCC1-Avi-His were cloned with an internal P2A sequence, upstream of an IRES2 containing a BirA-TagBFP2 fusion, in 'pTLCV2_XE-BFP'. Additionally, SLX4IP tagged with SUMO was cloned upstream of an IRES2-separated mOrange2 ('pTLCV2_SLX4IP-mOrange2'). A GFP-tagged SLX4 construct was included as a gift from Dr. Kaima Tsukada. Given the inherent toxicity observed with overexpression of SMX subcomplexes (Figure 5.12a, b), the inducible system was expected to provide a more controlled expression environment,

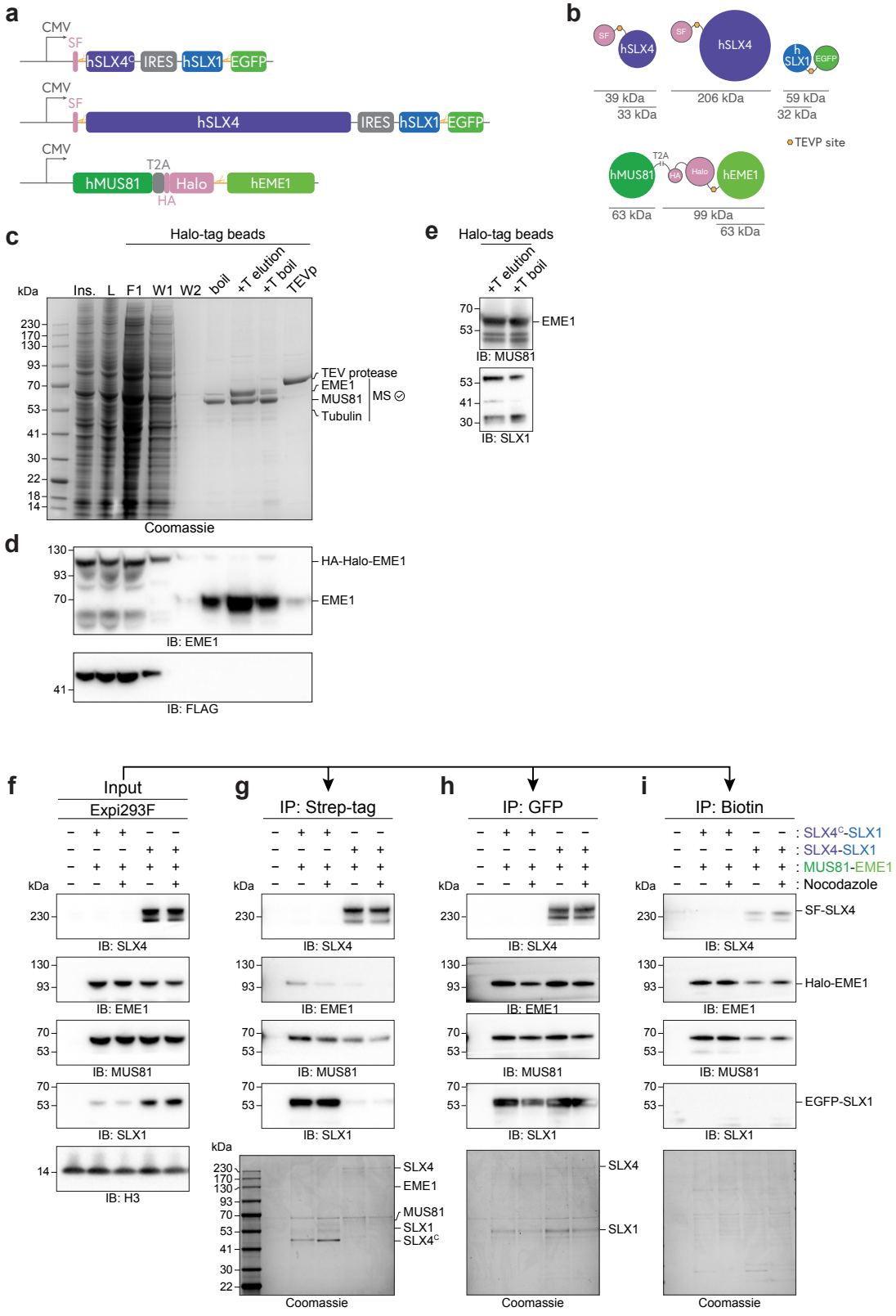

a
CMV
SF
hSLX4C
IRES
hSLX1
EGFP
hSLX4
hMUS81
T2A
Halo
HA
hEME1
b
39 kDa
33 kDa
206 kDa
59 kDa
32 kDa
TEVP site
63 kDa
99 kDa
c
Halo-tag beads
Ins. L F1 W1 W2 boil +T elution +T boil TEVp
TEV protease
EME1
MUS81
Tubulin
MS
Coomassie
d
HA-Halo-EME1
EME1
IB: EME1
IB: FLAG
e
+T elution
+T boil
IB: MUS81
IB: SLX1
f
Input
Expi293F
IB: SLX4
IB: EME1
IB: MUS81
IB: SLX1
IB: H3
g
IP: Strep-tag
SLX4
EME1
MUS81
SLX1
SLX4C
h
IP: GFP
i
IP: Biotin
: SLX4C-SLX1
: SLX4-SLX1
: MUS81-EME1
: Nocodazole
SF-SLX4
Halo-EME1
EGFP-SLX1

Figure 5.13 | Expression and purification of hSLX4^{C}-SLX1, hMUS81-EME1 complexes from Expi293F cells. a, Schematic representation of hSLX4^{C}-SLX1 and hMUS81-EME1 constructs used for mammalian expression under CMV promoters. Constructs include SF-hSLX4^{C} co-expressed with hSLX1-EGFP (IRES), and hMUS81 co-expressed with HA-Halo-hEME1 (*via* T2A). **b,** Molecular weights of individual components and tags. SF - TwinStrep-FLAG; Halo - Halo-tag; TEVP - TEV protease cleavage site. **c,** SDS-PAGE of Halo-tag purification fractions for hMUS81-EME1. Fractions include insoluble lysate (Ins.), flow-through (F1), washes (W1-W2), and elution fractions without (boil) or with TEV protease cleavage (+T elution) or boiled Halo beads post-TEV protease (+T boil). TEV protease at the concentration used for elution is loaded as a control (TEVp). **d,** Western blot analysis of hMUS81-EME1 purification fractions probing for EME1 (HA-Halo-EME1) and FLAG-tag. **e,** Western blot analysis of eluted Halo fractions showing EME1, MUS81, and SLX1 subunits. **f-i,** Immunoprecipitation (IP) experiments in Expi293F cells. **f,** Western blot analysis of input lysates for SLX4, EME1, MUS81, SLX1, and H3. **g,** Strep-tag IP showing the presence of SLX4, MUS81-EME1, and SLX1. **h,** GFP IP confirming SLX4-SLX1 and MUS81-EME1 interactions. **i,** Biotin IP demonstrating interactions among SF-SLX4, MUS81-EME1, and EGFP-SLX1. SDS-PAGE gels below each blot confirm the presence of protein complexes.

yielding a higher number of viable and expressing cells, albeit with likely reduced SMX expression. The process for creating a stable, inducible cell line is outlined in Figure 5.14b. As proof of principle, Expi293F cells inducibly expressing SLX4IP were established. Suspension Expi293F cells were adapted to adherent growth in Freestyle 293F medium (based on personal communication with Prof. Yin Dong and our own medium tests, data not shown) and transduced with lentivirus carrying 'pTLCV2_SLX4IP-mOrange2' (Figure 5.14b). Transductants were selected using increasing concentrations of puromycin, and cells expressing SLX4IP-mOrange2 were enriched by fluorescence-activated cell sorting (FACS) following doxycycline (Dox) induction (Figure 5.14c). Approximately 1% of transductants displayed leaky, non-Dox-induced expression of mOrange2, which increased to $\sim$15% upon Dox treatment (Figure 5.14c). This leaky expression may result from trace levels of tetracycline contamination in fetal bovine serum (FBS) used in the adherent culture medium, however this issue would be resolved when cells are switched to suspension culture in medium that does not require FBS. A sorted population of Dox-responsive cells was expanded and frozen for future experiments. Expression of SLX4IP in response to Dox was evaluated for this heterogeneous population. Cells were treated with Dox for 24 hours, and mOrange2 fluorescence was used to quickly assess expression quality by visualising a fraction of the harvested pellet (Figure 5.14d). Western blotting further confirmed Dox-induced SLX4IP expression in these cells (Figure 5.14e). The successful inducible expression of SLX4IP provides a foundation for extending this approach to the full SMX complex.

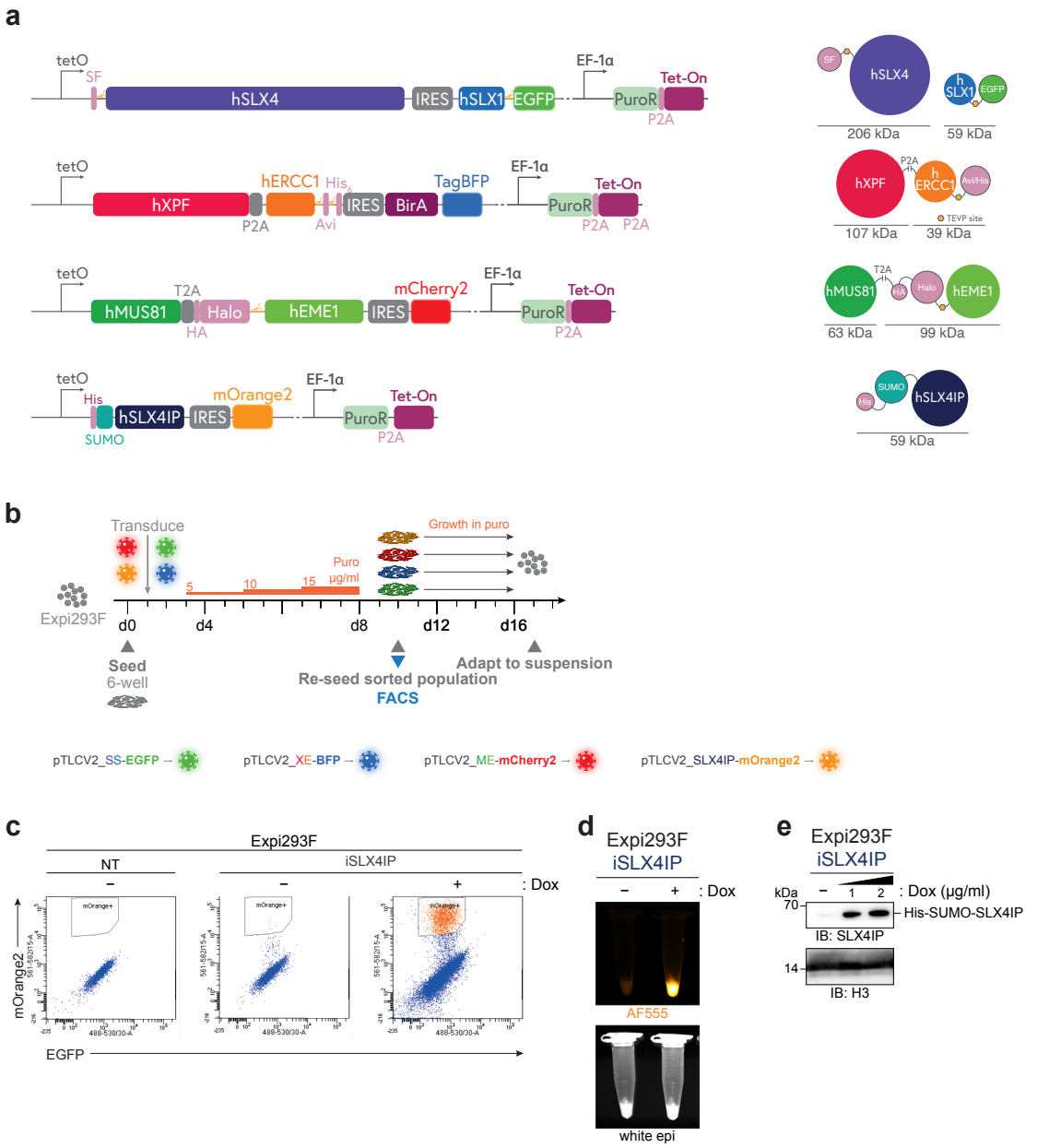


Figure 5.14 | Design of inducible hSLX4IP and hSMX constructs for generation of inducibly-expression cell lines. a, Schematic representation of hSLX4IP and hSMX constructs under Tet-On expression systems for mammalian expression. Components include SF-hSLX4 co-expressed with hSLX1-EGFP (IRES), hXPF-hERCC1 (co-expressed *via* P2A and Avi/His_{10} tags), Halo-hMUS81 co-expressed with hEME1-mCherry2 (*via* T2A), and His-SUMO-hSLX4IP-mOrange2 (IRES). Tet-On systems allow doxycycline-inducible expression, with puromycin resistance (PuroR) for selection. **b,** Workflow for generating stable Expi293F cell lines. Cells were seeded in six-well plates as adherent cells, transduced with lentiviral constructs, and subjected to increasing puromycin selection (5–15 μg/mL) over 12 days. Fluorescence-sorted populations were expanded and adapted to suspension culture. Constructs for transduction are colour-coded as indicated. **c,** Fluorescence-activated cell sorting (FACS) of Expi293F cells expressing Dox-inducible His-SUMO-hSLX4IP and mOrange2. Representative density plots show expression of mOrange2 in transduced cells with Dox induction. **d,** Fluorescence imaging of Expi293F cells expressing His-SUMO-hSLX4IP-mOrange2, visualised using the Alexa Fluor 555 (AF555) filter for mOrange2 and under white epi-illumination. **e,** Western blot analysis confirming doxycycline-inducible expression of His-SUMO-hSLX4IP in Expi293F cells, probed with anti-His and anti-H3 antibodies.

To establish an inducible expression system for SLX4, a pTLCV2 plasmid encoding EGFP-SLX4 under a Tet-ON-responsive promoter was utilised (Figure 5.15a, b). Initial transduction followed by puromycin selection and sorting for EGFP expression revealed some leaky expression ($\sim$1%), which increased to 13% upon Dox treatment (Figure 5.15b). After enrichment, the sorted population was grown, frozen, and subsequently induced with Dox for 24 hours to assess expression. Fluorescence imaging of harvested cell pellets confirmed the presence of EGFP-SLX4 (Figure 5.15c). To determine the effects of SLX4 overexpression, cell viability assays were performed. Expi293F cells inducibly overexpressing SLX4 exhibited reduced viability in the presence of Dox compared to cells inducibly expressing SLX4IP or GFP alone, although this pronounced sensitivity was observed in long-term treatments (Figure 5.15d). This supported the development of this inducible system, where Dox concentrations could be optimised to balance expression levels with cell viability. Additionally, shorter transient expression periods were suggested as a strategy to mitigate toxicity. Using Dox concentrations of 1 µg/ml and 1 µg/ml, GFP-SLX4 was successfully overexpressed after overnight treatment, though showing no pronounced effect of Dox-doseage (Figure 5.15e). While minor induction of SLX1 and XPF was observed, no significant changes in the expression of other SMX components were detected (Figure 5.15e). GFP immunoprecipitation under standard conditions successfully enriched SMX components, including SLX4, SLX1, XPF, and SLX4IP, but not EME1 (Figure 5.15f). This lack of EME1 detection may be attributed to the limitations of the antibody used, which is overcome when specifically overexpressing MUS81-EME1 (Figure 5.12e). To optimise expression conditions, cells were treated with nocodazole to induce G2/M arrest and/or SJG-136 to induce crosslinking damage, hypothesised to stabilise the SMX complex. However, neither treatment significantly altered complex assembly or yield (Figure 5.15f). A portion of the GFP immunoprecipitate was incubated with a fork substrate to test for nuclease activity, but no activity was observed (Figure 5.15g). Future efforts using larger culture volumes, akin to methods employed for purifying human adult muscle-type nicotinic receptors[845], may improve the yield and quality of SMX components by this system.

Following the successful expression and immunoprecipitation of SLX4, the next objective was to establish cell lines expressing combinations of SMX components. Initially, a sequential transduction and sorting strategy was considered, but to expedite the

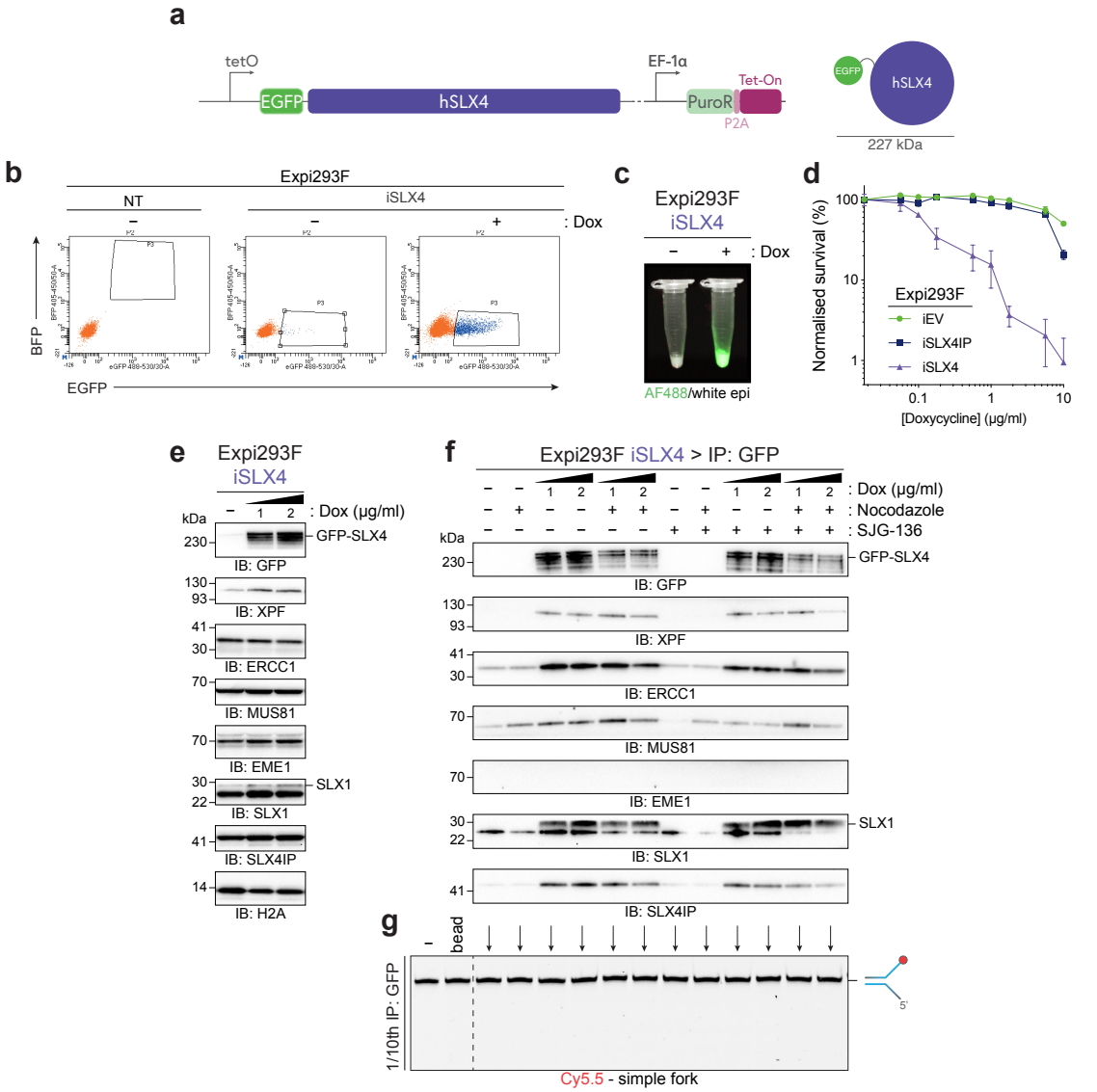


Figure 5.15 | Inducible expression and functional analysis of GFP-hSLX4 in Expi293F cells. **a,** Schematic representation of the GFP-hSLX4 inducible construct. GFP-hSLX4 is expressed under the control of a doxycycline-inducible tetO promoter with EF-1α driving constitutive expression of puromycin resistance (PuroR). The predicted molecular weight of GFP-hSLX4 is 227 kDa. **b,** Fluorescence-activated cell sorting (FACS) of Expi293F cells transduced with the GFP-hSLX4 construct, with or without doxycycline (+/- Dox). GFP fluorescence indicates successful induction of GFP-hSLX4 expression. NT: non-transduced control. **c,** Fluorescent imaging of cell pellets under white epi-illumination and Alexa Fluor 488 (AF488) channel, showing GFP fluorescence in doxycycline-treated cells. **d,** Dose-response survival analysis of Expi293F cells Dox-inducibly expressing GFP-hSLX4, hSLX4IP, or an empty vector (iEV). Cells were treated with increasing concentrations of doxycycline for 6 days after treatment, and survival was normalised to untreated controls. Error bars represent the mean $\pm$ SD of n = 3 biological repeats, each of 3 technical repeats. **e,** Western blot analysis of GFP-hSLX4 expression and its associated proteins in Expi293F cells treated with 1 or 2 μg/mL doxycycline. **f,** Immunoprecipitation (IP) of GFP-hSLX4 in Expi293F cells treated with 1 or 2 μg/mL doxycycline, with or without nocodazole or SJG-136. Western blotting confirms the interaction of GFP-hSLX4 with XPF, ERCC1, MUS81, EME1, SLX1, and SLX4IP. **g,** Endonuclease cleavage of a Cy5.5-labelled simple fork was examined *via* a gel-based assay. Substrate processing was assessed with the incubation of 1/10th of GFP immunoprecipitates from **f** with 10 nM DNA substrate for 1 h at 37 °C. Reaction products were then resolved on a 20% polyacrylamide denaturing gel, and fluorescence visualised.

process, combinations of lentiviruses were co-transduced into Expi293F cells. Cell lines expressing combinations of particular SMX components of interest were generated: 'iXE' (XPF-ERCC1), 'iSS-ME' (SLX4-SLX1 with MUS81-EME1), and 'iSMX' (entire SMX complex, including XPF-ERCC1, SLX4-SLX1, and MUS81-EME1). Immunofluorescence microscopy confirmed Dox-inducible expression of BFP (iXE), EGFP and mCherry2 (iSS-ME), and BFP, EGFP, and mCherry2 (iSMX) (Figure 5.16a). FACS analysis demonstrated 44.6% of live cells expressing BFP in the iXE population (Figure 5.16b). Although a similar fraction of cells appeared to be BFP+ in iSS-ME-transduced cells, which are not expected to express BFP, this was attributed to overlapping spectral signals between EGFP and BFP. In iSS-ME-transduced cells, EGFP+/mCherry2+ cells accounted for 0.23% of live cells, forming the required population for iSS-ME (Figure 5.16b, upper layer panel 2). For iSMX, triple-positive (BFP+/EGFP+/mCherry2+) cells comprised 2.7% of cells that were mCherry2+/EGFP+ (0.006% of all live cells) (Figure 5.16b, middle panel).

The establishment of these inducible cell lines is at a preliminary stage, with frozen stocks prepared for future experiments. Confirmation of subunit expression *via* Western blotting remains pending, but the inducible system offers promising potential for large-scale SMX complex expression in Expi293F cells. Induced expression in larger culture volumes represents a feasible next step in advancing this methodology.

5.4 | Exploring regulation of SLX4 sub-complexes

5.4.1 | Phosphorylation of SLX4 by CDK1-cyclin B

SMX assembly is tightly regulated throughout the cell cycle, with assembly occurring transiently during G2 to early mitosis before disassembling in anaphase. This regulation is mediated by cell-cycle-dependent post-translational modifications rather than variations in subunit expression[571]. In yeast, Mus81-Mms4^{EME1} activity peaks during G2/M, driven by phosphorylation by Cdc28$^{CDK1/2}$ and Cdc5^{PLK1}, which enhances Holliday junction (HJ) resolution[572,573]. Similarly, Eme1 phosphorylation by Cdc2^{CDK1} and Rad3^{ATR} is DNA damage-dependent, providing tight regulation of nuclease activity[575]. In human cells, phosphorylation of EME1 by CDK1 and PLK1 increases resolvase activity and promotes interaction with SLX4[554,565], although its precise role in HJ resolution remains unclear.

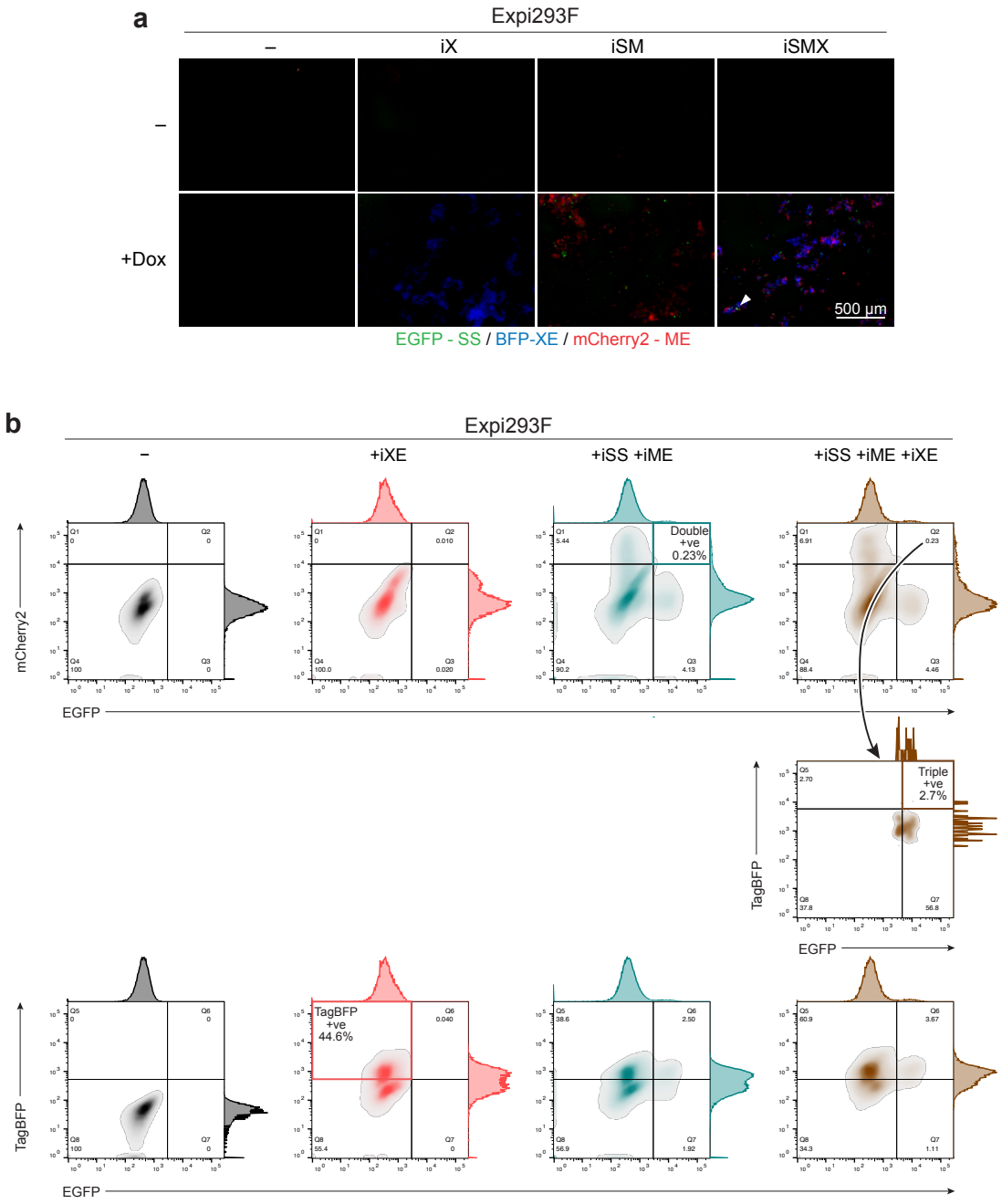


Figure 5.16 | Expi293F cell lines for inducible expression of hSMX components. a, Fluorescence imaging of Expi293F cells transduced with inducible constructs for XPF-ERCC1 (iX), SLX4-SLX1-MUS81-EME1 (iSM), or complete (iSMX) hSMX complexes. Fluorescent tags include EGFP for SS (hSLX4/hSLX1), BFP for XE (hXPF/hERCC1), and mCherry2 for ME (hMUS81/hEME1). Cells were treated with or without 2 µg/ml doxycycline (±Dox) to induce expression. Scale bar: 500 µm. **b,** Flow cytometry analysis of Expi293F cells transduced with inducible hSMX components. Top panels show mCherry2 *versus* EGFP fluorescence, and bottom panels show TagBFP *versus* EGFP fluorescence, for cells expressing combinations of iXE, iSS + iME, and iSS + iME + iXE. The right-most panel highlights the detection of triple-positive cells (EGFP, TagBFP, and mCherry2) indicative of the full iSMX complex; the mCherry2+/EGFP+ double-positive cells in iSS + iME + iXE were gated into the population seen in the middle right panel, denoted by the arrow.

CDK1 phosphorylation of SLX4's SAP domain (T1544, T1561, T1571) significantly enhances MUS81 interaction by increasing binding affinity ten-fold[577,578], with phosphomimetic mutants further supporting this effect[578]. Additionally, MUS81 is phosphorylated at S87 by CK2 during mitosis, promoting its interaction with SLX4[579]. Given previous successful purifications of SLX4-C (comprising the SLX1- and MUS81-interacting C-terminal region of SLX4 along with SLX1, MUS81, and EME1, Figure 5.8f), it is feasible to investigate whether similar phosphorylation events occur in larger SLX4 peptides that include SLX1.

To evaluate phosphorylation changes in Sf9 insect cells overexpressing SLX4-C, experiments were conducted under conditions that aimed to synchronise cells in G2/M. While Sf9 cells have been reported to arrest in G2/M following nocodazole treatment, significantly higher concentrations are required compared to human cells[846]. Partial G2/M synchronisation was achieved using 10 µg/ml nocodazole—200-fold greater than the standard dose for human cells—although robust arrest was not observed (Figure 5.17a, b). Sf9 cells were subsequently transduced with a baculovirus encoding SLX4-C (Figure 5.17c), with nocodazole added during the final 24 hours of expression. Phosphorylation of histone H3 at serine 10 (H3 pS10), a marker of mitotic chromosome condensation and transcriptional activation, increased at the highest nocodazole dose in uninfected cells (Figure 5.17d). Notably, H3 pS10 levels were markedly elevated in all conditions for baculovirus-infected Sf9 cells, possibly reflecting transcriptional activation driven by baculovirus expression. Despite indications that partial G2/M arrest may be achievable in Sf9 cells, no significant changes were observed in SLX4-C subunit expression, post-translational modification (PTM) status, or pulldown yields (Figure 5.17d). Purified SLX4-C was subsequently incubated with recombinant CDK1–cyclin B to assess phosphorylation changes. CDK1–cyclin B treatment induced a mobility shift in SLX4 on SDS-PAGE, indicative of phosphorylation (Figure 5.17e). This mobility shift was reversed by λ-phosphatase (λ-PPase) treatment, with the SLX4 band migrating beyond its untreated position, suggesting partial phosphorylation of SLX4 in insect cells prior to purification (Figure 5.17e, f). This observation may reflect a subset of Sf9 cells in mitosis harboring active CDK1. EME1 displayed a similar mobility shift pattern, whereas SLX1 and MUS81 did not (Figure 5.17e, f).

Mass spectrometry confirmed that CDK1–cyclin B phosphorylated SLX4 at T1561 and T1571 but did not identify phosphorylation at T1544 (Figure 5.17g), a site previously reported[577]. Although SDS-PAGE analysis revealed a mobility shift in untreated SLX4-C compared with SLX4-C treated with λ-PPase, no phosphorylated SLX4 peptides were detected in this state (Figure 5.17g, upper right panel). Multiple phosphorylation sites were identified on EME1, with minor changes observed following CDK1–cyclin B treatment, despite a notable band shift in SDS-PAGE for treated EME1 (Figure 5.17e, g). Future experiments will explore whether these phosphorylation events influence the nuclease activity of SLX4-C and their potential impact on SLX1.

5.4.2 | Exploring binding partners of SLX4 reveals an interaction with FANCJ

SLX4 is implicated in a broad spectrum of biological processes, largely mediated by interactions with its canonical binding partners. While these interactions are well-validated biochemically, their specific roles in defined biological contexts remain unclear. Telomeric accumulation of polyubiquitylated PCNA recruits SLX4, a means to trigger BIR at telomeres through nuclease incision[847]. If a direct PCNA interaction stabilises SLX4 at chromatin, then it follows that the inclusion of PCNA in SMX preparations may stabilise the complex, making it more structurally tractable. A recent study suggested that SLX4 is recruited to stalled replication forks by recognising RAD18-mediated monoubiquitylation of PCNA, a process negatively regulated by ATR phosphorylation of RAD18[605]. This ATR-dependent phosphorylation is thought to act as a gatekeeping mechanism, restricting SLX4 accumulation at stalled forks and thus preventing fork collapse driven by SLX1 nuclease activity[605]. Furthermore, the study proposed that this regulatory mechanism also limits SLX4 recruitment to telomeres in ALT-positive cells, safeguarding these structures from nuclease-induced instability[605].

Putative PCNA-interacting peptide (PIP box) motifs within SLX4 were identified in the study by searching through the SLX4 sequence. However, these sequences deviate from the canonical PIP box consensus (Qxxψxx(FY)(FY), ψ denoting a hydrophobic amino acid), with SLX4's putative PIP box sequences including ^{54}LCASFFQRVKK, ^{204}RTAQLVLQRMQ, and ^{267}AVALTLQQEFA. Structurally, this raised concerns, since

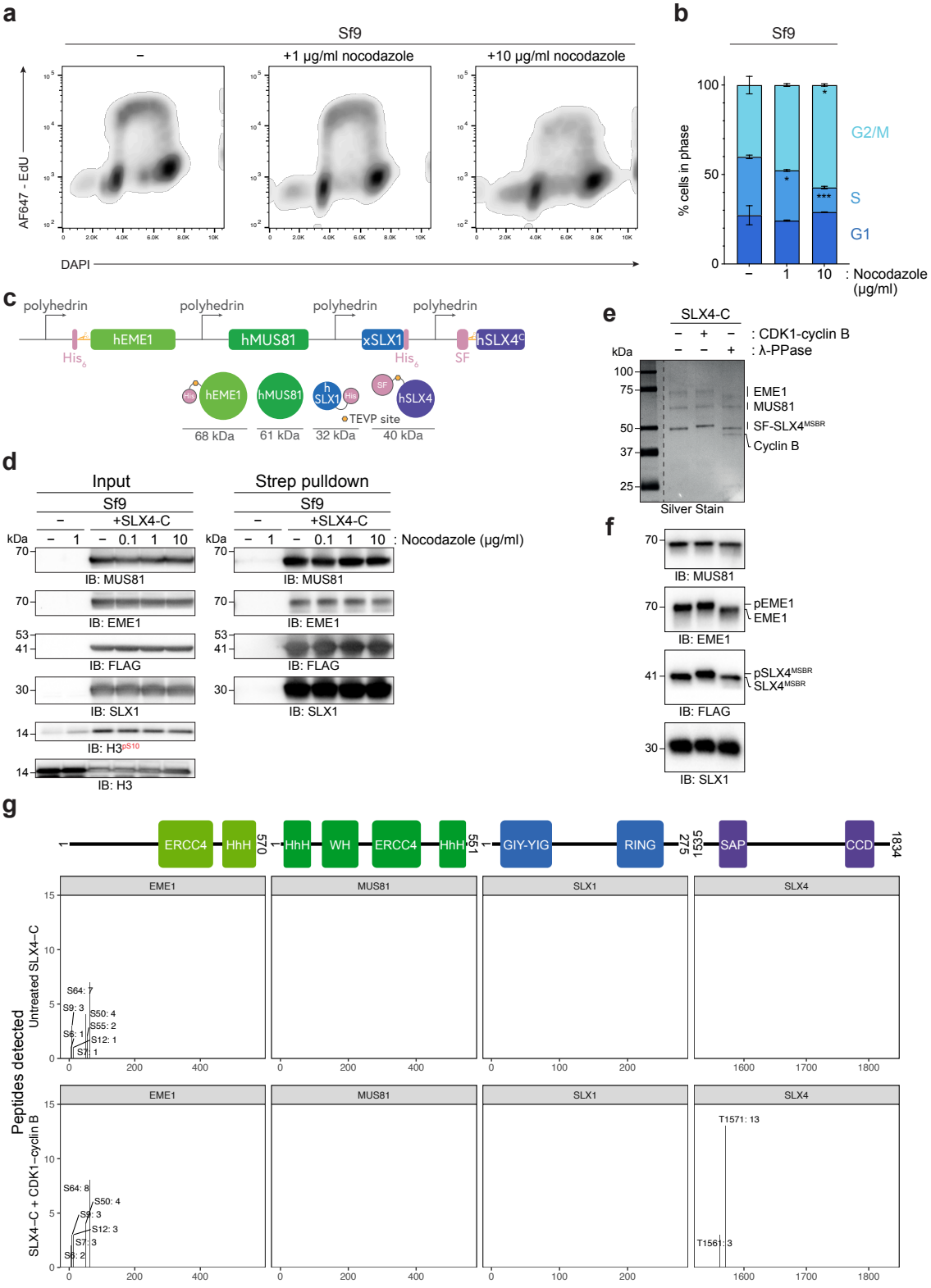

a
Sf9
–
+1 µg/ml nocodazole
+10 µg/ml nocodazole
AF647 - EdU
DAPI
b
Sf9
% cells in phase
G2/M
S
G1
: Nocodazole (µg/ml)
c
polyhedrin
hEME1
hMUS81
xSLX1
hSLX4C
His6
SF
TEVP site
68 kDa
61 kDa
32 kDa
40 kDa
d
Input
Strep pulldown
Sf9
+SLX4-C
: Nocodazole (µg/ml)
IB: MUS81
IB: EME1
IB: FLAG
IB: SLX1
IB: H3pS10
IB: H3
e
SLX4-C
: CDK1-cyclin B
: λ-PPase
Silver Stain
EME1
MUS81
SF-SLX4MSBR
Cyclin B
f
IB: MUS81
pEME1
EME1
IB: EME1
pSLX4MSBR
SLX4MSBR
IB: FLAG
IB: SLX1
g
ERCC4
HhH
WH
GIY-YIG
RING
SAP
CCD
EME1
MUS81
SLX1
SLX4
Peptides detected
Untreated SLX4–C
SLX4–C + CDK1–cyclin B

Figure 5.17 | Phosphorylation analysis of SLX4-C following nocodazole treatment or CDK1-cyclin B activation. a, Cell cycle analysis of Sf9 cells treated with 0, 1, or 10 µg/mL nocodazole for 16 h. Cells were stained with DAPI for DNA content and Alexa Fluor 647 (AF647)-EdU for S-phase detection. Representative density plots of AF647-EdU *versus* DAPI signal are shown. **b,** Quantification of the percentage of cells in G1, S, and G2/M phases based on flow cytometry profiles shown in **a**. Data represent mean ± SD of n = 3 biological repeats. **c,** Schematic representation of the SLX4-C construct expressed in Sf9 cells, including hEME1, hMUS81, xSLX1, and hSLX4^{C}. Molecular weights of individual components are shown in kilodaltons (kDa). **d,** Western blot analysis of Sf9 cell lysates (Input) and Strep-Tactin pull-downs of SLX4-C following treatment with increasing concentrations of nocodazole (0–10 µg/mL). Probing was performed for MUS81, EME1, FLAG (hSLX4^{C}), SLX1, and H3.3^{pS10} is present as a marker of mitotic progression. **e,** Silver stain analysis of SLX4-C complex treated with CDK1-cyclin B kinase and λ-phosphatase for 1 h to assess phosphorylation status. **f,** Western blot analysis of SLX4-C phosphorylation by band shift, using anti-MUS81, anti-EME1, anti-FLAG (for SLX4), and anti-SLX1 antibodies. **g,** Mass spectrometry analysis of phosphorylated peptides identified in EME1, MUS81, SLX1, and SLX4^{C} before (top) and after (bottom) CDK1-cyclin B treatment. Detected phosphorylation sites are mapped to their respective protein domains.
Note: CDK1-cyclin B/λ-PPase treatment and mass spectrometry was performed by Dr. Abimael Cruz-Migoni.

the folds of these sub-optimal PIP boxes are unlikely to resemble other PIP box motifs, prompting an analysis of AlphaFold models for these SLX4 peptides in complex with PCNA. This was performed alongside a well-characterised PIP box-containing peptide from p21/CDKN1A as a reference[848]. The peptides ^{54}LCASFFQRVKK and ^{204}RTAQLVLQRMQ from SLX4 displayed markedly different folds around PCNA compared to the p21 PIP box, while the ^{267}AVALTLQQEFA peptide exhibited a moderate degree of structural overlap with p21 (Figure 5.18a, b). However, the confidence scores for the positioning of all three SLX4 peptides were insufficient for AlphaFold to identify a reliable interaction using the AlphaBridge tool (Figure 5.18c). In contrast, the p21 peptide formed a clear and expected interaction interface with PCNA as predicted by AlphaFold/AlphaBridge (Figure 5.18c). Additionally, the study's reliance on 1% formaldehyde crosslinking prior to cell lysis in co-immunoprecipitation experiments raises concerns about artificial stabilisation of the interactions. To further investigate the potential interaction between SLX4 and PCNA, co-immunoprecipitation experiments were performed using GFP-tagged SLX4. Only a very minor fraction of PCNA could be detected in the GFP-SLX4 immunoprecipitate, requiring extended exposure for visualisation, regardless of whether PCNA was ubiquitylated or unmodified (Figure 5.18d). In a separate experiment, however, no PCNA was observed in the GFP-SLX4 immunoprecipitate using a shorter exposure time. This inconsistency between the two co-immunoprecipitation experiments may suggest that this interaction is weak or difficult to capture. In contrast, SNM1A

overexpression robustly induced PCNA monoubiquitylation, as expected[849], and SNM1A successfully pulled down both ubiquitylated and unmodified forms of PCNA, confirming the interaction between SNM1A and PCNA and serving as a positive control, showing that PCNA could be pulled down under these experimental conditions (Figure 5.18e). These findings raise significant doubts about a direct interaction between SLX4 and PCNA under the tested conditions[605]. Further studies are needed to determine whether the proposed interaction occurs under physiological conditions or whether additional factors or experimental conditions are required to facilitate this binding.

Direct physical or functional interactions between SLX4 and FANCJ remain largely unexplored, despite both proteins being implicated in the Fanconi anaemia (FA) pathway. FA complementation group J can be caused by biallelic mutations in FANCJ, but the stage at which FANCJ acts within the FA pathway has been suggested to involve its interaction with BRCA1, and hence is connected to HR[851,852]. However, its distinct function in HR is loosely defined, and its resolution of G4 DNA structures through its helicase activity and stabilisation of microsatellites appear as HR-independent responsibilities. DNA ICLs are processed in a manner that depends on FANCJ's helicase activity and interaction with MLH1[853]. The FANCJ interactome also includes RPA[854], BLM[855], TOPBP1[856] and CtIP[857]. Furthermore, FANCJ is actively required for SPRTN-dependent DPC proteolysis, promoting the replicative bypass of protein barriers either covalently or non-covalently attached to DNA[858].

The nature of their connection within this pathway remains unclear. Recent advances in machine learning and structural prediction, such as AlphaFold, have been employed to predict binary interactions within the DNA damage response (DDR)[775]. Among SLX4's established interactors, including XPF, SLX1, MSH2, and RTEL1, FANCJ has emerged as a notable candidate (Figure 5.19a). The predicted SLX4-FANCJ interaction features an extensive buried interface, indicating a potentially significant functional relationship (Figure 5.19b, c). High-throughput proximity biotinylation studies further corroborate this interaction, identifying FANCJ as a key component of the SLX4 interactome[859]. Additionally, analysis of the CRISPR DepMap database[750] highlights FANCJ as the third-highest co-dependent gene with SLX4, following XPF and FANCI (Figure 5.19e). In reverse, SLX4 ranks as the 14th-most co-dependent gene for FANCJ. These co-dependence

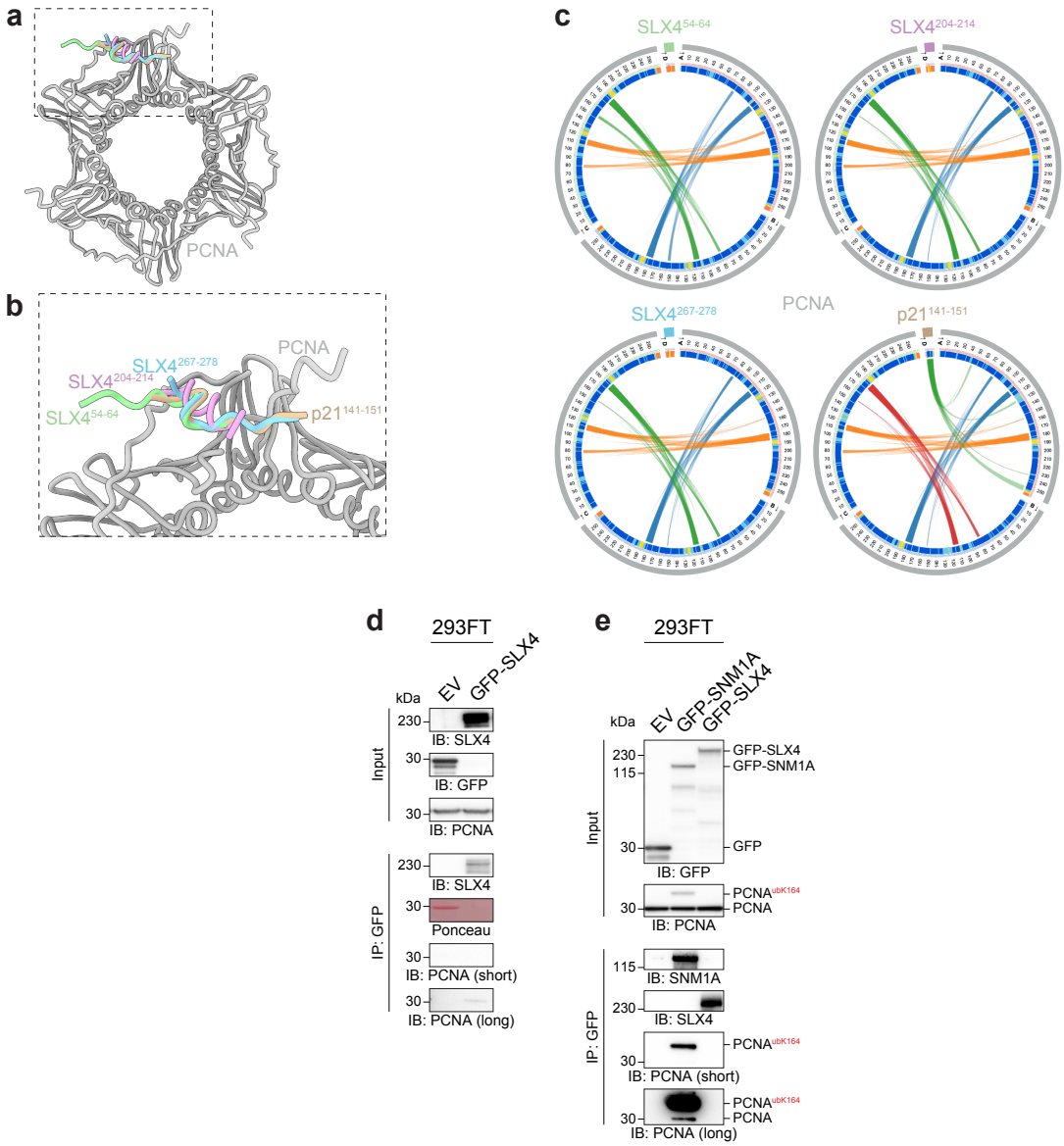


Figure 5.18 | SLX4-PCNA interactions revealed through structural modeling and co-immunoprecipitation. a, Structural model of the PCNA homotrimer, highlighting the binding regions of SLX4 (green, magenta, and blue) and p21 (orange) with PCNA. The boxed region indicates the detailed view shown in **b**. **b,** Close-up view of the SLX4 and p21 binding motifs interacting with PCNA. SLX4$^{54-64}$, SLX4$^{204-214}$, SLX4$^{267-278}$, and p21$^{141-151}$ are color-coded, highlighting their binding positions relative to the PCNA interface. **c,** Circos plots showing the predicted interaction networks of SLX4$^{54-64}$ (green), SLX4$^{204-214}$ (magenta), SLX4$^{267-278}$ (blue), and p21$^{141-151}$ (orange) with PCNA by AlphaBridge tool[850]. Lines indicate the interaction strength and specific contact residues. **d,** Co-immunoprecipitation (IP) of GFP-tagged SLX4 in 293FT cells demonstrates binding to endogenous PCNA. **e,** Co-IP of GFP-tagged SLX4 or SNM1A with PCNA in 293FT cells. The presence of acetylated PCNA (PCNA^{AcK164}) is highlighted in the IP fractions.

metrics, derived from Pearson correlations of knockout effects across diverse cell lines, suggest a degree of epistatic interaction between the two genes, warranting further investigation into their biological interplay.

Immunoprecipitation experiments using GFP-tagged SLX4 were conducted to investigate its interaction with FANCJ, with GFP-BRCA1 serving as a positive control[860]. GFP-SLX4 successfully pulled down several of its known interactors. However, MUS81 and SLX4IP were absent under all immunoprecipitation conditions (Figure 5.20b), despite being well-documented binding partners of SLX4[577,668]. Conversely, both MSH2 and TRF2 were consistently identified in GFP-SLX4 immunoprecipitates (Figure 5.20b). The absence of MUS81 and SLX4IP in these experiments is likely attributable to the high-salt wash conditions used in the protocol, which disrupt transient interactions. In contrast, the SLX4-FANCJ interaction remained stable under these stringent conditions (Figure 5.20c). Notably, SLX4 co-precipitated a higher-molecular-weight form of FANCJ, potentially indicative of a post-translationally modified variant not enriched by BRCA1 (Figure 5.20c). To assess whether the SLX4-FANCJ interaction was mediated by BRCA1, immunoprecipitations were repeated in 293FT cells following BRCA1 knockdown. BRCA1 was effectively depleted using siRNA (Figure 5.20d). Interestingly, BRCA1 depletion diminished the SLX4-dependent increase in SLX1 expression (Figure 5.20d). This relationship may reflect a mutual stabilisation of BRCA1 and SLX4 complexes, potentially mediated by the bridging function of BARD1[610]. However, FANCJ continued to co-purify with GFP-SLX4, albeit to a lesser extent, indicating that the SLX4-FANCJ interaction occurs independently of BRCA1.

The predicted interaction between SLX4 and FANCJ is suggested to involve a specific β-sheet within the BTB domain of SLX4 and an antiparallel β-sheet in FANCJ (Figure 5.19d). To map the interaction site within SLX4, truncation mutants of GFP-tagged SLX4 were expressed and analysed through co-immunoprecipitation experiments (Figure 5.21a). The expression of these GFP-SLX4 truncations was uneven across experiments, complicating the interpretation of interaction losses with FANCJ (Figure 5.21b). Nonetheless, the data provided insights into how different regions of SLX4 influence cellular tolerance to its overexpression. The epitope of the SLX4 antibody used in this study was mapped to the C-terminal region (residues 1489-1834) of SLX4, but outside the

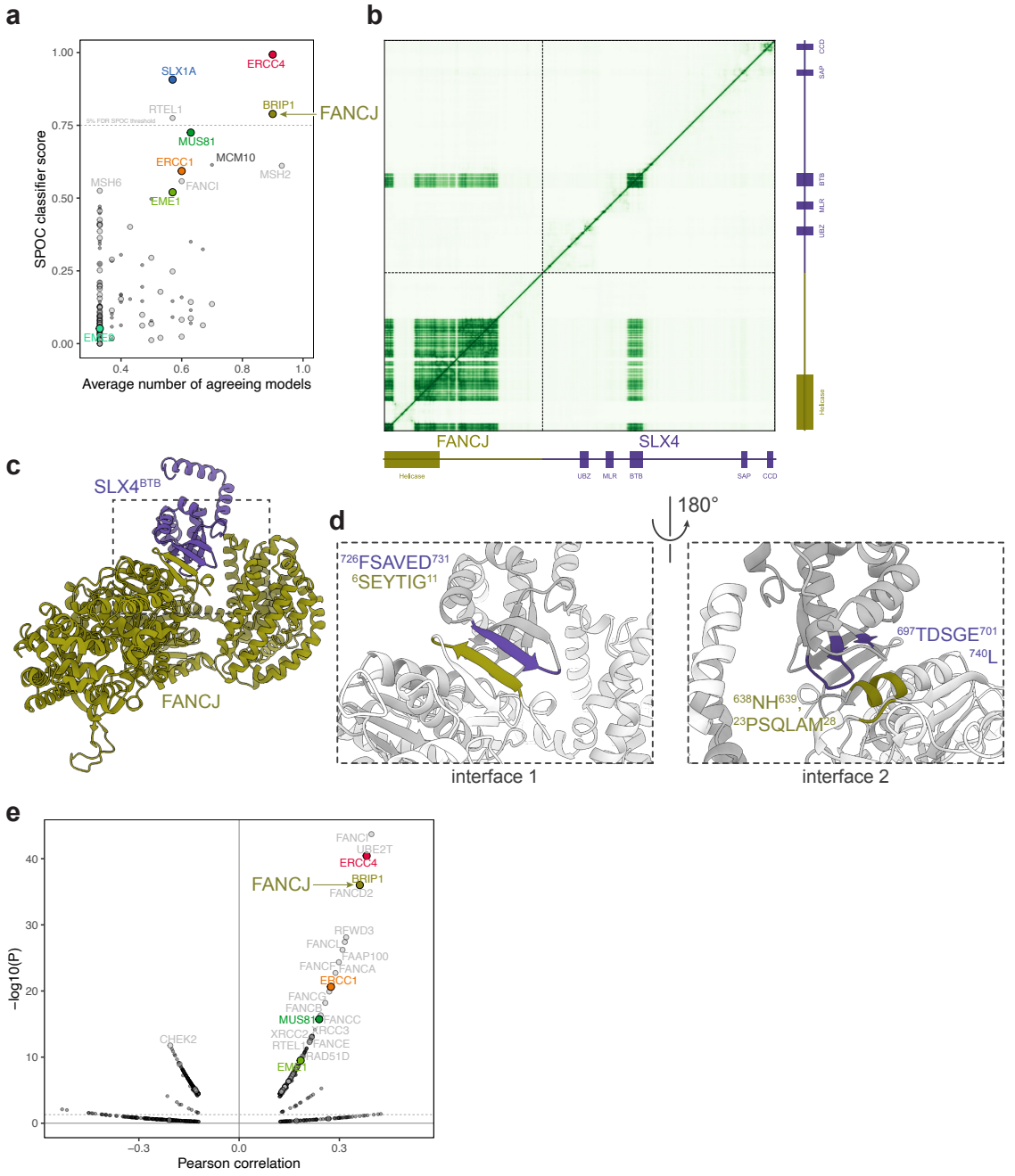


Figure 5.19 | Predicted interaction and functional dependency between SLX4 and FANCJ. **a,** SPOC classifier analysis[775] ranking SLX4-interacting proteins based on predicted agreement scores and the average number of structural models. FANCJ and MUS81 are highlighted as top interactors, alongside ERCC4 (XPF) and EME1. **b,** Contact map of predicted interactions between SLX4 and FANCJ generated by structural modelling *via* AlphaFold, with domains annotated. Key regions of interest include SLX4^{BTB} (purple) with FANCJ (gold). **c,** Structural model of the SLX4-FANCJ complex. SLX4^{BTB} is shown in purple, and FANCJ in gold. The boxed regions highlight interfaces 1 and 2, further detailed in **d**. **d,** Close-up views of interface 1 (left) and interface 2 (right). Interface 1 involves interactions between SLX4 residues ^{726}FSAVED^{731} and FANCJ residues ^{6}SEYTIG^{11}. Interface 2 highlights contacts between SLX4 residues ^{697}TDSGE^{701} and FANCJ residues ^{638}NH^{639} and ^{23}PSQLAM^{28}. **e,** Volcano plot mapping the Pearson correlation coefficient (x-axis) with statistical significance (-log$_{10}$(P)) for functional co-dependencies of SLX4 in the CRISPR DepMap[750].

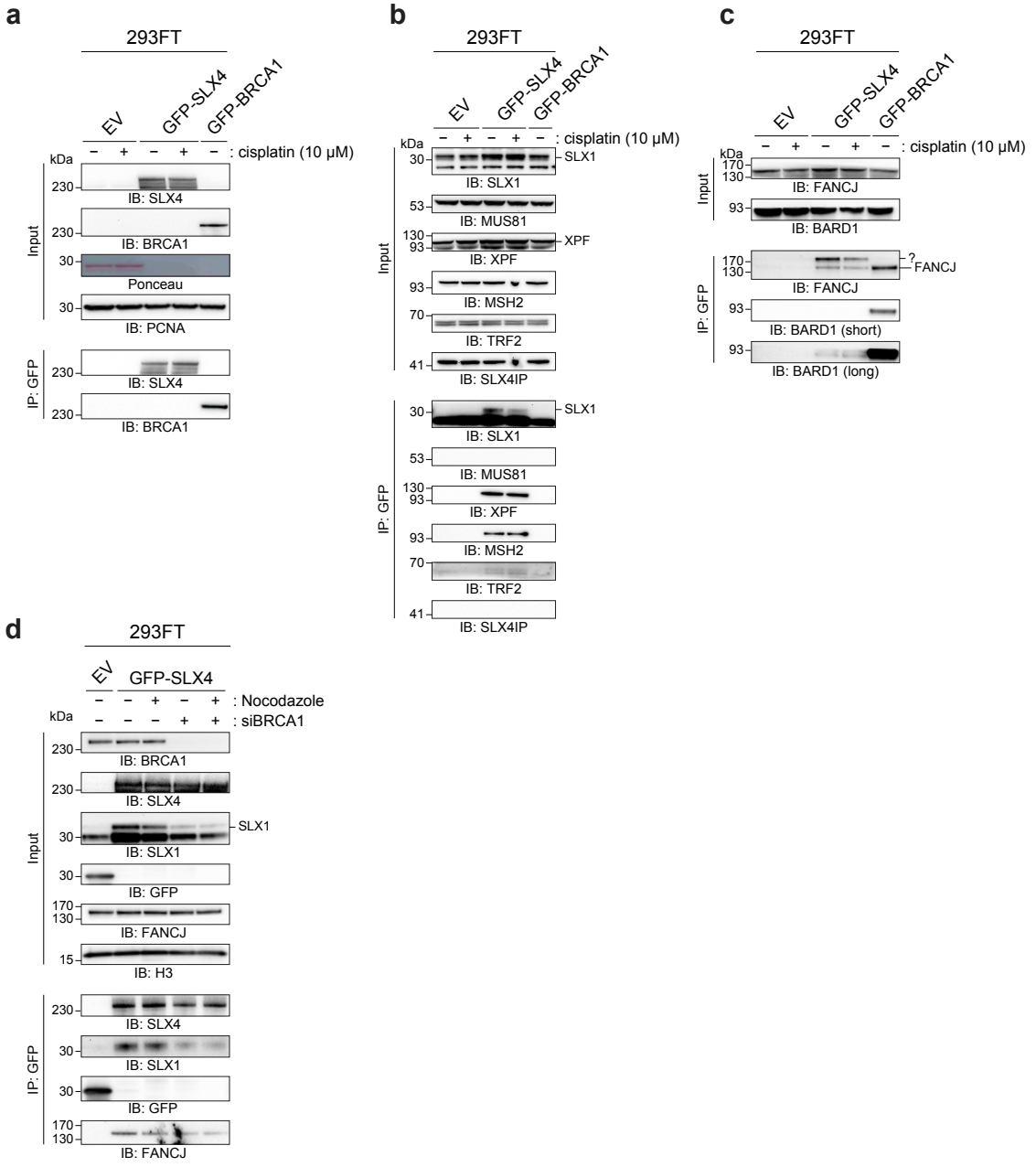


Figure 5.20 | SLX4 co-immunoprecipitation analysis in 293FT cells reveals interactions with DNA repair factors. a, Co-immunoprecipitation (IP) of GFP-SLX4 and GFP-BRCA1 in 293FT cells with or without 10 μM cisplatin treatment. Western blotting of input and GFP-IP fractions indicates that GFP-SLX4 specifically co-precipitates with BRCA1. **b,** SLX4 IP analysis probing for interacting proteins including SLX1, MUS81, XPF, MSH2, and TRF2. Input and GFP-IP fractions demonstrate robust SLX4 interactions with canonical SMX complex partners and telomere-associated protein TRF2. **c,** Western blot of FANCJ and BARD1 in GFP-SLX4 and GFP-BRCA1 IP fractions. Co-immunoprecipitation with GFP-SLX4 supports interactions with FANCJ, while both FANCJ and BARD1 are enriched in GFP-BRCA1 IP fractions. **d,** Co-immunoprecipitation of GFP-SLX4 in 293FT cells treated with nocodazole (to induce G2/M arrest) or BRCA1 siRNA knockdown. Western blotting reveals that SLX4 interactions with BRCA1 and FANCJ are modulated under these conditions, highlighting cell-cycle-dependent and BRCA1-dependent interactions.

SAP or CCD domains (Figure 5.21b). As expected, the interaction between SLX4 and XPF was abolished when the UBZ and MLR domains were deleted, confirming their necessity for this interaction (Figure 5.21b). Interestingly, a higher-molecular-weight variant of XPF was observed in immunoprecipitates of the D1 (lacking residues 1-559) and D3 (lacking residues 1093-1488) SLX4 truncation mutants (Figure 5.21b). This higher-weight XPF species was particularly enriched in the D3 GFP-SLX4 immunoprecipitate, which lacks the SUMO-interacting motifs (SIMs) (Figure 5.21b). These findings suggest a potential role for the SIMs of SLX4 in regulating putative post-translational modification of XPF. The interaction between SLX4 and SLX1 was disrupted in both the D4 GFP-SLX4 mutant (lacking residues 1489-1834) and SLX4$^{\Delta CCD}$ (Figure 5.21b), corroborating the importance of these regions for SLX1 binding[554,565]. Notably, the FANCJ-SLX4 interaction persisted across all truncation mutants tested (Figure 5.21a, b). The stability of the FANCJ-SLX4 binding could indicate the involvement of multiple interaction sites or an indirect association facilitated by other domains. Further biochemical and structural studies will be required to delineate the precise molecular determinants underpinning this interaction.

5.5 | Discussion

The study presents the progress that has been made in understanding the best strategies for purification and characterisation of SMX complexes and the interactions of SLX4 with DNA repair proteins, highlighting both advances and challenges in these areas. Previous attempts to purify SMX complexes faced challenges such as low yields, aggregation, and imbalances in subunit stoichiometry, even when using MultiBac and single-gene expression systems. The polyprotein strategy, which facilitates co-translational folding, showed promise in improving expression and solubility, but did not wholly resolve issues related to the stability of specific subunits. This instability and inconsistent expression underscores the inherent complexity of assembling and stabilising this large multi-protein complex (Figures 5.1–5.4).

Further trials with mouse homologues of SMX revealed advantages over the human and chimaeric constructs, including reduced aggregation and higher yields. However, even with mouse SMX, the separate expression of polyproteins resulted in incomplete complexes. Interestingly, SUMO-interacting motifs (SIMs) within SLX4 have been implicated in

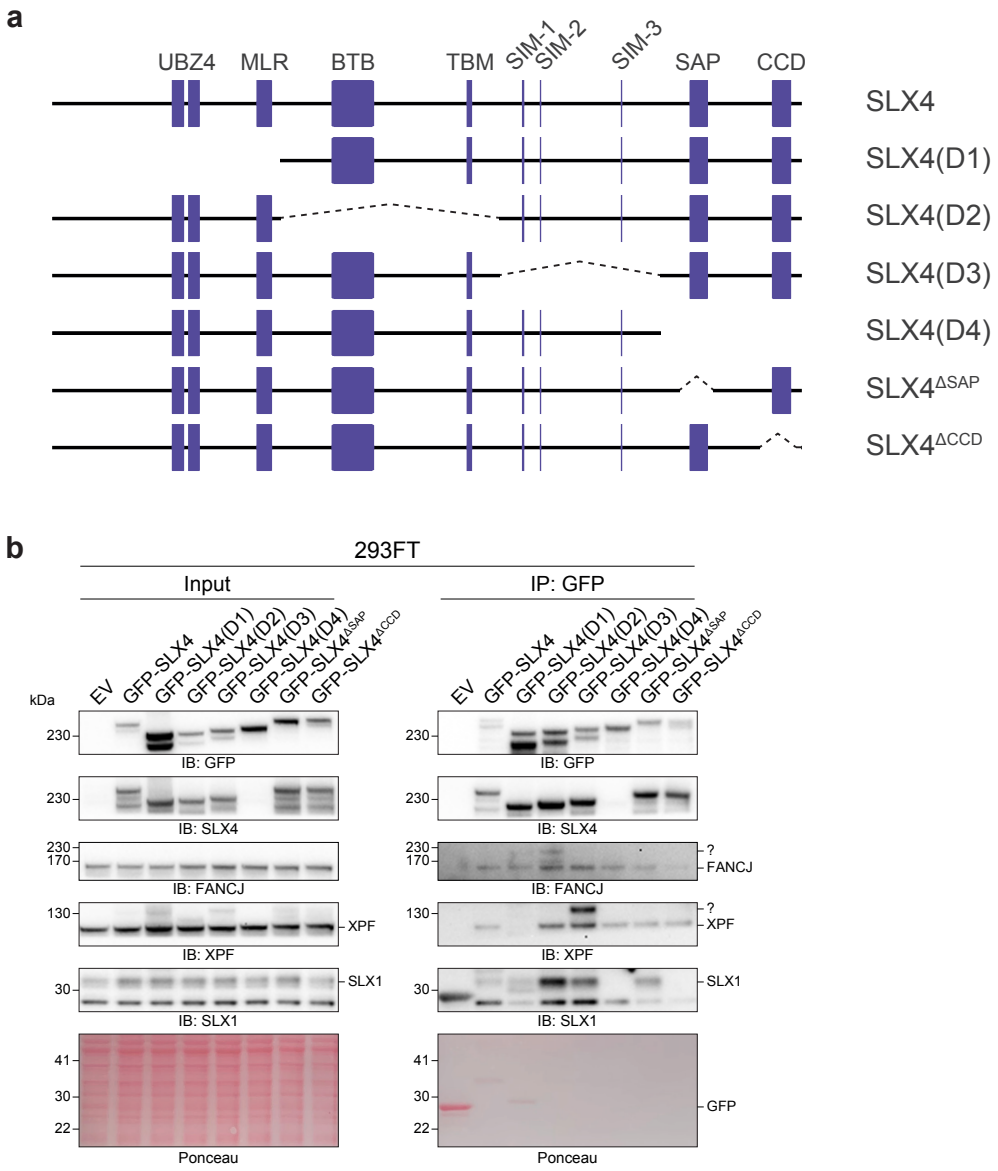


Figure 5.21 | Mapping SLX4 domains required for interactions with FANCJ and SLX1 in 293FT cells. a, Schematic representation of SLX4 truncation constructs highlighting key functional domains, including UBZ4 (ubiquitin-binding zinc finger 4), MLR (MUS81-LIG4 interaction region), BTB (broad-complex, tramtrack, bric-à-brac), TBM (Tel1-BRCT motif), SIM (SUMO-interacting motifs 1–3), SAP (SAF-A/B, Acinus, and PIAS), and CCD (coiled-coil domain). Specific truncation constructs, including SLX4(D1 - $\Delta 1-559$), SLX4(D2 - $\Delta 560-1092$), SLX4(D3 - $\Delta 1093-1488$), SLX4(D4 - $\Delta 1489-1834$), SLX4$^{\Delta SAP}$, and SLX4$^{\Delta CCD}$, are depicted to illustrate the retained and deleted regions. **b,** Co-immunoprecipitation (IP) analysis of GFP-tagged SLX4 truncation constructs expressed in 293FT cells. Western blot analysis of input and GFP-IP fractions shows that FANCJ interacts primarily with full-length SLX4, while deletion of specific domains alters this interaction. SLX1 and XPF interactions remain robust across multiple truncations, with input and IP fractions probed for GFP, SLX4, FANCJ, XPF, and SLX1. Ponceau staining is included as a loading control for total protein.

regulating post-translational modifications of certain subunits, such as XPF (Figure 5.21), which might complicate the decision to delete the central disordered region of SLX4. This finding suggests a potential avenue for improving complex stability, perhaps through the deletion of regions in SLX4 that otherwise regulate SMX subunits (Figures 5.7–5.9). Using a single-baculovirus approach might help resolve some of these assembly challenges, and further investigation into the role of SLX4's central domain and SIMs in subunit interactions could yield additional valuable insights.

To address limitations of insect-cell systems, the expression of human SMX complexes was explored in mammalian Expi293F cells. Transient transfections demonstrated that expression of the full SMX complex reduced cell viability, particularly with enzymatically-active forms (Figure 5.12). This toxicity suggested that the expression of nuclease-dead variants and shorter expression windows may help balance protein yield with cell survival. Despite these challenges, key SMX components were successfully expressed, purified, and characterised (Figure 5.12, 5.13). However, treatments such as nocodazole-induced cell cycle arrest or DNA damage induction did not significantly enhance complex assembly or yield, suggesting that SMX assembly may be independent of such interventions (Figures 5.11–5.12). Moving forward, inducible expression systems and stable cell lines represent promising strategies for scaling up protein production while mitigating toxicity, an important step to yield sufficient material for structural characterisation.

A pivotal aspect of this work was the detailed characterisation of SLX4's interactions with its binding partners and of its post-translational modifications. SLX4's phosphorylation by CDK1-cyclin B at specific residues, T1561 and T1571, was observed in a complex that includes MUS81 and SLX1 (Figure 5.17). This is in agreement with, but expands upon observations of similar phosphorylation within a shorter SLX4 peptide without $SLX1^{577}$. Meanwhile, co-immunoprecipitation and structural modelling identified a stable and previously unexplored interaction between SLX4 and FANCJ. Notably, this interaction persisted even after BRCA1 depletion, suggesting a direct and BRCA1-independent binding mechanism (Figures 5.19–5.21). These experiments were also carried out in the presence of benzonase, suggesting this interaction is not mediated *via* nucleic acids. While the physiological role of this interaction remains to be determined, characterising this will be the focus of future work. To confirm the validity of this interaction, future

initial experiments will involve carrying out reciprocal co-immunoprecipitations using over-expressed FANCJ.

Overall, the findings presented in this chapter advance our understanding of SMX complex assembly and SLX4's functional network while highlighting the challenges associated with purifying and stabilising these complexes. Further work is needed to refine expression and purification systems, explore the biochemical and structural determinants of SLX4-FANCJ interactions, and elucidate the role of post-translational modifications in regulating SMX nuclease activity. The insights gained from this study pave the way for deeper investigations into the role of SMX complexes in DNA repair and replication stress responses.

6 | Discussion

Contents

This study explored the biological roles of SLX4 and its associated structure-specific endonucleases, particularly SLX1 and XPF, as well as their interactions with formaldehyde-induced DNA damage. This research aimed to address questions spanning structural, biochemical, and cellular perspectives, providing new insights into aldehyde biology and the broader context of DNA repair pathways. The timeline of this thesis broadly followed a progression from biochemical to cellular studies. The early years were primarily dedicated to the design and cloning of constructs aimed at purifying the SMX complex, beginning with chimaeric assemblies, followed by murine constructs in the second year, and culminating in efforts to express the human complex in the final year. Recognising the significant technical challenges involved in achieving structural characterisation of SMX, a parallel stream of work, as in chapter 4, emerged in the third year focused on cell biology, particularly the functional analysis of XPF and SLX1. Around this time, new experimental directions were explored to investigate how formaldehyde impacts DNA crosslinking and the Fanconi anaemia pathway. Collaborative projects to address these questions, which are detailed in chapter 3, were also initiated during the third year and continued alongside ongoing expression and purification work.

6.1 Insights into Aldehyde Biology and Formaldehyde Damage

A major objective of this thesis was to establish tools to study endogenous aldehyde biology, hypothesising that aldehydes, likely generated during processes such as epigenetic reprogramming, could form DNA interstrand crosslinks (ICLs) that require detoxification or repair through pathways involving FA proteins and XPF. To investigate this, a model using induced pluripotent stem (iPS) cells was developed, as these can be differentiated into neurons, a process expected to release high levels of formaldehyde. Despite sequence-verified generation of *ADH5*$^{-/-}$ and *ALDH2*$^{-/-}$ clones, functional assays revealed no detectable increase in formaldehyde sensitivity or DNA damage markers like γH2AX, though given a short time in the collaborating lab, many of these experiments represent single biological repeats (Figure 3.5). One concern regarding these studies is that the gene deletions generated were not complete knockouts, however, it is also possible that some redundancy in detoxification pathways conceals the expected phenotypes. No induction of γH2AX was observed in iPSC-derived neurons lacking ADH5. This absence may indicate that any DNA damage from aldehyde accumulation had already been resolved by the time differentiation was complete. A more informative approach would be to assess whether the differentiation process itself is impaired in these knockout cells, potentially through the additional use of an ADH5 inhibitor (ADH5i). Alternatively, monitoring DNA damage, such as γH2AX induction, at various stages of differentiation could provide insight into whether aldehyde-induced genomic instability occurs during the differentiation process. This strategy would offer a clearer understanding of the temporal relationship between aldehyde metabolism, DNA damage, and cellular differentiation.

Further studies mapped formaldehyde-induced DNA damage to identify specific hotspots and compared these patterns to known crosslinking agents such as SJG-136. The damage signature in post-mitotic cells defined by crosslinking agents is repair-associated synthesis at gene bodies, specifically on the transcribed strand, showcasing a mostly uncharacterised transcription-coupled ICL repair pathway. The interstrand crosslinking efficiency of a range of drugs tested in neurons is associated with the strength of this particular transcription-associated damage signature; it is most well-observed in treatment with SJG-136 (a drug that predominantly creates ICLs), less so with cisplatin (which

generates only a few ICLs compared to other lesions); and almost entirely lost with illudin S, which generates single-strand adducts (unpublished results from collaborator Dr. William Nathan). SJG-136 crosslinks link two strands *via* N^3-position of guanines within the minor groove at consensus sequences of `GATC/CTAG` and therefore produce minimally distorting lesions that fail to engage any GG-NER, therefore explaining the reliance on transcription-coupled repair[861]. These findings indicated that formaldehyde alone did not create hotspots analogous to SJG-136. However, when combined with ADH5 inhibition, formaldehyde exhibited a pattern consistent with crosslink formation (Figure 3.6). This concept aligns well with findings from the Patel group, who have proposed a two-tiered protection mechanism against formaldehyde, *via* catabolism of aldehydes by ADH5/ALDH2 or failing that, the repair of aldehyde-induced crosslinks[9,757].

The broader relevance of formaldehyde-induced DNA damage differs significantly between post-mitotic and rapidly dividing cells, primarily due to the varying metabolic and repair environments. As discussed in chapter 3, one major endogenous source of formaldehyde is histone demethylation, a process that becomes particularly prominent during epigenetic reprogramming associated with cellular differentiation. In post-mitotic cells, such as terminally differentiated neurons, large-scale reprogramming events are generally considered absent, and thus, the release of formaldehyde through these mechanisms is expected to be minimal. Moreover, the FA pathway-dependent repair of general and formaldehyde-induced ICLs is predominantly active during S phase. However, some emerging evidence suggests that post-mitotic neurons are not necessarily protected from this formaldehyde-related genotoxicity. Neuronal plasticity, a key feature of learning and memory, is underpinned by dynamic epigenetic changes, including histone demethylation events that could plausibly generate bursts of local formaldehyde[758]. How such DNA damage is resolved in the absence of replication-coupled repair pathways remains largely unknown. Some studies have proposed that transcription-coupled repair mechanisms, particularly those involving CSB, may operate in neurons to resolve these lesions *via* an NER-like pathway[756]. Unlike in eukaryotes, where ICL repair often requires replication-coupled processes, bacterial systems predominantly rely on NER for the resolution of these lesions[862]. The genotoxic impact of formaldehyde in neurons is supported by murine studies showing that the neurological phenotypes associated with Cockayne syndrome

(CS) only manifest when CSB loss is combined with the deletion of the formaldehyde-detoxifying ADH5[58]. This synergy suggests that formaldehyde-induced lesions accumulate in neurons and that their repair becomes dependent on CSB-mediated processes. Furthermore, the high metabolic demand of the brain imposes a substantial requirement for reactive oxygen species (ROS) scavenging. As formaldehyde can react with glutathione and deplete antioxidant defences, impaired aldehyde clearance may exacerbate oxidative stress, indirectly increasing genotoxicity and the need for DNA repair[863]. Together, these findings imply that although the sources and pathways of formaldehyde damage differ between cell types, both post-mitotic and proliferative cells are vulnerable in distinct but biologically significant ways. Importantly, this mechanism may have implications for specific cancer types. Cancers arising in tissues with high turnover or frequent differentiation, such as the haematopoietic system or intestinal epithelium, may be particularly vulnerable to aldehyde-induced genome instability. As above, Fanconi anaemia, which prominently features bone marrow failure and predisposition to myeloid malignancies, provides a strong example where aldehyde-induced damage during differentiation may be causative[864]. Understanding the interplay between differentiation, aldehyde metabolism, and DNA repair may therefore offer new perspectives on tumour biology and potential therapeutic vulnerabilities.

HT-29 colorectal cancer cells provided another tractable model for studying aldehyde-induced damage during differentiation. In agreement with previous research[795,797], HT-29 cells can be differentiated into goblet cells through a simple medium change from RPMI to DMEM (Figure 3.7). Crucially, XPF-deficient cells showed reduced expression of MUC2, a goblet cell marker, which was further exacerbated by ADH5 inhibition (Figure 3.8). These findings imply that epigenetic remodelling during differentiation generates aldehyde-mediated crosslinks, which, in the absence of two-tier protection (XPF-deficiency and ADH5 inhibition)[9], impede proper differentiation. Another explanation may be that the clonal expansion of the HT-29 Δ*ERCC4* rendered this particular population of cells less stem-like, while the parental WT cells were not as recently derived from a single-cell clone. These experiments could be repeated in other verified HT-29 Δ*ERCC4* clones to verify there are no clonal effects. Furthermore, more quantitative methods could be used to support these findings, perhaps employing MUC2 flow cytometry or searching for DNA damage markers during the differentiation process using IF.

An alternative strategy for studying formaldehyde-induced crosslinking damage was developed that bypassed differentiation by directly engineering a system capable of releasing formaldehyde near DNA through the targeted activity of histone demethylases at specific genomic loci. Initial results demonstrated successful inducible expression and the induction of DNA damage at targeted promoters and telomeres, marked by γH2AX foci (Figure 3.11). This system represents a promising tool for investigating the contributions of endogenous aldehydes to genome instability and the role of XPF-mediated repair pathways in mitigating this damage. While the observed link between formaldehyde generation and XPF activity suggests the possibility of formaldehyde generating ICLs, another plausible explanation is the formation of DNA-protein crosslinks (DPCs). XPF may play a role in repairing these protein scars on DNA through a nucleotide excision repair (NER)-like pathway[865]. Determining whether ICLs or DPCs are predominantly implicated in the studies presented in this chapter remains a challenge. One potential approach is the alkaline comet assay; however, in the inducible LSD1-dCas9 system, the limited number of genomic loci targeted may render ICL detection difficult. Alternatively, genomic tools such as S1-END-seq (to identify XPF-dependent nicks at ICLs) could be employed, with careful timing of damage events during specific cell cycle phases or through the knockdown of key DPC repair proteins like SPRTN, to ascertain whether DPC repair is involved. Additionally, Prof. Jean-Yves Masson's group has developed immunofluorescence-based methods for inferring the presence of ICLs by probing for the colocalisation of γH2AX with repair proteins such as MRE11, RPA, HMCES, and EXO1. This approach is currently being implemented as part of ongoing collaborative work to elucidate the nature of the lesions generated in this system.

6.2 | Towards understanding XPF's genetic interactions

XPF-ERCC1 is a cornerstone of DNA damage repair, performing decisive incisions during nucleotide excision repair (NER) and ICL repair pathways. Patients with *ERCC4* mutations display diverse phenotypes, ranging from Xeroderma pigmentosum to Fanconi anaemia, emphasising its broad functionality. This study generated U2OS, A549, and HT-29 XPF-deficient cell lines (Figure 4.1) and evaluated their sensitivity to DNA-damaging agents

that have been less well characterised in an XPF-ERCC1-deficient context, including formaldehyde, SJG-136, and olaparib (Figure 4.2, Figure 4.3, Figure 4.4). The particular sensitivities were assayed alongside MUS81 and BRCA1 depletion, in efforts to better understand whether repair of these lesions by XPF occurred in the context of a pathway that also employed some form of HR.

The loss of XPF conferred resistance to trabectedin, consistent with this drug's dependence on transcription-coupled NER (TC-NER). Unexpectedly, BRCA1 depletion specifically reduced viability in XPF-deficient cells exposed to trabectedin (Figure 4.5), suggesting that trabectedin lesions not engaged by XPF are later substrates for HR, or that they can lead to replication stress, requiring BRCA1. Nevertheless, these cell lines confirmed expected functional phenotypes of XPF loss - extreme sensitivity towards crosslinking agents, formaldehyde and SJG-136, less pronounced but definitive sensitivity towards olaparib, although in a manner independent of MUS81/BRCA1. This validation provided sufficient confidence in the cell lines to perform a first CRISPR-KO screen to identify targets of XPF synthetic sickness/lethality (Sections 2.3.2.4, 2.5.4).

6.3 | Deletion of SLX1 in human cell lines

Despite its role within the SMX complex, SLX1 remains poorly characterised. Preliminary RNA interference experiments indicated that SLX1 depletion reduces colony outgrowth in *p53*-proficient cell lines, suggesting a role in cellular homeostasis mediated by a *p53*-dependent pathway (Figure 4.8). Efforts to generate SLX1-knockout cell lines revealed challenges due to its genomic context, characterised by near-identical paralogues (*SLX1A* and *SLX1B*). This aspect of SLX1 biology is especially unexplored, with genomics focusing on the other genes within this cluster, namely *BOLA2* in describing associations with haematological traits/anaemia and the copy number variations that strongly predict autism[835]. While this is a reasonable avenue of research since BOLA2 is involved in the maturation of iron-sulfur proteins[866], the intrinsic link between the FA pathway/ICL repair might also be explained through SLX1 changes during expansions of this cluster. Further links between neurodevelopmental conditions and SLX1 have been reported in the form of the requirement of SLX1 in enabling pathological expansion of trinucleotide repeats at the *FRAXA* locus through a MiDAS pathway[639]. What mechanistic role SLX1 plays in these

nucleotide repeat regulating processes, or if these associations are indirectly caused by perturbation to the genomic context around *SLX1* genes, requires attention. Nevertheless, targeting exons 2 and 3 *via* CRISPR-Cas9 produced putative knockouts in HAP1 and 293FT cells, which were validated by Sanger sequencing, RT-qPCR, and following extensive antibody optimisation and probing of nuclear extracts, Western blotting (Figure 4.11, Figure 4.12). Viability assays demonstrated variable responses to DNA-damaging agents, verifying functional knockout with MMC[511] (Figure 4.13. Sensitivity to cisplatin and SJG-136 was expected, implicating SLX1 in contributing to ICL repair. The identity of the nuclease responsible for the complementary nick to XPF that facilitates the 'unhooking' of an ICL has been the subject of ongoing debate and investigation[400]. SNM1A's epistasis with XPF strongly hints at its contribution to unhooking[445], but whether similar epistasis exists for SLX1 is an obvious outstanding experiment. It is conceivable that SLX1 is well-placed and capable of completing this reaction, given its notable promiscuity[675] and obligate heterodimerisation with SLX4, which is already involved in recruiting XPF to ICLs[435]. The sensitivity of SLX1-null clones towards ICL-inducing agents suggests that, at least in some manner, SLX1 is involved in this process. Complementation with inducible GFP-SLX1 partially restored sensitivity in 293FT cells but not in HAP1 cells (Figure 4.15). There are a few explanations that can be considered to explain these differences. First, and most plausibly, the continuous drug treatment (apart from cisplatin) adopted for these assays might mean that DNA damage outlasts the re-expression of SLX1 by doxycycline. Replenishment of doxycycline or removal of drug treatment after 24 h may help to mitigate this issue. Alternatively, SLX1 overexpression (Figure 4.14) may be particularly toxic in HAP1 cells. One driving factor between these cell line differences could be their relative cell volumes. In yeast, aneuploidy, causing proteomic imbalance, results in hypo-osmotic-like stress features[867]. In smaller cells, artificial overexpression could create similar imbalances relating to unexpected toxicity. Lastly, the observed phenotypes may represent non-targeted editing of genomic loci beyond just SLX1 loss or reflect impairment to nearby genes, such as *BOLA2*.

Since no RPE-1 cells with SLX1 deletion could be expanded and given the issues of complementation in HAP1 cells, inducible degradation of SLX1 was sought. As with knockouts, the genome organisation of *SLX1* and the existence of paralogues made many

knock-in strategies, such as Cre-Lox systems, impractical and instead, a degron knock-in approach was established. The minimally-disruptive FKBP tag (owing to its smaller size compared to other degrons) was knocked into the *SLX1* gene to tag its N-terminus, and clones were obtained with all alleles edited by knock-in events (Figure 4.16. Low native expression of SLX1, and the poor detectability of a single FLAG tag (or TwinStrep) meant that confirming knock-in and degradation at the protein level was challenging, requiring the transient overexpression of SLX4 to maximise differences in SLX1 protein stability upon dTag-13 addition between WT and knock-in clones (Figure 4.17). In one clone, RPE-1 $^{\mathrm{FKBP-SF}}$*SLX1* 6, depletion of SLX1 was mostly achieved upon dTag-13 addition and resulted in some sensitivity to SJG-136, akin to the HAP1 and 293FT SLX1 knockout cells (Figure 4.17). The fusion construct, including FKBP, has been redesigned to include tags with better low-expression detection (pCI-Neo:*SLX1*donor3, includes the HA epitope tag) and also swapped to invoke the multiple uses of the Halo-tag (pcDNA3.1-Zeo:*SLX1*donor4), which can be chemically linked to a fluorophore to permit cell sorting and can result in the ligand-induced degradation of its fusion, SLX1, as with FKBP.

6.4 | Working towards structural characterisation of the SMX complex

The structural characterisation of the SMX complex remains a long-standing goal, offering insights into SLX4‘s scaffolding mechanisms and the activity of its associated nucleases. High-resolution structures would improve understanding of the effect of disease mutations, especially in Fanconi anaemia, and aid drug discovery. The work in this thesis embraced many different strategies for purification, focusing ultimately on the promising utility of the human Expi293F suspension culture system (Figure 5.12) and with one particular mouse SMX construct (Section 5.2). Initial efforts using a human-frog chimaeric SMX complex faced issues of aggregation and subunit loss (Figure 5.2), and although refinements to purification protocols improved particle homogeneity and activity (Figure 5.4, Figure 5.5, Figure 5.6), challenges persisted due to the complex’s intrinsic flexibility and disordered regions. Furthermore, the non-physiological nature of a multi-species hybrid remained a concern during the project.

Shifting to murine SMX complexes provided a more tractable system due to their smaller disordered region in the central segment of SLX4. Attempts to co-express SXE and EMS components in insect cells revealed incomplete complex assembly, favouring the N-terminal subunits only and prompting a transition to a single-baculovirus system (Figure 5.8, Figure 5.9). The most significant barrier to complete SMX assembly was the unforeseen issue of expressing SLX1 from a separate plasmid/baculovirus to SLX4, as SLX4 is absolutely required for SLX1 stability[540]. With lytic baculoviruses, even under co-infection, given recommendations to ensure that the multiplicity of infection does not exceed 1, it is rare that an individual cell would be transduced with each plasmid. So, SLX1 stability would likely be compromised where it is expressed independently. Options exist to generate multi-baculovirus constructs from the individual pAceBac1 plasmids 'SXE' and 'EMS' through recombination[839]. The biGBac system is another option, allowing for 'mix and match' of genes as needed once set up correctly[868]. This system has been helpful in the purification and structural characterisation of several multi-protein complexes, such as APC/C[869] and the RAD51 paralogue complex[671]. In the interest of time, a single pFastBac-Dual vector was cloned, consisting of all murine SMX components. The resulting construct showed promise, with improved yields and stoichiometry (Figure 5.10), but further optimisation is required.

Insect cells have proven to be a useful tool for the purification of certain mammalian proteins due to their recapitulation of post-translational modifications, however this is not always perfectly faithful. The additional *in vitro* phosphorylation of the insect cell-purified C-terminal SLX4 subunits (SLX4-C) by CDK1-cyclin B suggests that insect cells fail to add all phosphorylation marks to the SMX subunits (Figure 5.17). The development of mammalian expression systems allowed for better recapitulation of post-translational modifications (PTMs) and native conditions. The SMX complex genes were cloned into three separate plasmids, each incorporating distinct affinity tags on the proteins to streamline downstream purification and enable clear visualisation *via* immunoblotting.

Transient transfections revealed that overexpression of SMX components significantly reduced cell viability, prompting the need to optimise transfection conditions and explore alternative expression systems (Figure 5.12). To minimise cytotoxic effects, transfection times were limited, and inducible expression systems were investigated as a more controlled

approach. Initial attempts to purify the complex using homemade GFP nanobody beads yielded suboptimal results, with preparations exhibiting contamination and bead quality inferior to commercially available alternatives (Figure 5.12). Further testing revealed that Strep-tagged beads provided the most consistent results, and these were identified as the preferred option for future experiments.

Given the challenges posed by potential toxicity, inducible cell lines were developed as a more feasible solution. A proof-of-concept inducible SLX4IP expressing cell line was successfully generated, demonstrating robust induction upon doxycycline treatment (Figure 5.14). Subsequently, an inducible SLX4 expressing line was created; however, its expression did not significantly enhance the levels of other SMX components, and nuclease activity on a model SMX substrate remained undetectable (Figure 5.15). To address this limitation, efforts were redirected towards creating combinatorial inducible systems. Preliminary work included generating and freezing triple-positive cell populations that express all SMX subunits in units of two (Figure 5.14, Figure 5.16), though the initial number of these cells was extremely low. Further optimisation, including re-sorting and expansion of these populations, will be required to achieve sufficient protein yields for downstream experiments.

Phosphorylation of SLX4 and EME1 was observed in insect cell preparations, with native phosphorylation levels further enhanced upon the addition of CDK1. This phosphorylation was completely reversed by treatment with λ-phosphatase and mass spectrometry identified, in agreement with previous work[577], a cluster of two phosphorylation sites on SLX4 at residues 1561 and 1571. The novelty in the repetition of this notable work is that these same phosphorylation marks are present in a larger complex of C-terminal SLX4, which also consists of SLX1. Whether these PTMs influence the activity of SLX1 is something that can be ascertained biochemically. Additionally, the cellular relevance of these sites warrants further investigation. Potential studies could include generating phospho-mimetic (*e.g.*, S1561D/S1571D) or phospho-null (*e.g.*, S1561A/S1571A) SLX4 mutants to assess changes in SMX complex activity, stability, or cellular localisation.

A study has recently proposed an interaction between PCNA and SLX4[605], but this hypothesis appears questionable based on AlphaFold modelling and the poor consensus of putative SLX4 'PIP boxes' to verified PIP boxes (Figure 5.18). The predicted structural

alignment and a co-immunoprecipitation performed with GFP-SLX4 suggest that any interaction, if present, is likely weak or transient. In contrast, an interaction between SLX4 and FANCJ appears more plausible, especially given their demonstrated mutual co-dependency in various repair pathways. Experimental data indicate a stable SLX4-FANCJ interaction, which persists even under high salt wash conditions, suggesting a strong and biologically meaningful direct association (Figure 5.20). Notably, co-immunoprecipitation assays revealed that SLX4 associates with a higher molecular weight form of FANCJ, potentially reflecting post-translational modifications such as phosphorylation, ubiquitylation, or acetylation on FANCJ that is recognised by SLX4. Unfortunately, efforts to map the SLX4 domain responsible for FANCJ binding using truncation mutants did not disrupt the interaction, suggesting the involvement of multiple redundant binding sites or the mediation of this interaction by co-factors within a larger protein complex (Figure 5.21). These findings highlight the complexity of SLX4's scaffold function and its ability to integrate various repair pathways.

Future studies should aim to identify the precise regions of this interaction, potentially through crosslinking mass spectrometry or more confined mutational analysis of both proteins. The higher molecular weight form of FANCJ may result from post-translational modifications (PTMs), such as phosphorylation at T1133 and S990, which mediate interactions with TopBP1 and BRCA1, respectively, or acetylation at K1249[870]. To confirm this, these sites could be mutated to non-modifiable residues (*e.g.*, alanine or arginine substitutions) to determine whether the higher molecular weight band disappears. Mass spectrometry could further identify specific modifications, providing insights into how PTMs regulate FANCJ's activity, interactions, and role in DNA repair processes. The link between two key processes in the FA pathway, the monoubiquitylation and stabilisation of FANCD2-FANCI and the recruitment of SLX4-XPF-ERCC1 for the primary incision, has yet to be coupled by compelling evidence (Section 1.4.4). FANCJ might represent a missing piece in this picture. It can be hypothesised that FANCD2-FANCI stabilisation on dsDNA prevents the helicase activity of FANCJ, which is likely recruited early to ICLs by its association with the replicative machinery through MMR complexes[455,853], which might otherwise cause its dissociation after catalysis. This retention of FANCJ might then be a key recruitment signal for SLX4-XPF. Assaying the recruitment of SLX4

to ICLs in FANCJ-null cells, or where SLX4-binding-deficient mutants can be created, could shed light on this hypothesis.

6.5 | Implications and future directions

One significant lesson learned from the expression of recombinant proteins using baculovirus-transduced insect cells is the notable variability in per-expression protein yield. Incorporating a downstream, in-frame fluorescent reporter (adopted for mouse SMX constructs) can significantly enhance the ability to monitor harvest quality. However, the inclusion of such tags requires careful consideration of the payload size. For the SMX complex, which has a total expression payload exceeding 500 kDa, the addition of a 27 kDa fluorescent tag is unlikely to introduce substantial challenges. Balancing these design factors is essential for optimising recombinant protein expression systems. Future structural studies may focus on stabilising the SMX complex, potentially through the incorporation of post-translational modification (PTM) mimetics or binding partners, to enable high-resolution characterisation. Investigating the interactions of SLX4 with FANCJ and other proteins could provide critical insights into their biological functions and therapeutic potential. Nonetheless, it is conceivable that the SMX complex is intrinsically too disordered to yield well-defined single particles suitable for structure reconstruction. Should this be the case, refining purifications to isolate more tractable subcomplexes remains a worthwhile endeavour, offering valuable mechanistic insights into the cellular functions of SLX4 and the broader SMX complex.

The generation of XPF- and SLX1-deficient cell lines, alongside their proficient counterparts, is intended to facilitate CRISPR-based screens to identify factors genetically linked to these components. While CRISPR screens are often regarded as hypothesis-independent and exploratory in nature, their application is particularly relevant in this context. For SLX1, many phenotypes are likely masked by functional redundancy with other, more highly expressed nucleases. For XPF-deficient cells, ongoing screens can be complemented by validation experiments involving knockdown approaches. To this end, HT-29 wild-type and Δ*ERCC4* lines expressing KRAB-dCas9 for CRISPR interference (CRISPRi)-based knockdown have been developed, ensuring robust experimental validation of candidate hits. The creation of a verified SLX1 degron-fusion cell line will enable CRISPR screens and

provide additional understanding of SLX1 biology. This approach is particularly significant given that SLX1 loss appears to induce cell death in a p53-dependent manner (Figure 4.8). Employing degron systems in both p53-proficient and p53-deficient cell lines could reveal the underlying mechanisms of this dependency and identify pathways contributing to SLX1-associated vulnerabilities. Further investigation into the specific vulnerabilities of SLX1-deficient or degron-targeted cells can be conducted using established techniques. For example, fluorescent reporter systems can elucidate potential defects in homologous recombination (HR), non-homologous end joining (NHEJ), or other double-strand break (DSB) repair pathways[871,872]. Additionally, comet assays could determine whether lesions persist in these cells and provide insights into the nature of these lesions, including the potential for unresolved crosslinks[873,874]. These complementary approaches will contribute to a comprehensive understanding of the genetic and functional landscape associated with XPF and SLX1 deficiencies.

The inducible formaldehyde generation system holds significant potential for establishing formaldehyde as an *in vivo* interstrand crosslinking agent. If this system can reliably generate ICLs, it would enable studies examining the differential impact of ICLs on the leading *versus* lagging strands during replication and aid in understanding mechanisms of ICL bypass[399,449]. For instance, guide RNAs targeting regions near a mapped origin of replication could help delineate strand-specific repair pathways, akin to recent work employing nicking Cas9 to study the same relationships for SSB repair[875,876]. Further functionalities of the system could include switching the dCas9 component to a nicking Cas9, allowing the generation of single-strand nicks around formaldehyde-generated lesions that could be evaluated for their dependency on XPF for repair. This would offer insights into whether the absence of XPF impacts the repair of nick-primed crosslinks differently from intact crosslinks. It may also aid identification of which nuclease performs the 5′ incision, or whether XPF itself does this. Additionally, CRISPR-based genetic screens could be employed to identify genomic regions where formaldehyde-induced crosslinks are particularly difficult to repair, potentially leading to elevated cytotoxicity. This approach could be especially informative in the context of transformed cells, providing insights into vulnerabilities that might be therapeutically exploited. Moreover, the system offers avenues for this therapeutic repurposing, such as swapping the catalytic domain of

the histone demethylase with enzymes capable of activating DNA-damaging prodrugs at specific genomic loci. This could facilitate precision-targeted approaches to generate or amplify DNA damage, contributing to the development of innovative therapeutic strategies.

Appendices

A | Additional Tables

A.1 | Plasmids

Table A.1 | Details of plasmids used. The plasmids used and how they were cloned are listed here. Within the ORF1-3 descriptions, any L^x sequence refers to a linker of x length. Other abbreviations: TEVP - TEV protease; t - TEV protease cleavage site; ECFP - enhanced cyan fluorescent protein; SF - TwinStrep-FLAG tag; T2A/P2A - self-cleaving peptide sequence; PuroR - puromycin resistance cassette; HygroR - hygromycin resistance cassette; BleoR - bleomycin/zeocin resistance cassette; rrTA-Advanced - mutant Tet repressor fused to minimal F-type transcriptional activation domain from VP16 protein; SUMO - SUMO-tag, consisting of a Ulp1 protease cleavage site; GFP-Nb - GFP nanobody; mCherry-Nb - mCherry nanobody; FKBP - FK506 binding protein; HA - hemagglutinin-tag; HiBiT - small 11-amino acid, epitope tag able to produce a bioluminescent signal when bound to $LgBiT^{877}$; KRAB - Kruppel associated box; eDHFR - *E. coli* dihydrofolate reductase.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P1	pKLPBac:χSXE	Insect	polyhedrin	1. L^4-TEVP-L^{12}-t 2. SF-L^6-xSLX4-L^8-t 3. L^{11}-xXPF-L^8-t 4. L^{11}-hERCC1-L^9-t 5. ECFP			Chimaeric SLX4-XPF-ERCC1 expression as polyprotein with TEV protease (TEVP)	This study
P2	pAceBac1:χEMS	Insect	polyhedrin x3	^{2}-His_6-L^{10}-t-L^1-hEME1	hMUS81	xSLX1-His_6	Chimaeric His6-EME1, MUS81, SLX1-His6 expression each under individual polyhedrin promoter	Gift of Dr. Denis Ptchelkine

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P3	pAceBac1:mSXE	Insect	polyhedrin	1. TEVP-tL^{11}-mSLX4-L^{8}-t 2. L^{10}-mXPF-L^{8}-t 3. L^{1}-SF-L^{8}-mERCC1			Mouse SLX4-XPF-ERCC1 expression as polyprotein with TEV protease (TEVP)	This study
P4	pAceBac1:mSΔXE	Insect	polyhedrin	1. TEVP-t 2. L^{11}-mSLX4$^{\Delta 759-1252}$-L^{8}-t 3. L^{10}-mXPF-L^{8}-t 4. L^{1}-SF-L^{8}-mERCC1			Mouse SLX4Δ-XPF-ERCC1 expression as polyprotein with TEV protease (TEVP)	This study
P5	pAceBac1:mEMS	Insect	polyhedrin	1. TEVP-t 2. L^{11}-mEME1-L^{5}-t 3. L^{8}-mMUS81-L^{5}-t 4. L^{7}-mSLX1-L^{3}-His$_{10}$			Mouse EME1-MUS81-SLX1 expression as polyprotein with TEV protease (TEVP)	This study

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P6	pFastBac-Dual:mSXE,EMS	Insect	polyhedrin (ORF1), p10 (ORF2)	1. L^2TEVP-L^5-t 2. L^{11}-mSLX4$^{\Delta 759-1252}$-L^8-t 3. L^{12}-mXPF-L^{10}-HaloTag-L^4-t 4. L^1-SF-L^4-EGFP-L^2-3c-L^2-mERCC1	1. mEME1-L^5-t 2. L^8-mMUS81-L^5-t 3. L^7-mSLX1-L^3-His_{10}		Dual mouse SLX4-XPF-ERCC1 and EME1-MUS81-SLX1 expression as polyproteins from opposing promoters, polyhedrin and p10	This study
P7	pEGFP-C1	Mammalian	CMV	EGFP			GFP empty vector (EV)	Gift of Dr. Lonnie Swift
P8	pX459	Mammalian	U6 (ORF1), CMV enhancer/chicken β-actin		3xFLAG-SpCas9-T2A-PuroR		Empty pX459	ref.[736]

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P9	pX459:sg*SLX1*-N1	Mammalian	U6 (ORF1), CMV enhancer/chicken β-actin	sg*SLX1*-N [GAAGGCCCGATGGGTCCCGCG]	3xFLAG-SpCas9-T2A-PuroR		pX459 expressing sgRNA targeting N-terminus of SLX1	This study
P10	pX458	Mammalian	U6 (ORF1), CMV enhancer/chicken β-actin		3xFLAG-SpCas9-T2A-EGFP		Empty pX458	Gift of Dr. Philip Hublitz
P11	pX458:sg*SLX1*-ex1-1	Mammalian	U6 (ORF1), CMV enhancer/chicken β-actin	sg*SLX1*-ex1-1 [GTACCGGGGCCGCGTCTACG]	3xFLAG-SpCas9-T2A-EGFP		pX458 expressing sgRNA targeting exon 1 of *SLX1* #1	This study
P12	pX458:sg*SLX1*-ex1-2	Mammalian	U6 (ORF1), CMV enhancer/chicken β-actin	sg*SLX1*-ex1-2 [CGTAGACGCGGCCCCGGTAC]	3xFLAG-SpCas9-T2A-EGFP		pX458 expressing sgRNA targeting exon 1 of *SLX1* #2	This study

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P13	pX458:sg*SLX1*-int1-1	Mammalian	U6 (ORF1), CMV enhancer/chicken β-actin	sg*SLX1*-int1-1 [CATCTCCGCGGCGGAACTCA]	3xFLAG-SpCas9-T2A-EGFP		pX458 expressing sgRNA targeting intron 1 of *SLX1* #1	This study
P14	pX458:sg*SLX1*-int1-2	Mammalian	U6 (ORF1), CMV enhancer/chicken β-actin	sg*SLX1*-int1-2 [GGGCTGCGACAGGGACATCA]	3xFLAG-SpCas9-T2A-EGFP		pX458 expressing sgRNA targeting intron 1 of *SLX1* #2	This study
P15	pX458-mRuby2:sg*SLX1*-int3-1	Mammalian	U6 (ORF1), CMV enhancer/chicken β-actin	sg*SLX1*-int3-1 [AGAGCCATCCATTCCTCGGG]	3xFLAG-SpCas9-T2A-mRuby2		pX458 expressing sgRNA targeting intron 3 of *SLX1* #1	This study
P16	pX458-mRuby2:sg*SLX1*-int3-2	Mammalian	U6 (ORF1), CMV enhancer/chicken β-actin	sg*SLX1*-int3-2 [TAGAGCCATCCATTCCTCGG]	3xFLAG-SpCas9-T2A-mRuby2		pX458 expressing sgRNA targeting intron 3 of *SLX1* #2	This study

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P17	pX458:sg*ERCC4*-1	Mammalian	U6 (ORF1), CMV enhancer/chicken β-actin	sg*ERCC4*-ex1fw [CAACACGCAGC-CGGCCGAGG]	3xFLAG-SpCas9-T2A-EGFP		pX458 expressing sgRNA targeting exon 1 of *ERCC4* + strand	Gift of Dr. Philip Hublitz
P18	pX458-mRuby2:sg*ERCC4*-2	Mammalian	U6 (ORF1), CMV enhancer/chicken β-actin	sg*ERCC4*-ex1rv [GTACTCCAGCAGCG-GCGCCA]	3xFLAG-SpCas9-T2A-mRuby2		pX458 expressing sgRNA targeting exon 1 of *ERCC4* - strand	Gift of Dr. Philip Hublitz
P19	pTwistCMV-Puro:hME	Mammalian	CMV, UbC (ORF2)	1. hMUS81-L^3-T2A 2. L^1-HA-L^2-HaloTag-L^3-t-hEME1	PuroR		pTwistCMV-Puro expressing human MUS81 and EME1 for transient expression in mammalian cells	Twist Bioscience

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P20	pCDNA3.1-Hygro:hSS	Mammalian	CMV; IRES2 between ORF1-2, SV40 (ORF3)	L^1-SF-L^6-t-hSLX4	L^8-hSLX1-L^4-t-EGFP	HygroR	pCDNA3.1-Hygro expressing human SLX4 and SLX1 for transient expression in mammalian cells	GenScript
P21	pCDNA3.1-Zeo:hXE	Mammalian	CMV; IRES2 between ORF1-2	1. hXPF-L^7-P2A 2. L^5-hERCC1-L^8-3c-L^1-AviTag-L^2-t-L^7-His_6	L^7-BirA-L^9-TagBFP	BleoR	pCDNA3.1-Zeo expressing human XPF and ERCC1 for transient expression in mammalian cells	GenScript

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P22	pTLCV2:MCS	Mammalian	tight TRE promoter; IRES2 between ORF1-2, EF-1α core (ORF2)		1. PuroR-L^3-P2A 2. rtTA-Advanced		pTLCV2 backbone plasmid, no fluorescent marker	Gift of Prof. Yin Dong
P23	pTLCV2:MCS-TagBFP	Mammalian	tight TRE promoter; IRES2 between ORF1-2, EF-1α core (ORF2)		TagBFP		pTLCV2 backbone plasmid, TagBFP after IRES2	Gift of Prof. Yin Dong
P24	pTLCV2:MCS-mCherry	Mammalian	tight TRE promoter; IRES2 between ORF1-2, EF-1α core (ORF2)		mCherry		pTLCV2 backbone plasmid, mCherry after IRES2	Gift of Prof. Yin Dong

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P25	pTLCV2:MCS-EGFP	Mammalian	tight TRE promoter; IRES2 between ORF1-2, EF-1α core (ORF2)	EGFP		as above for pTLCV2:MCS	pTLCV2 backbone plasmid, EGFP after IRES2	Gift of Prof. Yin Dong
P26	pTLCV2:MCS-mOrange2	Mammalian	tight TRE promoter; IRES2 between ORF1-2, EF-1α core (ORF2)	mOrange2		as above for pTLCV2:MCS	pTLCV2 backbone plasmid, mOrange2 after IRES2	This study
P27	pTLCV2:MCS-BirA-TagBFP	Mammalian	tight TRE promoter; IRES2 between ORF1-2, EF-1α core (ORF2)	L^7-BirA-L^9-TagBFP		as above for pTLCV2:MCS	pTLCV2 backbone plasmid, BirA-TagBFP after IRES2	This study

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P28	pTLCV2:hXE,BirA-TagBFP	Mammalian	tight TRE promoter; IRES2 between ORF1-2, EF-1α core (ORF2)	1. hXPF-L^7-P2A 2. L^5-hERCC1-L^8-3c-L^1-AviTag-L^2-t-L^7-His_6	L^7-BirA-L^9-TagBFP	as above for pTLCV2:MCS	pTLCV2 lentiviral vector for doxycycline-inducible human XPF and ERCC1 expression in mammalian cells, BirA-TagBFP after IRES2	This study
P29	pTLCV2:hSS-EGFP	Mammalian	tight TRE promoter; IRES2 between ORF1-2, EF-1α core (ORF2)	L^1-SF-L^6-t-hSLX4	L^8-hSLX1-L^4-t-EGFP	as above for pTLCV2:MCS	pTLCV2 lentiviral vector for doxycycline-inducible human SLX4 and SLX1 expression in mammalian cells	This study

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P30	pTLCV2:hME,mCherry	Mammalian	tight TRE promoter; IRES2 between ORF1-2, EF-1α core (ORF2)	1. hMUS81-L^3-T2A 2. L^1-HA-L^2-HaloTag-L^3-t-hEME1	L^7-mCherry	as above for pTLCV2:MCS	pTLCV2 lentiviral vector for doxycycline-inducible human MUS81 and EME1 expression in mammalian cells	This study
P31	pTLCV2:SLX4IP-mOrange2	Mammalian	tight TRE promoter; IRES2 between ORF1-2, EF-1α core (ORF2)	L^2-His_6-L^4-SUMO-L^8-hSLX4IP	L^7-mOrange2	as above for pTLCV2:MCS	pTLCV2 lentiviral vector for doxycycline-inducible human SLX4IP expression in mammalian cells	This study

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P32	pAceBac1:mSX^{D706A}E	Insect	polyhedrin	1. TEVP-t 2. L^{11}-mSLX4-L^{8}-t 3. L^{10}-mXPF^{D706A}-L^{8}-t 4. L^{1}-SF-L^{8}-mERCC1			Mouse SLX4Δ-XPF^{D706A}-ERCC1 expression as polyprotein with TEV protease (TEVP)	This study
P33	pAceBac1:mSΔX^{D706A}E	Insect	polyhedrin	1. TEVP-t 2. L^{11}-mSLX4$^{\Delta 759-1252}$-L^{8}-t 3. L^{10}-mXPF^{D706A}-L^{8}-t 4. L^{1}-SF-L^{8}-mERCC1			Mouse SLX4Δ-XPF^{D706A}-ERCC1 expression as polyprotein with TEV protease (TEVP)	This study

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P34	pAceBac1:mEM^{D307A}S	Insect	polyhedrin	1. TEVP-t 2. L^{11}-mEME1-L^{5}-t 3. L^{8}-mMUS81^{D307A}-L^{5}-t 4. L^{7}-mSLX1-L^{3}-His$_{10}$			Mouse EME1-MUS81^{D307A}-SLX1 expression as polyprotein with TEV protease (TEVP)	This study
P35	pAceBac1: mEMS^{R38A,E79A}	Insect	polyhedrin	1. TEVP-t 2. L^{11}-mEME1-L^{5}-t 3. L^{8}-mMUS81-L^{5}-t 4. L^{7}-mSLX1^{R38A,E79A}-L^{3}-His$_{10}$			Mouse EME1-MUS81-SLX1^{R38A,E79A} expression as polyprotein with TEV protease (TEVP)	This study

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P36	pAceBac1: $mEM^{D307A}S^{R38A,E79A}$	Insect	polyhedrin	1. TEVP-t 2. L^{11}-mEME1-L^{5}-t 3. L^{8}-$mMUS81^{D307A}$-L^{5}-t 4. L^{7}-$mSLX1^{R38A,E79A}$-L^{3}-His_{10}			Mouse EME1-$MUS81^{D307A}$-$SLX1^{R38A,E79A}$ expression as polyprotein with TEV protease (TEVP)	This study
P37	pET28a:GFP-Nb	Bacterial	T7	L^{3}-GFP-Nb-L^{4}-His_{6}			pET28a expression of GFP nanobody	This study
P38	pET28a:mCherry-Nb	Bacterial	T7	L^{3}-mCherry-Nb-L^{3}-His_{6}			pET28a expression of mCherry nanobody	This study
P39	pMA:*SLX1*donor1	Mammalian	T7 (not in use)	*SLX1*LHdonor-EGFP-L^{5}-P2A-L^{4}-FKBP-L^{7}-Twin-Strep-*SLX1*RHdonor			Plasmid donor for *SLX1* FKBP knock-in #1	Invitrogen

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P40	pCDNA6.2-V5-PL-DEST:*SLX1*donor2	Mammalian	T7 (not in use)	*SLX1*LHdonor-BleoR-L^3-P2A-EGFP-L^2-T2A-L^4-FKBP-L^5-SF-t-*SLX1*RHdonor			Plasmid donor for *SLX1* FKBP knock-in #2	This study
P41	pCI-Neo:*SLX1*donor3	Mammalian		*SLX1*LHdonor-BleoR-L^5-EGFP-L^2-P2A-L^2-HA-HA-L^3-HiBiT-L^2-FKBP-L^8-*SLX1*RHdonor			Plasmid donor for *SLX1* FKBP knock-in #3	GenScript
P42	pcDNA3.1-Zeo:*SLX1*donor4	Mammalian	CMV (in opposite direction to reading frame)	*SLX1*LHdonor-HA-HA-L^4-3xFLAG-L^5-HiBiT-HaloTag-L^4-t-L^2-*SLX1*Rhdonor			Plasmid donor for *SLX1* HaloTag knock-in #4	Invitrogen
P43	pCDNA5-FRTTO:EGFP-hSLX4	Mammalian	CMV	EGFP-L^4-hSLX4			Plasmid for transient overexpression of EGFP-hSLX4	Gift of Dr. Kaima Tsukada

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P44	pCDNA5-FRTTO:EGFP-hSLX4(D1)	Mammalian	CMV	EGFP-L^4-hSLX4$^{\Delta 2-559}$			Plasmid for transient overex-pression of EGFP-hSLX4(D1)	Gift of Dr. Kaima Tsukada
P45	pCDNA5-FRTTO:EGFP-hSLX4(D2)	Mammalian	CMV	EGFP-L^4-hSLX4$^{\Delta 560-1092}$			Plasmid for transient overex-pression of EGFP-hSLX4(D2)	Gift of Dr. Kaima Tsukada
P46	pCDNA5-FRTTO:EGFP-hSLX4(D3)	Mammalian	CMV	EGFP-L^4-hSLX4$^{\Delta 1093-1488}$			Plasmid for transient overex-pression of EGFP-hSLX4(D3)	Gift of Dr. Kaima Tsukada
P47	pCDNA5-FRTTO:EGFP-hSLX4(D4)	Mammalian	CMV	EGFP-L^4-hSLX4$^{\Delta 1489-1834}$			Plasmid for transient overex-pression of EGFP-hSLX4(D4)	Gift of Dr. Kaima Tsukada

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P48	pCDNA5-FRTTO:EGFP-hSLX4(ΔSAP)	Mammalian	CMV	EGFP-L^4-hSLX4$^{\Delta 1538-1618}$			Plasmid for transient overexpression of EGFP-hSLX4($^{\Delta SAP}$)	Gift of Dr. Kaima Tsukada
P49	pCDNA5-FRTTO:EGFP-hSLX4$^{\Delta SBD}$	Mammalian	CMV	EGFP-L^4-hSLX4$^{\Delta 1733-1820}$			Plasmid for transient overexpression of EGFP-hSLX4(ΔSBD)	Gift of Dr. Kaima Tsukada
P50	pTLCV2:EGFP-hSLX4	Mammalian	tight TRE promoter; IRES2 between ORF1-2, EF-1α core (ORF2)	EGFP-L^5-hSLX4$^{\Delta 1733-1820}$	as above for pTLCV2:MCS		pTLCV2 lentiviral vector for doxycycline-inducible human SLX4 expression in mammalian cells	Gift of Dr. Kaima Tsukada

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P51	pLenti-tetON:KRAB-dCas9-DHFR	Mammalian	tight TRE promoter (ORF1), EF-1α core (ORF2)	L^{10}-KRAB-L^{7}-dCas9-L^{1}-HA-L^{6}-SV40NLS-L^{2}-SV40NLS-L^{7}-eDHFR	TagRFP-L^{2}-T2A-Tet-ON 3G		Lentiviral for doxycycline-inducible expression of KRAB-dCas9 (CRISPRi), also controlled by degron from *E. coli* DHFR which can be stabilised by trimetho-prim	Addgene: 167935, ref[763]
P52	pX330a:dCas9-LSD1	Mammalian	CMV, chicken β-actin (ORF1)	L^{1}-SV40NLS-L^{3}-Cas9m4-L^{4}-SV40NLS-L^{7}-hLSD1			Backbone plasmid for subcloning of human LSD1 sequence	Addgene: 92362, ref[764]

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P53	pLenti-tetON:ADH1B-dCas9-DHFR	Mammalian	tight TRE promoter (ORF1), EF-1α core (ORF2)	L^{10}-hADH1B-L^{7}-dCas9-L^{1}-HA-L^{6}-SV40NLS-L^{2}-SV40NLS-L^{7}-eDHFR	TagRFP-L^{2}-T2A-Tet-ON 3G		Lentiviral for doxycycline-inducible expression of ADH1B-dCas9 (CRISPRf), also controlled by degron from *E. coli* DHFR which can be stabilised by trimetho-prim	This study

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P54	pLenti-tetON:ADH1G-dCas9-DHFR	Mammalian	tight TRE promoter (ORF1), EF-1α core (ORF2)	L^{10}-hADH1G-L^{7}-dCas9-L^{1}-HA-L^{6}-SV40NLS-L^{2}-SV40NLS-L^{7}-eDHFR	TagRFP-L^{2}-T2A-Tet-ON 3G		Lentiviral for doxycycline-inducible expression of ADH1G-dCas9 (CRISPRf), also controlled by degron from *E. coli* DHFR which can be stabilised by trimetho-prim	This study

Table A.1 continued: Details of plasmids used.

ID	Name	System	Promoter	ORF1	ORF2	ORF3	Notes	Source
P55	pLenti-tetON:LSD1-dCas9-DHFR	Mammalian	tight TRE promoter (ORF1), EF-1α core (ORF2)	L^{10}-hLSD1-L^{7}-dCas9-L^{1}-HA-L^{6}-SV40NLS-L^{2}-SV40NLS-L^{7}-eDHFR	TagRFP-L^{2}-T2A-Tet-ON 3G		Lentiviral for doxycycline-inducible expression of LSD1-dCas9 (CRISPRf), also controlled by degron from *E. coli* DHFR which can be stabilised by trimetho-prim	This study

B | Additional Figures

B.1 | Generation of CRISPRf (LSD1-dCas9/ADH1G-dCas9) by lentiviral transduction-FACS

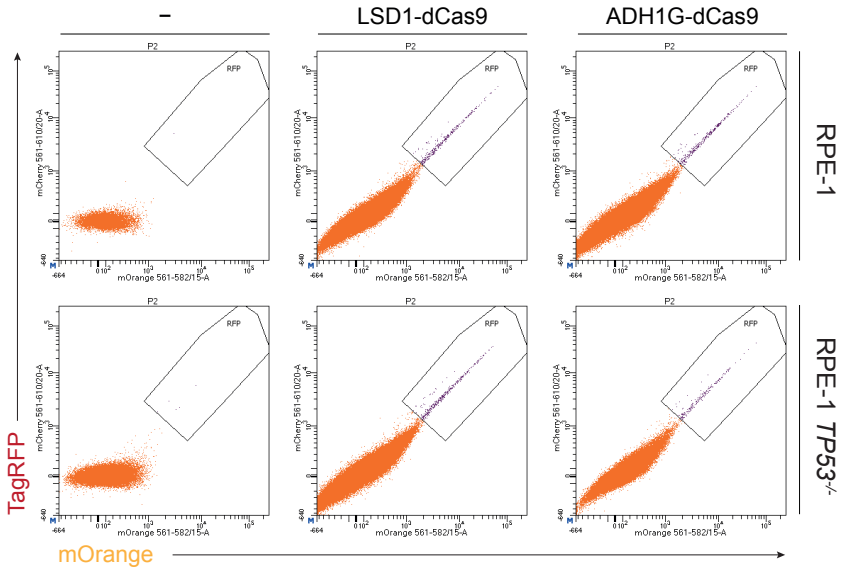


Figure B.1 | Single-cell sorting of LSD1-dCas9 and ADH1G-dCas9-transduced cells. Cells were transduced with lentivirus expressing ADH1G-dCas9 or LSD1-dCas9, sorted by tagRFP.

B.2 | Generation of SLX1 knockouts by transfection of RPE-1, RPE-1 *TP53*$^{-/-}$, HAP1 and 293FT cells with sg*SLX1*

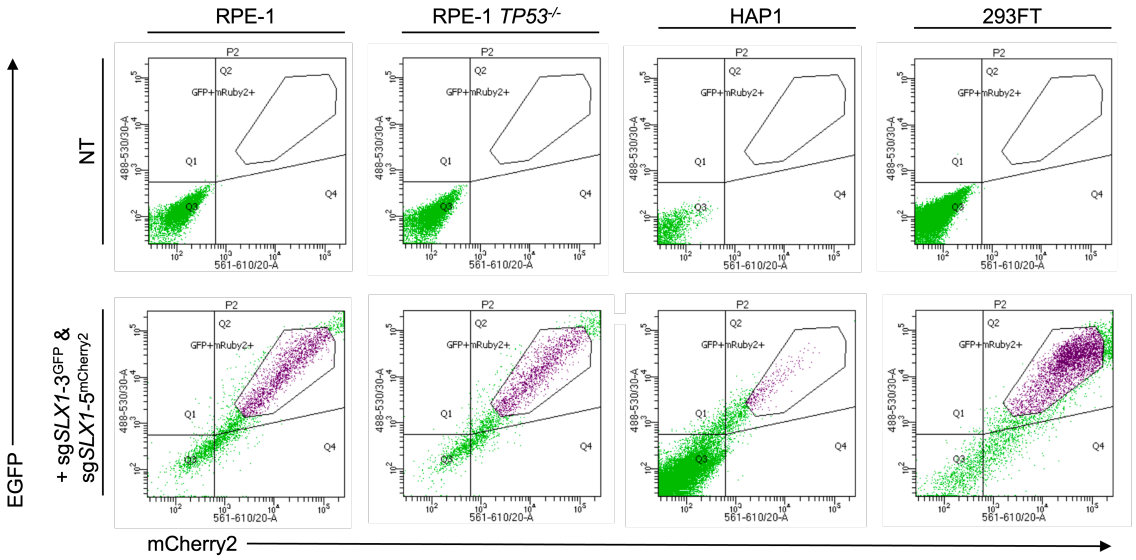


Figure B.2 | Single-cell sorting by FACS of RPE-1, RPE-1 *TP53*$^{-/-}$, HAP1 and 293FT for SLX1 knockout generation. Cells transfected with sg*SLX1*-3 (GFP) and sg*SLX1*-5 (mRuby2) were single-cell sorted for double-positive GFP^{+} and mRuby2^{+} cells by FACS.

B.3 | Generation of GFP-SLX1 complemented (and GFP-EV) transduced populations in 293FT Δ*SLX1* 1 cells.

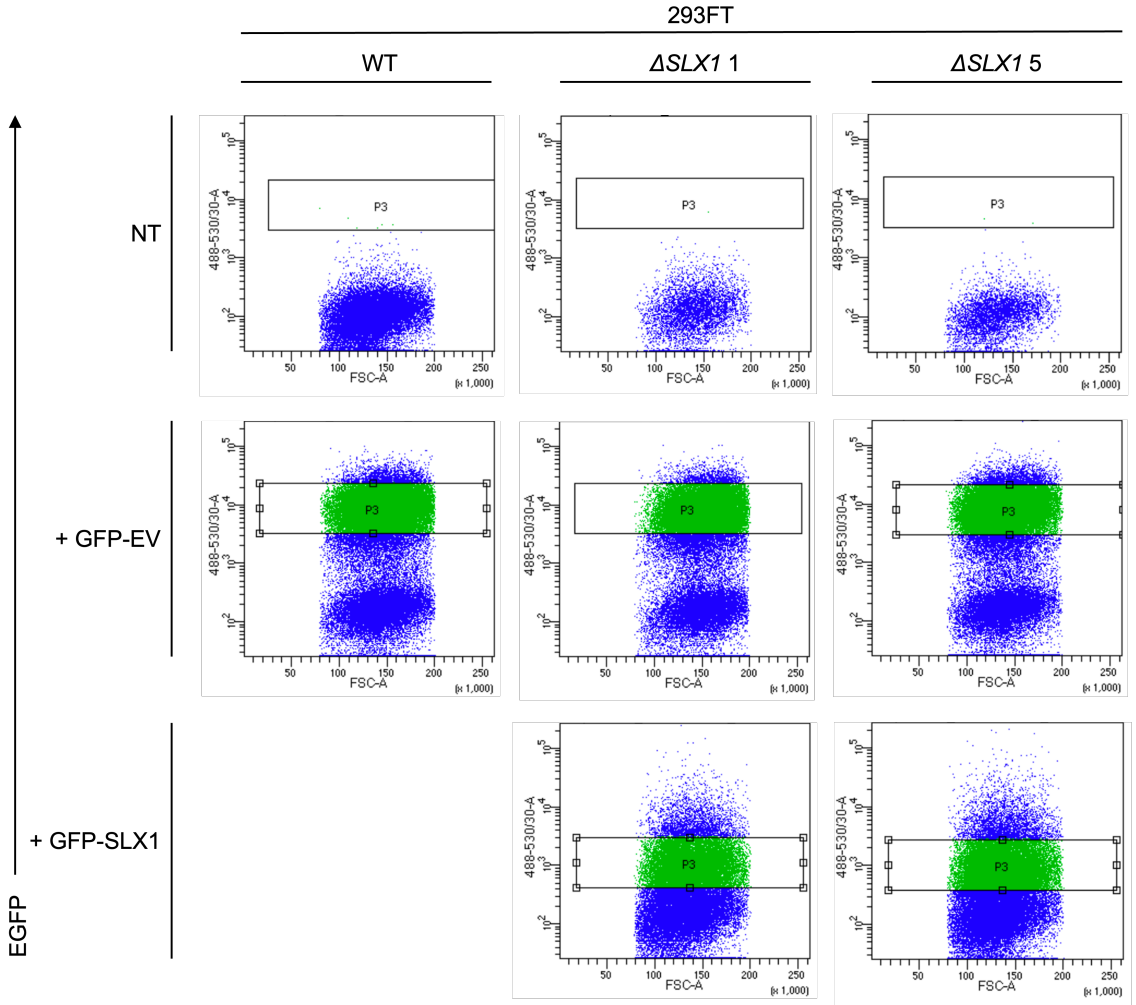


Figure B.3 | Heterogenous population sorting by FACS of 293FT Δ*SLX1* 1 for GFP-SLX1 complementation. Cells were transduced with lentivirus for the Dox-inducible expression of GFP-EV and GFP-SLX1 and population-sorted by FACS for GFP expression after overnight treatment with 2 µg/ml of Dox.

B.4 | Generation of GFP-SLX1 complemented (and GFP-EV) transduced populations in HAP1 Δ*SLX1* 37 cells.

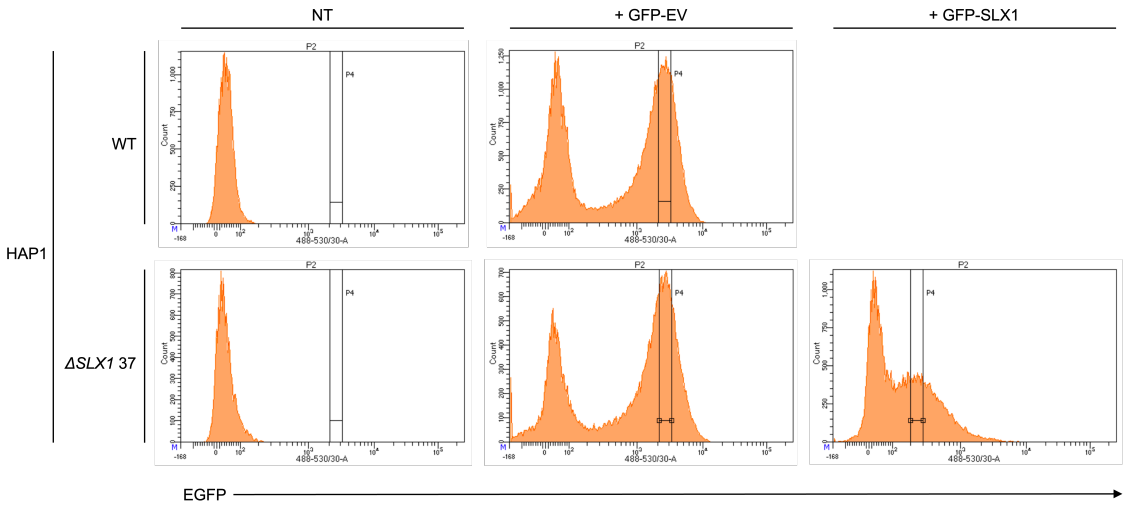


Figure B.4 | Heterogenous population sorting by FACS of HAP1 Δ*SLX1* 37 for GFP-SLX1 complementation. Cells were transduced with lentivirus for the Dox-inducible expression of GFP-EV and GFP-SLX1 and population-sorted by FACS for GFP expression after overnight treatment with 2 µg/ml of Dox.

B.5 | FACS of RPE-1 cells for SLX1 knock-in.

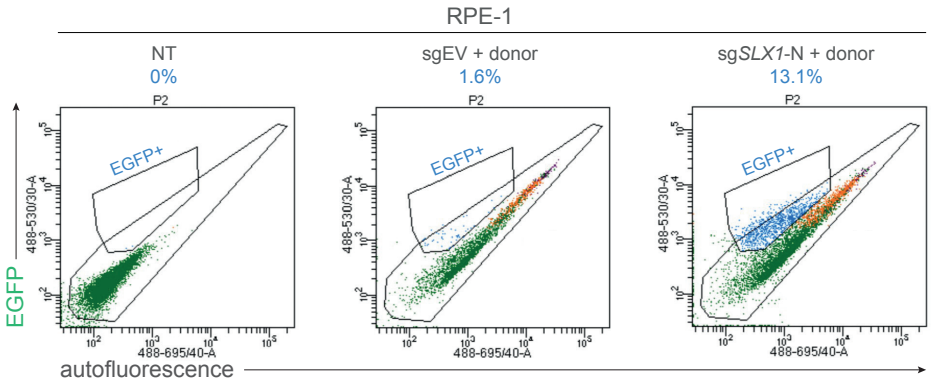


Figure B.5 | Single-cell sorting by FACS of RPE-1 cells for SLX1 knock-in generation. Cells were transfected with sg*SLX1*-N and donor plasmid pCDNA6.2-V5-PL-DEST:*SLX1*donor2, before gradual zeocin selection and later single-cell sorted for GFP^+ cells by FACS (FACS in this case collected the off-diagonal GFP^+ cells, that are interpreted as minimally fluorescent above any background fluorescence).

B.6 | Human and *Xenopus* SLX4 sequence alignment

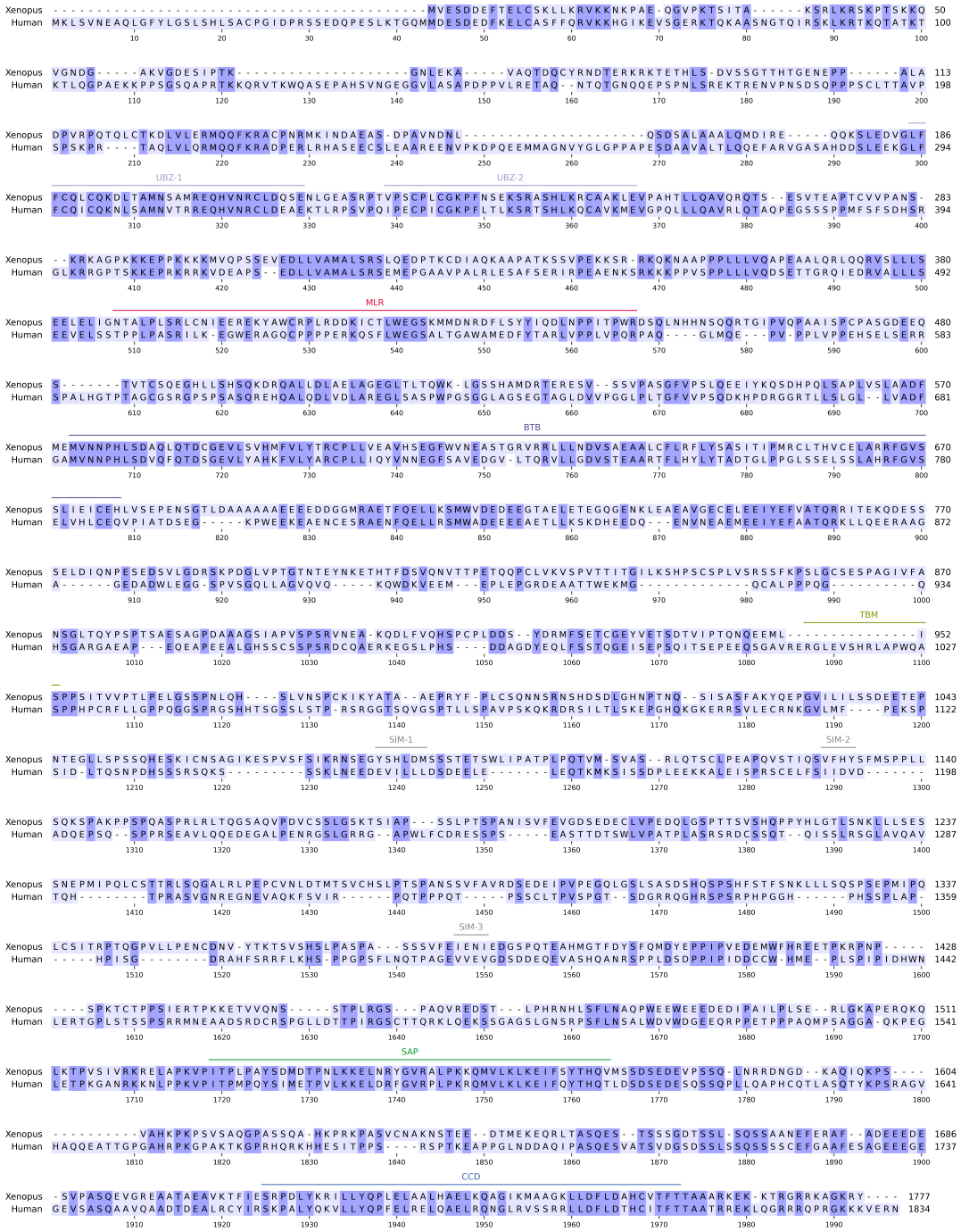


Figure B.6 | Human-*Xenopus* SLX4 sequence alignment. Sequence alignment of human and *Xenopus* SLX4 orthologues. Sequences were aligned using the parasail pairwise semi-global alignment tool (`https://github.com/jeffdaily/parasail`) and visualised using the pyMSAviz 0.5.0 package (`https://github.com/jeffdaily/parasail`). Domain boundaries corresponding to the human sequence are highlighted.

B.7 | PEI transfection efficiency

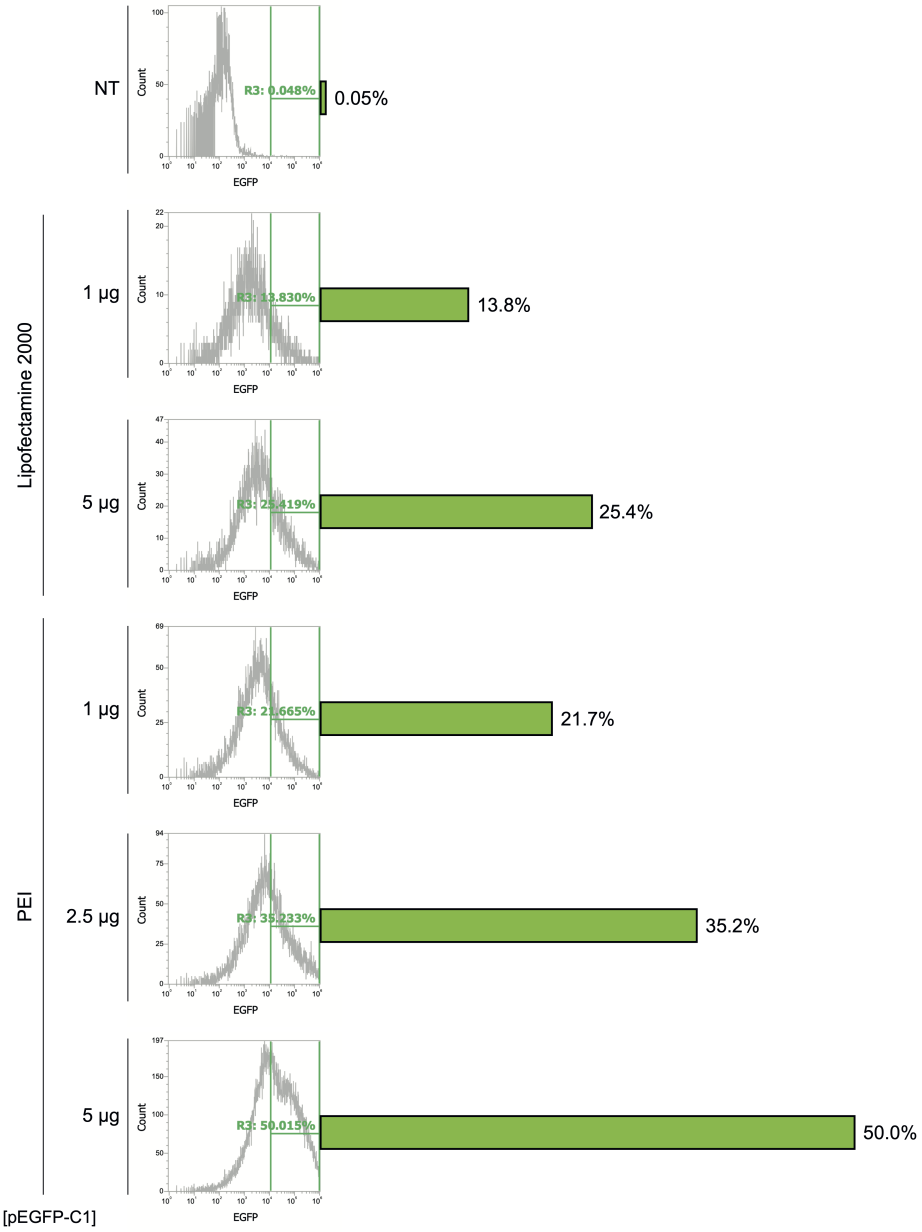


Figure B.7 | Comparison of transfection efficiency between lipofectamine and PEI. Expi293F cells were transfected with pEGFP-C1 using lipofectamine 2000 or polyethylenimine (PEI) and harvested next day for analysis of EGFP expression by flow cytometry.

References

1. Alcón, P. *et al.* FANCD2-FANCI is a clamp stabilized on DNA by monoubiquitination of FANCD2 during DNA repair. *Nature Structural & Molecular Biology* **27,** 240–248 (2020).
2. Rennie, M. L. *et al.* Differential functions of FANCI and FANCD2 ubiquitination stabilize ID2 complex on DNA. *EMBO reports* **21,** e50133 (2020).
3. Shakeel, S. *et al.* Structure of the Fanconi anaemia monoubiquitin ligase complex. *Nature* **575,** 234–237 (2019).
4. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).
5. Shah, S. *et al.* Assessment of SLX4 Mutations in Hereditary Breast Cancers. *PLoS ONE* **8,** e66961 (2013).
6. Gohil, D. & Roy, R. Beyond Nucleotide Excision Repair: The Importance of XPF in Base Excision Repair and Its Impact on Cancer, Inflammation, and Aging. *International Journal of Molecular Sciences* **25,** 13616 (2024).
7. Kim, Y. *et al.* Mutations of the SLX4 gene in Fanconi anemia. *Nature genetics* **43,** 142–146 (2011).
8. Deans, A. J. & West, S. C. DNA interstrand crosslink repair and cancer. *Nature reviews. Cancer* **11,** 467–480 (2011).
9. Dingler, F. A. *et al.* Two Aldehyde Clearance Systems Are Essential to Prevent Lethal Formaldehyde Accumulation in Mice and Humans. *Molecular Cell* **80,** 996–1012.e9 (2020).
10. Garaycoechea, J. I. & Patel, K. J. Why does the bone marrow fail in Fanconi anemia? *Blood* **123,** 26–34 (2014).
11. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171,** 737–738 (1953).
12. Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **168,** 644–656 (2017).
13. Grillari, J., Katinger, H. & Voglauer, R. Contributions of DNA interstrand cross-links to aging of cells and organisms. *Nucleic Acids Research* **35,** 7566–7576 (2007).
14. Mehta, P. A. & Ebens, C. in *GeneReviews®* (eds Adam, M. P. *et al.*) (University of Washington, Seattle, Seattle (WA), 1993).
15. McCulloch, S. D. & Kunkel, T. A. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Research* **18.** Publisher: Nature Publishing Group, 148–161 (2008).
16. Ciccia, A. & Elledge, S. J. The DNA Damage Response: Making It Safe to Play with Knives. *Molecular Cell* **40.** Publisher: Elsevier, 179–204 (2010).

17. Voitkun, V. & Zhitkovich, A. Analysis of DNA-protein crosslinking activity of malondialdehyde in vitro. *Mutation Research* **424,** 97–106 (1999).

18. Rastogi, R. P., Richa, Kumar, A., Tyagi, M. B. & Sinha, R. P. Molecular Mechanisms of Ultraviolet Radiation-Induced DNA Damage and Repair. *Journal of Nucleic Acids* **2010,** 592980 (2010).

19. McKenzie, R. L., Björn, L. O., Bais, A. & Ilyasd, M. Changes in biologically active ultraviolet radiation reaching the Earth's surface. *Photochemical & Photobiological Sciences* **2.** Publisher: The Royal Society of Chemistry, 5–15 (2003).

20. Huss, V. A. R., Festl, H. & Schleifer, K. H. Studies on the spectrophotometric determination of DNA hybridization from renaturation rates. *Systematic and Applied Microbiology* **4,** 184–192 (1983).

21. Whitmore, S. E., Potten, C. S., Chadwick, C. A., Strickland, P. T. & Morison, W. L. Effect of photoreactivating light on UV radiation-induced alterations in human skin. *Photodermatology, Photoimmunology & Photomedicine* **17,** 213–217 (2001).

22. Harm, H. & Rupert, C. S. Analysis of photoenzymatic repair of UV lesions in DNA by single light flashes: I. In vitro studies with haemophilus influenzae transforming DNA and yeast photoreactivating enzyme. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **6,** 355–370 (1968).

23. Nejedlý, K., Kittner, R., Pospísilová, S. & Kypr, J. Crosslinking of the complementary strands of DNA by UV light: dependence on the oligonucleotide composition of the UV irradiated DNA. *Biochimica Et Biophysica Acta* **1517,** 365–375 (2001).

24. Sevilla, M. D., Becker, D., Kumar, A. & Adhikary, A. Gamma and Ion-Beam Irradiation of DNA: Free Radical Mechanisms, Electron Effects, and Radiation Chemical Track Structure. *Radiation physics and chemistry (Oxford, England : 1993)* **128,** 60–74 (2016).

25. Belli, M. & Tabocchini, M. A. Ionizing Radiation-Induced Epigenetic Modifications and Their Relevance to Radiation Protection. *International Journal of Molecular Sciences* **21,** 5993 (2020).

26. *Sources and effects of ionizing radiation: United Nations Scientific Committee on the Effects of Atomic Radiation: UNSCEAR 2008 report to the General Assembly, with scientific annexes* (ed United Nations) OCLC: ocn680422162 (United Nations, New York, 2010). 2 pp.

27. Schauer, D. A. & Linton, O. W. NCRP REPORT NO. 160, IONIZING RADIATION EXPOSURE OF THE POPULATION OF THE UNITED STATES, MEDICAL EXPOSURE—ARE WE DOING LESS WITH MORE, AND IS THERE A ROLE FOR HEALTH PHYSICISTS? *Health Physics* **97,** 1 (2009).

28. Porstendörfer, J. Properties and behaviour of radon and thoron and their decay products in the air. *Journal of Aerosol Science* **25,** 219–263 (1994).

29. Hutchinson, F. in *Progress in Nucleic Acid Research and Molecular Biology* (eds Cohn, W. E. & Moldave, K.) 115–154 (Academic Press, 1985).

30. Eccles, L. J., O'Neill, P. & Lomax, M. E. Delayed repair of radiation induced clustered DNA damage: Friend or foe? *Mutation Research* **711,** 134–141 (2011).

31. Hada, M. & Georgakilas, A. G. Formation of clustered DNA damage after high-LET irradiation: a review. *Journal of Radiation Research* **49,** 203–210 (2008).

32. Ali, Y. F., Cucinotta, F. A., Ning-Ang, L. & Zhou, G. Cancer Risk of Low Dose Ionizing Radiation. *Frontiers in Physics* **8.** Publisher: Frontiers (2020).

33. Coskun, E. *et al.* Aflatoxin-Guanine DNA Adducts and Oxidatively Induced DNA Damage in Aflatoxin-Treated Mice in Vivo as Measured by Liquid Chromatography-Tandem Mass Spectrometry with Isotope Dilution. *Chemical research in toxicology* **32,** 80–89 (2019).

34. Niu, Z.-S., Niu, X.-J. & Wang, W.-H. Genetic alterations in hepatocellular carcinoma: An update. *World Journal of Gastroenterology* **22,** 9069–9095 (2016).

35. Vu, A. T. *et al.* Polycyclic Aromatic Hydrocarbons in the Mainstream Smoke of Popular U.S. Cigarettes. *Chemical research in toxicology* **28,** 1616–1626 (2015).

36. Smith, S. L. War! What is it good for? Mustard gas medicine. *CMAJ : Canadian Medical Association Journal* **189,** E321 (2017).

37. Anand, U. *et al.* Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. *Genes & Diseases* **10,** 1367–1401 (2023).

38. Shin, H.-J. *et al.* Doxorubicin-induced necrosis is mediated by poly-(ADP-ribose) polymerase 1 (PARP1) but is independent of p53. *Scientific Reports* **5.** Publisher: Nature Publishing Group, 15798 (2015).

39. Josting, A. *et al.* Secondary myeloid leukemia and myelodysplastic syndromes in patients treated for Hodgkin's disease: a report from the German Hodgkin's Lymphoma Study Group. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* **21,** 3440–3446 (2003).

40. Wheaton, W. W. & Chandel, N. S. Hypoxia. 2. Hypoxia regulates cellular metabolism. *American Journal of Physiology - Cell Physiology* **300,** C385–C393 (2011).

41. Eleftheriadis, T., Pissas, G., Filippidis, G., Liakopoulos, V. & Stefanidis, I. Reoxygenation induces reactive oxygen species production and ferroptosis in renal tubular epithelial cells by activating aryl hydrocarbon receptor. *Molecular Medicine Reports* **23,** 41 (2021).

42. Turrens, J. F. Mitochondrial formation of reactive oxygen species. *The Journal of Physiology* **552,** 335–344 (2003).

43. Franco, R. & Cidlowski, J. A. Apoptosis and glutathione: beyond an antioxidant. *Cell Death & Differentiation* **16,** 1303–1314 (2009).

44. Hahm, J. Y., Park, J., Jang, E.-S. & Chi, S. W. 8-Oxoguanine: from oxidative damage to epigenetic and epitranscriptional modification. *Experimental & Molecular Medicine* **54.** Publisher: Nature Publishing Group, 1626–1642 (2022).

45. Cheng, K. C., Cahill, D. S., Kasai, H., Nishimura, S. & Loeb, L. A. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G—-T and A—-C substitutions. *The Journal of Biological Chemistry* **267,** 166–172 (1992).

46. Hsiao, Y.-C. *et al.* DNA Damage and the Gut Microbiome: From Mechanisms to Disease Outcomes. *DNA* **3,** 13–32 (2023).

47. Pleguezuelos-Manzano, C. *et al.* Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. *Nature* **580,** 269–273 (2020).

48. Pöltl, L. *et al.* Microbiota-derived genotoxin tilimycin generates colonic stem cell mutations. *Cell Reports* **42.** Publisher: Elsevier (2023).

49. Xue, M. *et al.* Structure elucidation of colibactin and its DNA cross-links. *Science* **365.** Publisher: American Association for the Advancement of Science, eaax2685 (2019).

50. Jiang, Y. *et al.* Reactivity of an Unusual Amidase May Explain Colibactin's DNA Cross-Linking Activity. *Journal of the American Chemical Society* **141,** 11489–11496 (2019).

51. Volpe, M. R. *et al.* A small molecule inhibitor prevents gut bacterial genotoxin production. *Nature Chemical Biology* **19.** Publisher: Nature Publishing Group, 159–167 (2023).

52. Cao, Y. *et al.* Commensal microbiota from patients with inflammatory bowel disease produce genotoxic metabolites. *Science (New York, N.Y.)* **378,** eabm3233 (2022).

53. Liao, Y. *et al.* Cordycepin induces cell cycle arrest and apoptosis by inducing DNA damage and up-regulation of p53 in Leukemia cells. *Cell Cycle* **14,** 761–771 (2015).

54. Glover, T. W., Wilson, T. E. & Arlt, M. F. Fragile sites in cancer: more than meets the eye. *Nature Reviews. Cancer* **17,** 489–501 (2017).

55. Le Tallec, B. *et al.* Updating the mechanisms of common fragile site instability: how to reconcile the different views? *Cellular and molecular life sciences: CMLS* **71,** 4489–4494 (2014).

56. Kooistra, S. M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nature Reviews Molecular Cell Biology* **13.** Publisher: Nature Publishing Group, 297–311 (2012).

57. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nature Reviews Genetics* **7.** Publisher: Nature Publishing Group, 715–727 (2006).

58. Mulderrig, L. *et al.* Aldehyde-driven transcriptional stress triggers an anorexic DNA damage response. *Nature* **600,** 158–163 (2021).

59. Shen, X. *et al.* A Surge of DNA Damage Links Transcriptional Reprogramming and Hematopoietic Deficit in Fanconi Anemia. *Molecular Cell* **80,** 1013–1024.e6 (2020).

60. Zeman, M. K. & Cimprich, K. A. Causes and Consequences of Replication Stress. *Nature cell biology* **16,** 2–9 (2014).

61. Saxena, S. & Zou, L. Hallmarks of DNA replication stress. *Molecular Cell* **82,** 2298–2314 (2022).

62. Parsels, L. A. *et al.* PARP1 Trapping and DNA Replication Stress Enhance Radiosensitization with Combined WEE1 and PARP Inhibitors. *Molecular cancer research : MCR* **16,** 222–232 (2018).

63. Westhorpe, R., Roske, J. J. & Yeeles, J. T. P. Mechanisms controlling replication fork stalling and collapse at topoisomerase 1 cleavage complexes. *Molecular Cell* **84,** 3469–3481.e7 (2024).

64. Pommier, Y., Nussenzweig, A., Takeda, S. & Austin, C. Human topoisomerases and their roles in genome stability and organization. *Nature Reviews Molecular Cell Biology* **23.** Publisher: Nature Publishing Group, 407–427 (2022).

65. Moon, J., Han, J. H., Kim, D. Y., Jung, M.-j. & Kim, S. K. Effects of deficient of the Hoogsteen base-pairs on the G-quadruplex stabilization and binding mode of a cationic porphyrin. *Biochemistry and Biophysics Reports* **2,** 29–35 (2015).

66. García-Muse, T. & Aguilera, A. Transcription–replication conflicts: how they occur and how they are resolved. *Nature Reviews Molecular Cell Biology* **17.** Publisher: Nature Publishing Group, 553–563 (2016).

67. Lalonde, M., Trauner, M., Werner, M. & Hamperl, S. Consequences and Resolution of Transcription–Replication Conflicts. *Life* **11,** 637 (2021).

68. Petermann, E., Orta, M. L., Issaeva, N., Schultz, N. & Helleday, T. Hydroxyurea-Stalled Replication Forks Become Progressively Inactivated and Require Two Different RAD51-Mediated Pathways for Restart and Repair. *Molecular Cell* **37,** 492–502 (2010).

69. Musiałek, M. W. & Rybaczek, D. Hydroxyurea—The Good, the Bad and the Ugly. *Genes* **12,** 1096 (2021).

70. Bucknall, R. A., Moores, H., Simms, R. & Hesp, B. Antiviral Effects of Aphidicolin, a New Antibiotic Produced by Cephalosporium aphidicola. *Antimicrobial Agents and Chemotherapy* **4,** 294–298 (1973).

71. Slater, M. L. Effect of Reversible Inhibition of Deoxyribonucleic Acid Synthesis on the Yeast Cell Cycle. *Journal of Bacteriology* **113.** Publisher: American Society for Microbiology, 263–270 (1973).

72. Krokan, H., Wist, E. & Krokan, R. H. Aphidicolin inhibits DNA synthesis by DNA polymerase alpha and isolated nuclei by a similar mechanism. *Nucleic Acids Research* **9,** 4709–4719 (1981).

73. Wold, M. S. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annual Review of Biochemistry* **66,** 61–92 (1997).

74. Simoneau, A. & Zou, L. An extending ATR-CHK1 circuitry: the replication stress response and beyond. *Current Opinion in Genetics & Development* **71,** 92–98 (2021).

75. Rundle, S., Bradbury, A., Drew, Y. & Curtin, N. J. Targeting the ATR-CHK1 Axis in Cancer Therapy. *Cancers* **9.** Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, 41 (2017).

76. Buisson, R., Boisvert, J. L., Benes, C. H. & Zou, L. Distinct but Concerted Roles of ATR, DNA-PK, and Chk1 in Countering Replication Stress during S Phase. *Molecular Cell* **59,** 1011–1024 (2015).

77. Toledo, L., Neelsen, K. J. & Lukas, J. Replication Catastrophe: When a Checkpoint Fails because of Exhaustion. *Molecular Cell* **66,** 735–749 (2017).

78. Toledo, L. I. *et al.* ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *Cell* **155,** 1088–1103 (2013).

79. Conti, B. A. & Smogorzewska, A. Mechanisms of direct replication restart at stressed replisomes. *DNA repair* **95,** 102947 (2020).

80. Petermann, E. & Helleday, T. Pathways of mammalian replication fork restart. *Nature Reviews Molecular Cell Biology* **11.** Publisher: Nature Publishing Group, 683–687 (2010).

81. Bandyopadhyay, D. & Mishra, P. P. Revealing the DNA Unwinding Activity and Mechanism of Fork Reversal by RecG While Exposed to Variants of Stalled Replication-fork at Single-Molecular Resolution. *Journal of Molecular Biology* **434,** 167822 (2022).

82. Kolinjivadi, A. M. *et al.* Smarcal1-Mediated Fork Reversal Triggers Mre11-Dependent Degradation of Nascent DNA in the Absence of Brca2 and Stable Rad51 Nucleofilaments. *Molecular Cell* **67,** 867–881.e7 (2017).

83. Lemaçon, D. *et al.* MRE11 and EXO1 nucleases degrade reversed forks and elicit MUS81-dependent fork rescue in BRCA2-deficient cells. *Nature Communications* **8,** 860 (2017).

84. Mijic, S. *et al.* Replication fork reversal triggers fork degradation in BRCA2-defective cells. *Nature Communications* **8,** 859 (2017).

85. Taglialatela, A. *et al.* Restoration of Replication Fork Stability in BRCA1- and BRCA2-Deficient Cells by Inactivation of SNF2-Family Fork Remodelers. *Molecular Cell* **68,** 414–430.e8 (2017).

86. Neelsen, K. J. & Lopes, M. Replication fork reversal in eukaryotes: from dead end to dynamic response. *Nature Reviews. Molecular Cell Biology* **16,** 207–220 (2015).

87. Chappidi, N. *et al.* Fork Cleavage-Religation Cycle and Active Transcription Mediate Replication Restart after Fork Stalling at Co-transcriptional R-Loops. *Molecular Cell* **77,** 528–541.e8 (2020).

88. Di Marco, S. *et al.* RECQ5 Helicase Cooperates with MUS81 Endonuclease in Processing Stalled Replication Forks at Common Fragile Sites during Mitosis. *Molecular Cell* **66,** 658–671.e8 (2017).

89. Vujanovic, M. *et al.* Replication Fork Slowing and Reversal upon DNA Damage Require PCNA Polyubiquitination and ZRANB3 DNA Translocase Activity. *Molecular Cell* **67,** 882–890.e5 (2017).

90. Yates, M. *et al.* SMARCAL1 ubiquitylation controls its association with RPA-coated ssDNA and promotes replication fork stability. *PLoS biology* **22,** e3002552 (2024).

91. Bugreev, D. V., Rossi, M. J. & Mazin, A. V. Cooperation of RAD51 and RAD54 in regression of a model replication fork. *Nucleic Acids Research* **39,** 2153–2164 (2011).

92. Flaus, A., Martin, D. M. A., Barton, G. J. & Owen-Hughes, T. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Research* **34,** 2887–2905 (2006).

93. Yan, Z. *et al.* A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. *Molecular Cell* **37,** 865–878 (2010).

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

95. Hashimoto, Y., Ray Chaudhuri, A., Lopes, M. & Costanzo, V. Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis. *Nature Structural & Molecular Biology* **17.** Publisher: Nature Publishing Group, 1305–1311 (2010).

96. Schlacher, K. *et al.* Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145,** 529–542 (2011).

97. Wassing, I. E. & Esashi, F. RAD51: Beyond the break. *Seminars in Cell & Developmental Biology. Genome stability* **113,** 38–46 (2021).

98. Liberi, G. *et al.* Rad51-dependent DNA structures accumulate at damaged replication forks in sgs1 mutants defective in the yeast ortholog of BLM RecQ helicase. *Genes & Development* **19.** Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab, 339–350 (2005).

99. Kwon, Y. *et al.* DNA binding and RAD51 engagement by the BRCA2 C-terminus orchestrate DNA repair and replication fork preservation. *Nature Communications* **14.** Publisher: Nature Publishing Group, 432 (2023).

100. Esashi, F. *et al.* CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature* **434,** 598–604 (2005).

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

102. Pathania, S. *et al.* BRCA1 haploinsufficiency for replication stress suppression in primary cells. *Nature Communications* **5.** Publisher: Nature Publishing Group, 5496 (2014).

103. Chen, X., Bosques, L., Sung, P. & Kupfer, G. M. A novel role for non-ubiquitinated FANCD2 in response to hydroxyurea-induced DNA damage. *Oncogene* **35.** Publisher: Nature Publishing Group, 22–34 (2016).

104. Taniguchi, T. *et al.* S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. *Blood* **100,** 2414–2420 (2002).

105. Michl, J., Zimmer, J., Buffa, F. M., McDermott, U. & Tarsounas, M. FANCD2 limits replication stress and genome instability in cells lacking BRCA2. *Nature Structural & Molecular Biology* **23,** 755–757 (2016).

106. Kais, Z. *et al.* FANCD2 Maintains Fork Stability in BRCA1/2-Deficient Tumors and Promotes Alternative End-Joining DNA Repair. *Cell Reports* **15,** 2488–2499 (2016).

107. Lossaint, G. *et al.* FANCD2 binds MCM proteins and controls replisome function upon activation of s phase checkpoint signaling. *Molecular Cell* **51,** 678–690 (2013).

108. Tan, W. *et al.* Monoubiquitination by the human Fanconi anemia core complex clamps FANCI:FANCD2 on DNA in filamentous arrays. *eLife* **9,** e54128 (2020).

109. Wang, R., Wang, S., Dhar, A., Peralta, C. & Pavletich, N. P. DNA clamp function of the monoubiquitinated Fanconi anaemia ID complex. *Nature* **580,** 278–282 (2020).

110. Jha, J. K., Ramachandran, R. & Chattoraj, D. K. Opening the Strands of Replication Origins—Still an Open Question. *Frontiers in Molecular Biosciences* **3.** Publisher: Frontiers (2016).

111. Mirkin, E. V. & Mirkin, S. M. Replication fork stalling at natural impediments. *Microbiology and molecular biology reviews: MMBR* **71,** 13–35 (2007).

112. Brambati, A., Colosio, A., Zardoni, L., Galanti, L. & Liberi, G. Replication and transcription on a collision course: eukaryotic regulation mechanisms and implications for DNA stability. *Frontiers in Genetics* **6.** Publisher: Frontiers (2015).

113. Aguilera, A. & García-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Molecular Cell* **46,** 115–124 (2012).

114. Andrs, M. *et al.* Excessive reactive oxygen species induce transcription-dependent replication stress. *Nature Communications* **14.** Publisher: Nature Publishing Group, 1791 (2023).

115. Safari, M. *et al.* R-Loop–Mediated ssDNA Breaks Accumulate Following Short-Term Exposure to the HDAC Inhibitor Romidepsin. *Molecular Cancer Research* **19,** 1361–1374 (2021).

116. Wu, W. *et al.* RTEL1 suppresses G-quadruplex-associated R-loops at difficult-to-replicate loci in the human genome. *Nature Structural & Molecular Biology* **27,** 424–437 (2020).

117. Castillo Bosch, P. *et al.* FANCJ promotes DNA synthesis through G-quadruplex structures. *The EMBO journal* **33,** 2521–2533 (2014).

118. Lee, W. T. C. *et al.* Single-molecule imaging reveals replication fork coupled formation of G-quadruplex structures hinders local replication stress signaling. *Nature Communications* **12.** Publisher: Nature Publishing Group, 2525 (2021).

119. Wu, Y., Shin-ya, K. & Brosh, R. M. FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Molecular and Cellular Biology* **28,** 4116–4128 (2008).

120. Sanchez, A. *et al.* Transcription-replication conflicts as a source of common fragile site instability caused by BMI1-RNF2 deficiency. *PLoS genetics* **16,** e1008524 (2020).

121. Yang, Y. *et al.* Transcription-replication conflicts in primordial germ cells necessitate the Fanconi anemia pathway to safeguard genome stability. *Proceedings of the National Academy of Sciences of the United States of America* **119,** e2203208119 (2022).

122. Wulfridge, P. & Sarma, K. Intertwining roles of R-loops and G-quadruplexes in DNA repair, transcription and genome organization. *Nature Cell Biology* **26.** Publisher: Nature Publishing Group, 1025–1036 (2024).

123. Lang, K. S. & Merrikh, H. Topological stress is responsible for the detrimental outcomes of head-on replication-transcription conflicts. *Cell Reports* **34,** 108797 (2021).

124. Roy, D., Zhang, Z., Lu, Z., Hsieh, C.-L. & Lieber, M. R. Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Molecular and Cellular Biology* **30,** 146–159 (2010).

125. Drolet, M., Bi, X. & Liu, L. F. Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *The Journal of Biological Chemistry* **269,** 2068–2074 (1994).

126. Pommier, Y., Sun, Y., Huang, S.-Y. N. & Nitiss, J. L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nature Reviews. Molecular Cell Biology* **17,** 703–721 (2016).

127. Yasuhara, T. *et al.* RAP80 suppresses the vulnerability of R-loops during DNA double-strand break repair. *Cell Reports* **38,** 110335 (2022).

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

129. Watanabe, G. & Lieber, M. R. Dynamics of the Artemis and DNA-PKcs Complex in the Repair of Double-Strand Breaks. *Journal of Molecular Biology* **434,** 167858 (2022).

130. Sowers, L. C., Shaw, B. R., Veigl, M. L. & Sedwick, W. D. DNA base modification: ionized base pairs and mutagenesis. *Mutation Research* **177,** 201–218 (1987).

131. Bauer, N. C., Corbett, A. H. & Doetsch, P. W. The current state of eukaryotic DNA base damage and repair. *Nucleic Acids Research* **43,** 10083–10101 (2015).

132. Christmann, M., Verbeek, B., Roos, W. P. & Kaina, B. O(6)-Methylguanine-DNA methyltransferase (MGMT) in normal tissues and tumors: enzyme activity, promoter methylation and immunohistochemistry. *Biochimica Et Biophysica Acta* **1816,** 179–190 (2011).

133. Velu, C. S., Niture, S. K., Bailey, N. I. & Srivenugopal, K. S. MGMT protein turnover in human tumors is mediated by the ubiquitin-proteasome pathway: Phosphorylation-dependent Ub-conjugation by the Skp2-SCF complex. *Cancer Research* **65,** 1283 (2005).

134. Garg, P. & Burgers, P. M. Ubiquitinated proliferating cell nuclear antigen activates translesion DNA polymerases eta and REV1. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 18361–18366 (2005).

135. Plosky, B. S. *et al.* Controlling the subcellular localization of DNA polymerases iota and eta via interactions with ubiquitin. *The EMBO journal* **25,** 2847–2855 (2006).

136. Wit, N. *et al.* Roles of PCNA ubiquitination and TLS polymerases and in the bypass of methyl methanesulfonate-induced DNA damage. *Nucleic Acids Research* **43,** 282–294 (2015).

137. Pillaire, M.-J., Bétous, R. & Hoffmann, J.-S. Role of DNA polymerase in the maintenance of genomic stability. *Molecular & Cellular Oncology* **1.** Publisher: Taylor & Francis, e29902 (2014).

138. Chang, D. J. & Cimprich, K. A. DNA damage tolerance: when it's OK to make mistakes. *Nature Chemical Biology* **5,** 82–90 (2009).

139. Friedman, J. I. & Stivers, J. T. Detection of Damaged DNA Bases by DNA Glycosylase Enzymes. *Biochemistry* **49,** 4957–4967 (2010).

140. Jacobs, A. L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121,** 1–20 (2012).

141. Slupphaug, G. *et al.* A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384,** 87–92 (1996).

142. Mol, C. D. *et al.* Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* **80,** 869–878 (1995).

143. Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* **373,** 487–493 (1995).

144. Izumi, T. & Mellon, I. in *Genome Stability* (eds Kovalchuk, I. & Kovalchuk, O.) 275–302 (Academic Press, Boston, 2016).

145. Bennett, R. A. O. & Demple, B. in *Encyclopedia of Biological Chemistry (Second Edition)* (eds Lennarz, W. J. & Lane, M. D.) 1–8 (Academic Press, Waltham, 2013).

146. Maher, R. L., Wallace, S. S. & Pederson, D. S. The lyase activity of bifunctional DNA glycosylases and the 3-diesterase activity of APE1 contribute to the repair of oxidized bases in nucleosomes. *Nucleic Acids Research* **47,** 2922–2931 (2019).

147. Wang, L. *et al.* Recent advances in DNA glycosylase assays. *Chinese Chemical Letters* **33,** 3603–3612 (2022).

148. Morland, I., Luna, L., Gustad, E., Seeberg, E. & Bjørås, M. Product inhibition and magnesium modulate the dual reaction mode of hOgg1. *DNA Repair* **4,** 381–387 (2005).

149. Allinson, S. L., Dianova, I. I. & Dianov, G. L. DNA polymerase is the major dRP lyase involved in repair of oxidative base lesions in DNA by mammalian cell extracts. *The EMBO Journal* **20,** 6919–6926 (2001).

150. Suh, D., Wilson, D. M. & Povirk, L. F. 3'-phosphodiesterase activity of human apurinic/apyrimidinic endonuclease at DNA double-strand break ends. *Nucleic Acids Research* **25,** 2495–2500 (1997).

151. Parsons, J. L., Dianova, I. I. & Dianov, G. L. APE1 is the major 3'-phosphoglycolate activity in human cell extracts. *Nucleic Acids Research* **32,** 3531–3536 (2004).

152. Winters, T. A., Henner, W. D., Russell, P. S., McCullough, A. & Jorgensen, T. J. Removal of 3'-phosphoglycolate from DNA strand-break damage in an oligonucleotide substrate by recombinant human apurinic/apyrimidinic endonuclease 1. *Nucleic Acids Research* **22,** 1866–1873 (1994).

153. Jilani, A. *et al.* Molecular cloning of the human gene, PNKP, encoding a polynucleotide kinase 3'-phosphatase and evidence for its role in repair of DNA strand breaks caused by oxidative damage. *The Journal of Biological Chemistry* **274,** 24176–24186 (1999).

154. Karimi-Busheri, F. *et al.* Molecular characterization of a human DNA kinase. *The Journal of Biological Chemistry* **274,** 24187–24194 (1999).

155. Robertson, A. B., Klungland, A., Rognes, T. & Leiros, I. DNA repair in mammalian cells: Base excision repair: the long and short of it. *Cellular and molecular life sciences: CMLS* **66,** 981–993 (2009).

156. DeMott, M. S. *et al.* Covalent Trapping of Human DNA Polymerase by the Oxidative DNA Lesion 2-Deoxyribonolactone*. *Journal of Biological Chemistry* **277,** 7637–7640 (2002).

157. Wilson, S. H. *et al.* Base excision repair and design of small molecule inhibitors of human DNA polymerase. *Cellular and molecular life sciences: CMLS* **67,** 3633–3647 (2010).

158. Pascucci, B., Stucki, M., Jónsson, Z. O., Dogliotti, E. & Hübscher, U. Long Patch Base Excision Repair with Purified Human Proteins: DNA LIGASE I AS PATCH SIZE MEDIATOR FOR DNA POLYMERASES AND *. *Journal of Biological Chemistry* **274,** 33696–33702 (1999).

159. Matsumoto, Y. *et al.* Reconstitution of Proliferating Cell Nuclear Antigen-dependent Repair of Apurinic/Apyrimidinic Sites with Purified Human Proteins*. *Journal of Biological Chemistry* **274,** 33703–33708 (1999).

160. Stucki, M. *et al.* Mammalian base excision repair by DNA polymerases delta and epsilon. *Oncogene* **17,** 835–843 (1998).

161. Fortini, P. *et al.* Different DNA polymerases are involved in the short- and long-patch base excision repair in mammalian cells. *Biochemistry* **37,** 3575–3580 (1998).

162. Tang, Q., Kamble, P. & Çağlayan, M. DNA ligase I variants fail in the ligation of mutagenic repair intermediates with mismatches and oxidative DNA damage. *Mutagenesis* **35,** 391–404 (2020).

163. Kamble, P., Hall, K., Chandak, M., Tang, Q. & Çağlayan, M. DNA ligase I fidelity mediates the mutagenic ligation of pol oxidized and mismatch nucleotide insertion products in base excision repair. *Journal of Biological Chemistry* **296.** Publisher: Elsevier (2021).

164. Cappelli, E. *et al.* Involvement of XRCC1 and DNA Ligase III Gene Products in DNA Base Excision Repair*. *Journal of Biological Chemistry* **272,** 23970–23975 (1997).

165. Caldecott, K. W. DNA single-strand break repair. *Experimental Cell Research* **329,** 2–8 (2014).

166. Caldecott, K. W. XRCC1 protein; Form and function. *DNA Repair. Cutting-edge Perspectives in Genomic Maintenance VI* **81,** 102664 (2019).

167. Pogozelski, W. K. & Tullius, T. D. Oxidative Strand Scission of Nucleic Acids: Routes Initiated by Hydrogen Abstraction from the Sugar Moiety. *Chemical Reviews* **98.** Publisher: American Chemical Society, 1089–1108 (1998).

168. Masson, M. *et al.* XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage. *Molecular and Cellular Biology* **18,** 3563–3571 (1998).

169. Okano, S., Lan, L., Caldecott, K. W., Mori, T. & Yasui, A. Spatial and temporal cellular responses to single-strand breaks in human cells. *Molecular and Cellular Biology* **23,** 3974–3981 (2003).

170. Harrision, D., Gravells, P., Thompson, R. & Bryant, H. E. Poly(ADP-Ribose) Glycohydrolase (PARG) vs. Poly(ADP-Ribose) Polymerase (PARP) – Function in Genome Maintenance and Relevance of Inhibitors for Anti-cancer Therapy. *Frontiers in Molecular Biosciences* **7,** 191 (2020).

171. Kim, M. Y., Mauro, S., Gévry, N., Lis, J. T. & Kraus, W. L. NAD+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. *Cell* **119,** 803–814 (2004).

172. Nick McElhinny, S. A. *et al.* Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 4949–4954 (2010).

173. Williams, J. S. & Kunkel, T. A. Ribonucleotide Incorporation by Eukaryotic B-Family Replicases and Its Implications for Genome Stability. *Annual Review of Biochemistry* **91,** 133–155 (2022).

174. Rydberg, B. & Game, J. Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 16654–16659 (2002).

175. Eder, P. S., Walder, R. Y. & Walder, J. A. Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. *Biochimie* **75,** 123–126 (1993).

176. Sparks, J. L. *et al.* RNase H2-Initiated Ribonucleotide Excision Repair. *Molecular Cell* **47,** 980–986 (2012).

177. Vilar, E. & Gruber, S. B. Microsatellite instability in colorectal cancer—the stable evidence. *Nature Reviews Clinical Oncology* **7.** Publisher: Nature Publishing Group, 153–162 (2010).

178. Li, K., Luo, H., Huang, L., Luo, H. & Zhu, X. Microsatellite instability: a review of what the oncologist should know. *Cancer Cell International* **20,** 16 (2020).

179. Subramanian, S., Mishra, R. K. & Singh, L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biology* **4,** R13 (2003).

180. Bacher, J. W., Abdel Megid, W. M., Kent-First, M. G. & Halberg, R. B. Use of mononucleotide repeat markers for detection of microsatellite instability in mouse tumors. *Molecular Carcinogenesis* **44,** 285–292 (2005).

181. Van Wietmarschen, N. *et al.* Repeat expansions confer WRN dependence in microsatellite-unstable cancers. *Nature* **586,** 292–298 (2020).

182. Depienne, C. & Mandel, J.-L. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *The American Journal of Human Genetics* **108,** 764–785 (2021).

183. Peprah, E. Fragile X syndrome: the FMR1 CGG repeat distribution among world populations. *Annals of Human Genetics* **76,** 178–191 (2012).

184. Duyao, M. *et al.* Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genetics* **4,** 387–392 (1993).

185. Aldous, S. G. *et al.* A CAG repeat threshold for therapeutics targeting somatic instability in Huntington's disease. *Brain* **147,** 1784–1798 (2024).

186. Gul, I. S., Hulpiau, P., Saeys, Y. & van Roy, F. Metazoan evolution of the armadillo repeat superfamily. *Cellular and molecular life sciences: CMLS* **74,** 525–541 (2017).

187. LeClerc, J. E., Li, B., Payne, W. L. & Cebula, T. A. Promiscuous Origin of a Chimeric Sequence in the Escherichia coli O157:H7 Genome. *Journal of Bacteriology* **181.** Publisher: American Society for Microbiology, 7614–7617 (1999).

188. Matic, I. *et al.* Highly Variable Mutation Rates in Commensal and Pathogenic Escherichia coli. *Science* **277.** Publisher: American Association for the Advancement of Science, 1833–1834 (1997).

189. Kolodner, R. Biochemistry and genetics of eukaryotic mismatch repair. *Genes & Development* **10.** Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab, 1433–1442 (1996).

190. Marsischky, G. T., Filosi, N., Kane, M. F. & Kolodner, R. Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. *Genes & Development* **10.** Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab, 407–420 (1996).

191. Kunkel, T. A. & Erie, D. A. DNA MISMATCH REPAIR*. *Annual Review of Biochemistry* **74.** Publisher: Annual Reviews, 681–710 (Volume 74, 2005 2005).

192. Umar, A. *et al.* Requirement for PCNA in DNA Mismatch Repair at a Step Preceding DNA Resynthesis. *Cell* **87.** Publisher: Elsevier, 65–73 (1996).

193. Gu, L., Hong, Y., McCulloch, S., Watanabe, H. & Li, G.-M. ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair. *Nucleic Acids Research* **26,** 1173–1178 (1998).

194. Tishkoff, D. X. *et al.* Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts withMSH2. *Proceedings of the National Academy of Sciences* **94.** Publisher: Proceedings of the National Academy of Sciences, 7487–7492 (1997).

195. Tishkoff, D. X., Amin, N. S., Viars, C. S., Arden, K. C. & Kolodner, R. D. Identification of a Human Gene Encoding a Homologue of Saccharomyces cerevisiae EXO1, an Exonuclease Implicated in Mismatch Repair and Recombination1. *Cancer Research* **58,** 5027–5031 (1998).

196. Schofield, M. J. & Hsieh, P. DNA mismatch repair: molecular mechanisms and biological function. *Annual Review of Microbiology* **57,** 579–608 (2003).

197. Longley, M. J., Pierce, A. J. & Modrich, P. DNA polymerase delta is required for human mismatch repair in vitro. *The Journal of Biological Chemistry* **272,** 10917–10921 (1997).

198. Zhang, Y. *et al.* Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell* **122,** 693–705 (2005).

199. Petruseva, I., Evdokimov, A. N. & Lavrik, O. I. Molecular Mechanism of Global Genome Nucleotide Excision Repair. *Acta Naturae* **6,** 23–34 (2014).

200. Duan, M., Speer, R. M., Ulibarri, J., Liu, K. J. & Mao, P. Transcription-coupled nucleotide excision repair: New insights revealed by genomic approaches. *DNA repair* **103,** 103126 (2021).

201. Mu, H., Geacintov, N. E., Broyde, S., Yeo, J.-E. & Schärer, O. D. Molecular basis for damage recognition and verification by XPC-RAD23B and TFIIH in nucleotide excision repair. *DNA repair* **71,** 33–42 (2018).

202. Nishi, R. *et al.* Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. *Molecular and Cellular Biology* **25,** 5664–5674 (2005).

203. Araki, M. *et al.* Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. *The Journal of Biological Chemistry* **276,** 18665–18672 (2001).

204. Cheon, N. Y., Kim, H.-S., Yeo, J.-E., Schärer, O. D. & Lee, J. Y. Single-molecule visualization reveals the damage search mechanism for the human NER protein XPC-RAD23B. *Nucleic Acids Research* **47,** 8337–8347 (2019).

205. Chu, G. & Chang, E. Xeroderma pigmentosum group E cells lack a nuclear factor that binds to damaged DNA. *Science (New York, N.Y.)* **242,** 564–567 (1988).

206. Keeney, S., Chang, G. J. & Linn, S. Characterization of a human DNA damage binding protein implicated in xeroderma pigmentosum E. *The Journal of Biological Chemistry* **268,** 21293–21300 (1993).

207. Scrima, A. *et al.* Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* **135,** 1213–1223 (2008).

208. Wittschieben, B. Ø., Iwai, S. & Wood, R. D. DDB1-DDB2 (xeroderma pigmentosum group E) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/apyrimidinic sites, and compound lesions in DNA. *The Journal of Biological Chemistry* **280,** 39982–39989 (2005).

209. Bergink, S. *et al.* Recognition of DNA damage by XPC coincides with disruption of the XPC–RAD23 complex. *Journal of Cell Biology* **196,** 681–688 (2012).

210. Compe, E. & Egly, J.-M. TFIIH: when transcription met DNA repair. *Nature Reviews Molecular Cell Biology* **13.** Publisher: Nature Publishing Group, 343–354 (2012).

211. Bernardes de Jesus, B. M., Bjørås, M., Coin, F. & Egly, J. M. Dissection of the molecular defects caused by pathogenic mutations in the DNA repair factor XPC. *Molecular and Cellular Biology* **28,** 7225–7235 (2008).

212. Rimel, J. K. & Taatjes, D. J. The essential and multifunctional TFIIH complex. *Protein Science: A Publication of the Protein Society* **27,** 1018–1037 (2018).

213. Tapias, A. *et al.* Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. *The Journal of Biological Chemistry* **279,** 19074–19083 (2004).

214. Mathieu, N., Kaczmarek, N., Rüthemann, P., Luch, A. & Naegeli, H. DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH. *Current biology: CB* **23,** 204–212 (2013).

215. Camenisch, U., Dip, R., Schumacher, S. B., Schuler, B. & Naegeli, H. Recognition of helical kinks by xeroderma pigmentosum group A protein triggers DNA excision repair. *Nature Structural & Molecular Biology* **13.** Publisher: Nature Publishing Group, 278–284 (2006).

216. Winkler, G. S. *et al.* TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair. *The Journal of Biological Chemistry* **275,** 4258–4266 (2000).

217. Fan, L. *et al.* XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* **133,** 789–800 (2008).

218. Wolski, S. C. *et al.* Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLoS biology* **6,** e149 (2008).

219. Coin, F. *et al.* Nucleotide excision repair driven by the dissociation of CAK from TFIIH. *Molecular Cell* **31,** 9–20 (2008).

220. Sandrock, B. & Egly, J. M. A yeast four-hybrid system identifies Cdk-activating kinase as a regulator of the XPD helicase, a subunit of transcription factor IIH. *The Journal of Biological Chemistry* **276,** 35328–35333 (2001).

221. Marteijn, J. A., Lans, H., Vermeulen, W. & Hoeijmakers, J. H. J. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature Reviews Molecular Cell Biology* **15.** Publisher: Nature Publishing Group, 465–481 (2014).

222. Marietta, C. & Brooks, P. J. Transcriptional bypass of bulky DNA lesions causes new mutant RNA transcripts in human cells. *EMBO reports* **8,** 388–393 (2007).

223. Hendriks, G. *et al.* Transcription-dependent cytosine deamination is a novel mechanism in ultraviolet light-induced mutagenesis. *Current biology: CB* **20,** 170–175 (2010).

224. Mayne, L. V. & Lehmann, A. R. Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Research* **42,** 1473–1478 (1982).

225. Tantin, D., Kansal, A. & Carey, M. Recruitment of the putative transcription-repair coupling factor CSB/ERCC6 to RNA polymerase II elongation complexes. *Molecular and Cellular Biology* **17,** 6803–6814 (1997).

226. Schwertman, P. *et al.* UV-sensitive syndrome protein UVSSA recruits USP7 to regulate transcription-coupled repair. *Nature Genetics* **44,** 598–602 (2012).

227. Charlet-Berguerand, N. *et al.* RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. *The EMBO Journal* **25,** 5481–5491 (2006).

228. Menoni, H., Hoeijmakers, J. H. J. & Vermeulen, W. Nucleotide excision repair-initiating proteins bind to oxidative DNA lesions in vivo. *The Journal of Cell Biology* **199,** 1037–1046 (2012).

229. Nardo, T. *et al.* A UV-sensitive syndrome patient with a specific CSA mutation reveals separable roles for CSA in response to UV and oxidative DNA damage. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 6209–6214 (2009).

230. Tornaletti, S., Reines, D. & Hanawalt, P. C. Structural characterization of RNA polymerase II complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template DNA. *The Journal of Biological Chemistry* **274,** 24124–24130 (1999).

231. Yang, K. B. *et al.* Persistence of backtracking by human RNA polymerase II. *Molecular Cell* **84,** 897–909.e4 (2024).

232. Vidaković, A. T. *et al.* Regulation of the RNAPII Pool Is Integral to the DNA Damage Response. *Cell* **180.** Publisher: Elsevier, 1245–1261.e21 (2020).

233. Nakazawa, Y. *et al.* Ubiquitination of DNA Damage-Stalled RNAPII Promotes Transcription-Coupled Repair. *Cell* **180.** Publisher: Elsevier, 1228–1244.e24 (2020).

234. Muñoz, J. C., Beckerman, I., Choudhary, R., Bouvier, L. A. & Muñoz, M. J. DNA Damage-Induced RNAPII Degradation and Its Consequences in Gene Expression. *Genes* **13,** 1951 (2022).

235. Heuvel, D. v. d. *et al.* STK19 facilitates the clearance of lesion-stalled RNAPII during transcription-coupled DNA repair. *Cell* **187.** Publisher: Elsevier, 7107–7125.e25 (2024).

236. Staresincic, L. *et al.* Coordination of dual incision and repair synthesis in human nucleotide excision repair. *The EMBO journal* **28,** 1111–1120 (2009).

237. Tsodikov, O. V. *et al.* Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *The EMBO journal* **26,** 4768–4776 (2007).

238. Orelli, B. *et al.* The XPA-binding domain of ERCC1 is required for nucleotide excision repair but not other DNA repair pathways. *The Journal of Biological Chemistry* **285,** 3705–3712 (2010).

239. De Laat, W. L. *et al.* DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes & Development* **12,** 2598–2609 (1998).

240. Zotter, A. *et al.* Recruitment of the nucleotide excision repair endonuclease XPG to sites of UV-induced dna damage depends on functional TFIIH. *Molecular and Cellular Biology* **26,** 8868–8879 (2006).

241. Ogi, T. *et al.* Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Molecular Cell* **37,** 714–727 (2010).

242. Paul-Konietzko, K., Thomale, J., Arakawa, H. & Iliakis, G. DNA Ligases I and III Support Nucleotide Excision Repair in DT40 Cells with Similar Efficiency. *Photochemistry and Photobiology* **91,** 1173–1180 (2015).

243. Hanasoge, S. & Ljungman, M. H2AX phosphorylation after UV irradiation is triggered by DNA repair intermediates and is mediated by the ATR kinase. *Carcinogenesis* **28,** 2298–2304 (2007).

244. DiGiovanna, J. J. & Kraemer, K. H. Shining a light on xeroderma pigmentosum. *The Journal of Investigative Dermatology* **132,** 785–796 (2012).

245. Niedernhofer, L. J., Bohr, V. A., Sander, M. & Kraemer, K. H. Xeroderma pigmentosum and other diseases of human premature aging and DNA repair: molecules to patients. *Mechanisms of Ageing and Development* **132,** 340–347 (2011).

246. Kraemer, K. H. *et al.* Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship. *Neuroscience* **145,** 1388–1396 (2007).

247. Jelinek, M., Jurajda, M. & Duris, K. Oxidative Stress in the Brain: Basic Concepts and Treatment Strategies in Stroke. *Antioxidants* **10,** 1886 (2021).

248. Shadfar, S., Brocardo, M. & Atkin, J. D. The Complex Mechanisms by Which Neurons Die Following DNA Damage in Neurodegenerative Diseases. *International Journal of Molecular Sciences* **23,** 2484 (2022).

249. Karikkineth, A. C., Scheibye-Knudsen, M., Fivenson, E., Croteau, D. L. & Bohr, V. A. Cockayne syndrome: Clinical features, model systems and pathways. *Ageing Research Reviews. Monogenic Accelerated Aging Disorders with Perturbations to Normal DNA and Chromosome Function* **33,** 3–17 (2017).

250. Ciaffardini, F. *et al.* The cockayne syndrome B protein is essential for neuronal differentiation and neuritogenesis. *Cell Death & Disease* **5.** Publisher: Nature Publishing Group, e1268–e1268 (2014).

251. Chintaluri, C. & Vogels, T. P. Metabolically regulated spiking could serve neuronal energy homeostasis and protect from reactive oxygen species. *Proceedings of the National Academy of Sciences of the United States of America* **120,** e2306525120 (2023).

252. Sallmyr, A. & Tomkinson, A. E. Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. *The Journal of Biological Chemistry* **293,** 10536–10546 (2018).

253. Scully, R., Panday, A., Elango, R. & Willis, N. A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews. Molecular Cell Biology* **20,** 698–714 (2019).

254. Zirkle, R. E. & Bloom, W. Irradiation of parts of individual cells. *Science (New York, N.Y.)* **117,** 487–493 (1953).

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

256. Li, X. & Heyer, W.-D. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Research* **18.** Publisher: Nature Publishing Group, 99–113 (2008).

257. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell cycle (Georgetown, Tex.)* **7,** 2902–2906 (2008).

258. Saha, T., Sundaravinayagam, D. & Di Virgilio, M. Charting a DNA Repair Roadmap for Immunoglobulin Class Switch Recombination. *Trends in Biochemical Sciences* **46,** 184–199 (2021).

259. Malu, S., Malshetty, V., Francis, D. & Cortes, P. Role of non-homologous end joining in V(D)J recombination. *Immunologic Research* **54,** 233–246 (2012).

260. Castañeda-Zegarra, S. *et al.* Genetic interaction between the non-homologous end-joining factors during B and T lymphocyte development: In vivo mouse models. *Scandinavian Journal of Immunology* **92,** e12936 (2020).

261. Hung, P. J. *et al.* Deficiency of XLF and PAXX prevents DNA double-strand break repair by non-homologous end joining in lymphocytes. *Cell Cycle (Georgetown, Tex.)* **16,** 286–295 (2017).

262. Mikhova, M. *et al.* Single-molecule imaging reveals the kinetics of non-homologous end-joining in living cells. *Nature Communications* **15.** Publisher: Nature Publishing Group, 10159 (2024).

263. Britton, S., Coates, J. & Jackson, S. P. A new method for high-resolution imaging of Ku foci to decipher mechanisms of DNA double-strand break repair. *The Journal of Cell Biology* **202,** 579–595 (2013).

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

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

266. Ochi, T. *et al.* DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science (New York, N.Y.)* **347,** 185–188 (2015).

267. Kumar, V., Alt, F. W. & Frock, R. L. PAXX and XLF DNA repair factors are functionally redundant in joining DNA breaks in a G1-arrested progenitor B-cell line. *Proceedings of the National Academy of Sciences of the United States of America* **113,** 10619–10624 (2016).

268. Chen, S. *et al.* Structural basis of long-range to short-range synaptic transition in NHEJ. *Nature* **593.** Publisher: Nature Publishing Group, 294–298 (2021).

269. Graham, T. G., Walter, J. C. & Loparo, J. J. Two-Stage Synapsis of DNA Ends during Non-Homologous End Joining. *Molecular cell* **61,** 850–858 (2016).

270. Liu, L. *et al.* Autophosphorylation transforms DNA-PK from protecting to processing DNA ends. *Molecular Cell* **82,** 177–189.e4 (2022).

271. Stinson, B. M., Moreno, A. T., Walter, J. C. & Loparo, J. J. A Mechanism to Minimize Errors during Non-homologous End Joining. *Molecular Cell* **77,** 1080–1091.e8 (2020).

272. Chang, H. H. Y., Watanabe, G. & Lieber, M. R. Unifying the DNA end-processing roles of the artemis nuclease: Ku-dependent artemis resection at blunt DNA ends. *The Journal of Biological Chemistry* **290,** 24036–24050 (2015).

273. Mirman, Z., Cai, S. & de Lange, T. CST/Pol/primase-mediated fill-in synthesis at DSBs. *Cell Cycle (Georgetown, Tex.)* **22,** 379–389 (2023).

274. Setiaputra, D. & Durocher, D. Shieldin - the protector of DNA ends. *EMBO reports* **20,** e47560 (2019).

275. Song, B., Yang, S., Hwang, G.-H., Yu, J. & Bae, S. Analysis of NHEJ-Based DNA Repair after CRISPR-Mediated DNA Cleavage. *International Journal of Molecular Sciences* **22,** 6397 (2021).

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

277. Jiang, W. *et al.* Differential phosphorylation of DNA-PKcs regulates the interplay between end-processing and end-ligation during nonhomologous end-joining. *Molecular Cell* **58,** 172–185 (2015).

278. Lamarche, B. J., Orazio, N. I. & Weitzman, M. D. The MRN complex in Double-Strand Break Repair and Telomere Maintenance. *FEBS letters* **584,** 3682–3695 (2010).

279. Dinkelmann, M. *et al.* Multiple functions of MRN in end-joining pathways during isotype class switching. *Nature Structural & Molecular Biology* **16,** 808–813 (2009).

280. Williams, R. S. *et al.* Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* **135,** 97–109 (2008).

281. Paull, T. T. & Gellert, M. The 3' to 5' exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. *Molecular Cell* **1,** 969–979 (1998).

282. Hu, C. *et al.* The Contribution of Germline Predisposition Gene Mutations to Clinical Subtypes of Invasive Breast Cancer From a Clinical Genetic Testing Cohort. *Journal of the National Cancer Institute* **112,** 1231–1241 (2020).

283. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).

284. Lord, C. J., McDonald, S., Swift, S., Turner, N. C. & Ashworth, A. A high-throughput RNA interference screen for DNA repair determinants of PARP inhibitor sensitivity. *DNA repair* **7,** 2010–2019 (2008).

285. Hu, J. & Crickard, J. B. All who wander are not lost: the search for homology during homologous recombination. *Biochemical Society Transactions* **52,** 367–377 (2024).

286. Tye, S., Ronson, G. E. & Morris, J. R. A fork in the road: Where homologous recombination and stalled replication fork protection part ways. *Seminars in Cell & Developmental Biology* **113,** 14–26 (2021).

287. Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiology and molecular biology reviews: MMBR* **63,** 349–404 (1999).

288. Daboussi, F. *et al.* A homologous recombination defect affects replication-fork progression in mammalian cells. *Journal of Cell Science* **121,** 162–166 (Pt 2 2008).

289. Lee, J.-H. & Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science (New York, N.Y.)* **308,** 551–554 (2005).

290. Lee, J.-H. & Paull, T. T. Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. *Science (New York, N.Y.)* **304,** 93–96 (2004).

291. Garcia, V., Phelps, S. E. L., Gray, S. & Neale, M. J. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature* **479.** Publisher: Nature Publishing Group, 241–244 (2011).

292. Wang, W., Daley, J. M., Kwon, Y., Krasner, D. S. & Sung, P. Plasticity of the Mre11-Rad50-Xrs2-Sae2 nuclease ensemble in the processing of DNA-bound obstacles. *Genes & Development* **31,** 2331–2336 (2017).

293. Reginato, G., Cannavo, E. & Cejka, P. Physiological protein blocks direct the Mre11-Rad50-Xrs2 and Sae2 nuclease complex to initiate DNA end resection. *Genes & Development* **31,** 2325–2330 (2017).

294. Chen, L., Nievera, C. J., Lee, A. Y.-L. & Wu, X. Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair. *The Journal of Biological Chemistry* **283,** 7713–7720 (2008).

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

296. Lengsfeld, B. M., Rattray, A. J., Bhaskara, V., Ghirlando, R. & Paull, T. T. Sae2 is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 complex. *Molecular Cell* **28,** 638–651 (2007).

297. Sartori, A. A. *et al.* Human CtIP promotes DNA end resection. *Nature* **450,** 509–514 (2007).

298. Limbo, O. *et al.* Ctp1 is a cell-cycle-regulated protein that functions with Mre11 complex to control double-strand break repair by homologous recombination. *Molecular Cell* **28,** 134–146 (2007).

299. Kimble, M. T., Johnson, M. J., Nester, M. R. & Symington, L. S. Long-range DNA end resection supports homologous recombination by checkpoint activation rather than extensive homology generation. *eLife* **12** (eds Aguilera, A., Struhl, K. & Smolka, M. B.) Publisher: eLife Sciences Publications, Ltd, e84322 (2023).

300. Shim, E. Y. *et al.* Saccharomyces cerevisiae Mre11/Rad50/Xrs2 and Ku proteins regulate association of Exo1 and Dna2 with DNA breaks. *The EMBO journal* **29,** 3370–3380 (2010).

301. Sturzenegger, A. *et al.* DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *The Journal of Biological Chemistry* **289,** 27314–27326 (2014).

302. Ceppi, I. *et al.* Mechanism of BRCA1–BARD1 function in DNA end resection and DNA protection. *Nature* **634.** Publisher: Nature Publishing Group, 492–500 (2024).

303. Salunkhe, S. *et al.* Promotion of DNA end resection by BRCA1–BARD1 in homologous recombination. *Nature* **634.** Publisher: Nature Publishing Group, 482–491 (2024).

304. Chen, H., Lisby, M. & Symington, L. S. RPA coordinates DNA end resection and prevents formation of DNA hairpins. *Molecular Cell* **50,** 589–600 (2013).

305. Sung, P. & Klein, H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature Reviews Molecular Cell Biology* **7.** Publisher: Nature Publishing Group, 739–750 (2006).

306. Yang, H., Li, Q., Fan, J., Holloman, W. K. & Pavletich, N. P. The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA–ssDNA junction. *Nature* **433.** Publisher: Nature Publishing Group, 653–657 (2005).

307. Game, J. C. & Mortimer, R. K. A genetic study of x-ray sensitive mutants in yeast. *Mutation Research* **24,** 281–292 (1974).

308. Prakash, R., Zhang, Y., Feng, W. & Jasin, M. Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harbor Perspectives in Biology* **7,** a016600 (2015).

309. Drost, R. M. & Jonkers, J. Preclinical mouse models for BRCA1-associated breast cancer. *British Journal of Cancer* **101.** Publisher: Nature Publishing Group, 1651–1657 (2009).

310. Foo, T. K. & Xia, B. BRCA1-Dependent and Independent Recruitment of PALB2-BRCA2-RAD51 in the DNA Damage Response and Cancer. *Cancer Research* **82,** 3191–3197 (2022).

311. Zhao, W. *et al.* BRCA1-BARD1 promotes RAD51-mediated homologous DNA pairing. *Nature* **550,** 360–365 (2017).

312. Liu, S. *et al.* In vivo tracking of functionally tagged Rad51 unveils a robust strategy of homology search. *Nature Structural & Molecular Biology* **30.** Publisher: Nature Publishing Group, 1582–1591 (2023).

313. Ito, K., Murayama, Y., Takahashi, M. & Iwasaki, H. Two three-strand intermediates are processed during Rad51-driven DNA strand exchange. *Nature Structural & Molecular Biology* **25.** Publisher: Nature Publishing Group, 29–36 (2018).

314. Nassif, N., Penney, J., Pal, S., Engels, W. R. & Gloor, G. B. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Molecular and Cellular Biology* **14,** 1613–1625 (1994).

315. Mueller, J. E., Clyman, J., Huang, Y. J., Parker, M. M. & Belfort, M. Intron mobility in phage T4 occurs in the context of recombination-dependent DNA replication by way of multiple pathways. *Genes & Development* **10,** 351–364 (1996).

316. Miura, T. *et al.* Homologous Recombination via Synthesis-Dependent Strand Annealing in Yeast Requires the Irc20 and Srs2 DNA Helicases. *Genetics* **191,** 65–78 (2012).

317. Westmoreland, J. W. & Resnick, M. A. Coincident resection at both ends of random, -induced double-strand breaks requires MRX (MRN), Sae2 (Ctp1), and Mre11-nuclease. *PLoS genetics* **9,** e1003420 (2013).

318. Ho, H. N. & West, S. C. Generation of double Holliday junction DNAs and their dissolution/resolution within a chromatin context. *Proceedings of the National Academy of Sciences* **119.** Publisher: Proceedings of the National Academy of Sciences, e2123420119 (2022).

319. Liberi, G. & Foiani, M. The double life of Holliday junctions. *Cell Research* **20.** Publisher: Nature Publishing Group, 611–613 (2010).

320. Wyatt, H. D. M. & West, S. C. Holliday junction resolvases. *Cold Spring Harbor Perspectives in Biology* **6,** a023192 (2014).

321. Wechsler, T., Newman, S. & West, S. C. Aberrant chromosome morphology in human cells defective for Holliday junction resolution. *Nature* **471,** 642–646 (2011).

322. Ip, S. C. Y. *et al.* Identification of Holliday junction resolvases from humans and yeast. *Nature* **456,** 357–361 (2008).

323. Svendsen, J. M. *et al.* Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. *Cell* **138,** 63–77 (2009).

324. Wyatt, H. D. M., Laister, R. C., Martin, S. R., Arrowsmith, C. H. & West, S. C. The SMX DNA Repair Tri-nuclease. *Molecular Cell* **65,** 848–860.e11 (2017).

325. Bizard, A. H. & Hickson, I. D. The Dissolution of Double Holliday Junctions. *Cold Spring Harbor Perspectives in Biology* **6,** a016477 (2014).

326. Liu, L. & Malkova, A. Break-induced replication: unraveling each step. *Trends in Genetics* **38.** Publisher: Elsevier, 752–765 (2022).

327. Kramara, J., Osia, B. & Malkova, A. Break-Induced Replication: The Where, The Why, and The How. *Trends in genetics: TIG* **34,** 518–531 (2018).

328. Yan, C. T. *et al.* IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449,** 478–482 (2007).

329. Truong, L. N. *et al.* Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 7720–7725 (2013).

330. Koole, W. *et al.* A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nature Communications* **5,** 3216 (2014).

331. Wood, R. D. & Doublié, S. DNA polymerase (POLQ), double-strand break repair, and cancer. *DNA Repair. Cutting-edge Perspectives in Genomic Maintenance III* **44,** 22–32 (2016).

332. Wyatt, D. W. *et al.* Essential Roles for Polymerase -Mediated End Joining in the Repair of Chromosome Breaks. *Molecular Cell* **63,** 662–673 (2016).

333. Kent, T., Mateos-Gomez, P. A., Sfeir, A. & Pomerantz, R. T. Polymerase is a robust terminal transferase that oscillates between three different mechanisms during end-joining. *eLife* **5,** e13740 (2016).

334. Mateos-Gomez, P. A. *et al.* Mammalian polymerase promotes alternative NHEJ and suppresses recombination. *Nature* **518,** 254–257 (2015).

335. Yousefzadeh, M. J. *et al.* Mechanism of suppression of chromosomal instability by DNA polymerase POLQ. *PLoS genetics* **10,** e1004654 (2014).

336. Mateos-Gomez, P. A. *et al.* The helicase domain of Pol counteracts RPA to promote alt-NHEJ. *Nature Structural & Molecular Biology* **24,** 1116–1123 (2017).

337. Fijen, C. *et al.* Sequential requirements for distinct Pol domains during theta-mediated end joining. *Molecular Cell* **84,** 1460–1474.e6 (2024).

338. Mengwasser, K. E. *et al.* Genetic Screens Reveal FEN1 and APEX2 as BRCA2 Synthetic Lethal Targets. *Molecular Cell* **73,** 885–899.e6 (2019).

339. Fleury, H. *et al.* The APE2 nuclease is essential for DNA double-strand break repair by microhomology-mediated end joining. *Molecular Cell* **83,** 1429–1445.e8 (2023).

340. Chan, S. H., Yu, A. M. & McVey, M. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. *PLoS genetics* **6,** e1001005 (2010).

341. Arana, M. E., Seki, M., Wood, R. D., Rogozin, I. B. & Kunkel, T. A. Low-fidelity DNA synthesis by human DNA polymerase theta. *Nucleic Acids Research* **36,** 3847–3856 (2008).

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

343. Brambati, A. *et al.* RHINO directs MMEJ to repair DNA breaks in mitosis. *Science* **381.** Publisher: American Association for the Advancement of Science, 653–660 (2023).

344. Ivanov, E. L., Sugawara, N., Fishman-Lobell, J. & Haber, J. E. Genetic requirements for the single-strand annealing pathway of double-strand break repair in Saccharomyces cerevisiae. *Genetics* **142,** 693–704 (1996).

345. Orel, N., Kyryk, A. & Puchta, H. Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. *The Plant Journal: For Cell and Molecular Biology* **35,** 604–612 (2003).

346. Sugawara, N., Ira, G. & Haber, J. E. DNA length dependence of the single-strand annealing pathway and the role of Saccharomyces cerevisiae RAD59 in double-strand break repair. *Molecular and Cellular Biology* **20,** 5300–5309 (2000).

347. Benitez, A. *et al.* FANCA Promotes DNA Double-Strand Break Repair by Catalyzing Single-Strand Annealing and Strand Exchange. *Molecular Cell* **71,** 621–628.e4 (2018).

348. Papp, D. *et al.* Massive contractions of myotonic dystrophy type 2-associated CCTG tetranucleotide repeats occur via double-strand break repair with distinct requirements for DNA helicases. *G3 (Bethesda, Md.)* **14,** jkad257 (2024).

349. Ho, V. *et al.* Aberrant Expression of RAD52, Its Prognostic Impact in Rectal Cancer and Association with Poor Survival of Patients. *International Journal of Molecular Sciences* **21,** 1768 (2020).

350. Tomkinson, A. E., Bardwell, A. J., Bardwell, L., Tappe, N. J. & Friedberg, E. C. Yeast DNA repair and recombination proteins Rad1 and Rad10 constitute a single-stranded-DNA endonuclease. *Nature* **362,** 860–862 (1993).

351. Evans, E., Sugawara, N., Haber, J. E. & Alani, E. The Saccharomyces cerevisiae Msh2 mismatch repair protein localizes to recombination intermediates in vivo. *Molecular Cell* **5,** 789–799 (2000).

352. Motycka, T. A., Bessho, T., Post, S. M., Sung, P. & Tomkinson, A. E. Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. *The Journal of Biological Chemistry* **279,** 13634–13639 (2004).

353. Al-Minawi, A. Z., Saleh-Gohari, N. & Helleday, T. The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. *Nucleic Acids Research* **36,** 1–9 (2008).

354. Li, F. *et al.* Microarray-based genetic screen defines SAW1, a gene required for Rad1/Rad10-dependent processing of recombination intermediates. *Molecular Cell* **30,** 325–335 (2008).

355. McVey, M., Khodaverdian, V. Y., Meyer, D., Cerqueira, P. G. & Heyer, W.-D. Eukaryotic DNA Polymerases in Homologous Recombination. *Annual Review of Genetics* **50,** 393–421 (2016).

356. Hicks, W. M., Kim, M. & Haber, J. E. Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science (New York, N.Y.)* **329,** 82–85 (2010).

357. Zhang, W.-W. & Matlashewski, G. Single-Strand Annealing Plays a Major Role in Double-Strand DNA Break Repair following CRISPR-Cas9 Cleavage in Leishmania. *mSphere* **4,** e00408–19 (2019).

358. Vu, T. V., Das, S., Nguyen, C. C., Kim, J. & Kim, J.-Y. Single-strand annealing: Molecular mechanisms and potential applications in CRISPR-Cas-based precision genome editing. *Biotechnology Journal* **17,** e2100413 (2022).

359. Ronson, G. E. *et al.* Mechanisms of synthetic lethality between BRCA1/2 and 53BP1 deficiencies and DNA polymerase theta targeting. *Nature Communications* **14.** Publisher: Nature Publishing Group, 7834 (2023).

360. Sfeir, A. & Symington, L. S. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends in Biochemical Sciences* **40,** 701–714 (2015).

361. Schimmel, J., Kool, H., van Schendel, R. & Tijsterman, M. Mutational signatures of non-homologous and polymerase theta-mediated end-joining in embryonic stem cells. *The EMBO journal* **36,** 3634–3649 (2017).

362. Saito, S., Maeda, R. & Adachi, N. Dual loss of human POLQ and LIG4 abolishes random integration. *Nature Communications* **8,** 16112 (2017).

363. Balboni, B. *et al.* Novel Insights into RAD52's Structure, Function, and Druggability for Synthetic Lethality and Innovative Anticancer Therapies. *Cancers* **15,** 1817 (2023).

364. Schrempf, A., Slyskova, J. & Loizou, J. I. Targeting the DNA Repair Enzyme Polymerase in Cancer Therapy. *Trends in Cancer* **7,** 98–111 (2021).

365. Zatreanu, D. *et al.* Pol inhibitors elicit BRCA-gene synthetic lethality and target PARP inhibitor resistance. *Nature Communications* **12,** 3636 (2021).

366. Enriquez-Rios, V. *et al.* DNA-PKcs, ATM, and ATR Interplay Maintains Genome Integrity during Neurogenesis. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **37,** 893–905 (2017).

367. Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual Review of Biochemistry* **79,** 181–211 (2010).

368. Cortez, D., Guntuku, S., Qin, J. & Elledge, S. J. ATR and ATRIP: partners in checkpoint signaling. *Science (New York, N.Y.)* **294,** 1713–1716 (2001).

369. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *The Journal of Biological Chemistry* **273,** 5858–5868 (1998).

370. Stucki, M. *et al.* MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123,** 1213–1226 (2005).

371. Jungmichel, S. *et al.* The molecular basis of ATM-dependent dimerization of the Mdc1 DNA damage checkpoint mediator. *Nucleic Acids Research* **40,** 3913–3928 (2012).

372. Liu, J. *et al.* Structural mechanism of the phosphorylation-dependent dimerization of the MDC1 forkhead-associated domain. *Nucleic Acids Research* **40,** 3898–3912 (2012).

373. Kuo, L. J. & Yang, L.-X. Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *In Vivo (Athens, Greece)* **22,** 305–309 (2008).

374. Iacovoni, J. S. *et al.* High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. *The EMBO journal* **29,** 1446–1457 (2010).

375. Ahn, J. Y., Schwarz, J. K., Piwnica-Worms, H. & Canman, C. E. Threonine 68 phosphorylation by ataxia telangiectasia mutated is required for efficient activation of Chk2 in response to ionizing radiation. *Cancer Research* **60,** 5934–5936 (2000).

376. Matsuoka, S. *et al.* Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 10389–10394 (2000).

377. Thada, V. & Cortez, D. Common motifs in ETAA1 and TOPBP1 required for ATR kinase activation. *The Journal of Biological Chemistry* **294,** 8395–8402 (2019).

378. Liu, Y., Vidanes, G., Lin, Y. C., Mori, S. & Siede, W. Characterization of a Saccharomyces cerevisiae homologue of Schizosaccharomyces pombe Chk1 involved in DNA-damage-induced M-phase arrest. *Molecular & general genetics: MGG* **262,** 1132–1146 (2000).

379. Ghelli Luserna di Rorà, A., Cerchione, C., Martinelli, G. & Simonetti, G. A WEE1 family business: regulation of mitosis, cancer progression, and therapeutic target. *Journal of Hematology & Oncology* **13,** 126 (2020).

380. Schmidt, M. *et al.* Regulation of G2/M Transition by Inhibition of WEE1 and PKMYT1 Kinases. *Molecules (Basel, Switzerland)* **22,** 2045 (2017).

381. Gorecki, L., Andrs, M. & Korabecny, J. Clinical Candidates Targeting the ATR-CHK1-WEE1 Axis in Cancer. *Cancers* **13,** 795 (2021).

382. D'Angiolella, V. *et al.* Cyclin F-mediated degradation of ribonucleotide reductase M2 controls genome integrity and DNA repair. *Cell* **149,** 1023–1034 (2012).

383. Bouwman, P. *et al.* 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. *Nature Structural & Molecular Biology* **17,** 688–695 (2010).

384. Zimmermann, M., Lottersberger, F., Buonomo, S. B., Sfeir, A. & de Lange, T. 53BP1 regulates DSB repair using Rif1 to control 5' end resection. *Science (New York, N.Y.)* **339,** 700–704 (2013).

385. Chapman, J. R., Sossick, A. J., Boulton, S. J. & Jackson, S. P. BRCA1-associated exclusion of 53BP1 from DNA damage sites underlies temporal control of DNA repair. *Journal of Cell Science* **125,** 3529–3534 (Pt 15 2012).

386. Hu, Q. *et al.* Mechanisms of BRCA1-BARD1 nucleosome recognition and ubiquitylation. *Nature* **596,** 438–443 (2021).

387. Nakamura, K. *et al.* H4K20me0 recognition by BRCA1-BARD1 directs homologous recombination to sister chromatids. *Nature Cell Biology* **21,** 311–318 (2019).

388. Becker, J. R. *et al.* BARD1 reads H2A lysine 15 ubiquitination to direct homologous recombination. *Nature* **596,** 433–437 (2021).

389. Bunting, S. F. *et al.* 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* **141,** 243–254 (2010).

390. Hurley, R. M. *et al.* 53BP1 as a potential predictor of response in PARP inhibitor-treated homologous recombination-deficient ovarian cancer. *Gynecologic Oncology* **153,** 127–134 (2019).

391. Mayor, P., Gay, L. M., Lele, S. & Elvin, J. A. BRCA1 reversion mutation acquired after treatment identified by liquid biopsy. *Gynecologic Oncology Reports* **21,** 57–60 (2017).

392. Paes Dias, M. *et al.* Loss of nuclear DNA ligase III reverts PARP inhibitor resistance in BRCA1/53BP1 double-deficient cells by exposing ssDNA gaps. *Molecular Cell* **81,** 4692–4708.e9 (2021).

393. Gupta, R. *et al.* DNA Repair Network Analysis Reveals Shieldin as a Key Regulator of NHEJ and PARP Inhibitor Sensitivity. *Cell* **173.** Publisher: Elsevier, 972–988.e23 (2018).

394. Dev, H. *et al.* Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. *Nature Cell Biology* **20.** Publisher: Nature Publishing Group, 954–965 (2018).

395. Ghezraoui, H. *et al.* 53BP1 cooperation with the REV7–shieldin complex underpins DNA structure-specific NHEJ. *Nature* **560.** Publisher: Nature Publishing Group, 122–127 (2018).

396. Noordermeer, S. M. *et al.* The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* **560.** Publisher: Nature Publishing Group, 117–121 (2018).

397. Niedernhofer, L. J., Daniels, J. S., Rouzer, C. A., Greene, R. E. & Marnett, L. J. Malondialdehyde, a product of lipid peroxidation, is mutagenic in human cells. *The Journal of Biological Chemistry* **278,** 31426–31433 (2003).

398. Warren, G. M. & Shuman, S. Structure and in vivo psoralen DNA crosslink repair activity of mycobacterial Nei2. *mBio* **15.** Publisher: American Society for Microbiology, e01248–24 (2024).

399. Zhang, J. *et al.* DNA interstrand cross-link repair requires replication-fork convergence. *Nature Structural & Molecular Biology* **22,** 242–247 (2015).

400. Semlow, D. R. & Walter, J. C. Mechanisms of Vertebrate DNA Interstrand Cross-Link Repair. *Annual Review of Biochemistry* **90.** _eprint: https://doi.org/10.1146/annurev-biochem-080320-112510, 107–135 (2021).

401. Nalepa, G. & Clapp, D. W. Fanconi anaemia and cancer: an intricate relationship. *Nature Reviews. Cancer* **18,** 168–185 (2018).

402. Lobitz, S. & Velleuer, E. Guido Fanconi (1892-1979): a jack of all trades. *Nature Reviews. Cancer* **6,** 893–898 (2006).

403. Fanconi, G. Familiare infantile perniciosaartige anamia (pernicioses blutbild und konstitution). *Jahb Kinderheikd* **117,** 257–280 (1927).

404. Fanconi, G. HYPOTHESIS OF CHROMOSOMAL TRANSLOCATION AS A GENETIC INTERPRETATION OF FANCONI'S FAMILIAL CONSTITUTIONAL PANMYELOPATHY. *Helvetica Paediatrica Acta* **19,** 29–33 (1964).

405. Kook, H. *et al.* Fanconi anemia screening by diepoxybutane and mitomicin C tests in Korean children with bone marrow failure syndromes. *Journal of Korean Medical Science* **13,** 623–628 (1998).

406. Ruggiero, J. L. *et al.* Cutaneous Findings in Fanconi Anemia. *Journal of the American Academy of Dermatology* **85,** 1253–1258 (2021).

407. Tsui, V. & Crismani, W. The Fanconi Anemia Pathway and Fertility. *Trends in Genetics* **35.** Publisher: Elsevier, 199–214 (2019).

408. Berezney, R., Dubey, D. D. & Huberman, J. A. Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma* **108,** 471–484 (2000).

409. Scaramuzza, S. *et al.* TRAIP resolves DNA replication-transcription conflicts during the S-phase of unperturbed cells. *Nature Communications* **14.** Publisher: Nature Publishing Group, 5071 (2023).

410. Harley, M. E. *et al.* TRAIP promotes DNA damage response during genome replication and is mutated in primordial dwarfism. *Nature Genetics* **48,** 36–43 (2016).

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

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

413. Imani Nejad, M. *et al.* Unhooking of an interstrand cross-link at DNA fork structures by the DNA glycosylase NEIL3. *DNA repair* **86,** 102752 (2020).

414. Fullbright, G., Rycenga, H. B., Gruber, J. D. & Long, D. T. p97 Promotes a Conserved Mechanism of Helicase Unloading during DNA Cross-Link Repair. *Molecular and Cellular Biology* **36,** 2983–2994 (2016).

415. Long, D. T., Joukov, V., Budzowska, M. & Walter, J. C. BRCA1 promotes unloading of the CMG helicase from a stalled DNA replication fork. *Molecular Cell* **56,** 174–185 (2014).

416. Amunugama, R. *et al.* Replication Fork Reversal during DNA Interstrand Crosslink Repair Requires CMG Unloading. *Cell Reports* **23.** Publisher: Elsevier, 3419–3428 (2018).

417. Basbous, J. & Constantinou, A. A tumor suppressive DNA translocase named FANCM. *Critical Reviews in Biochemistry and Molecular Biology* **54,** 27–40 (2019).

418. Singh, T. R. *et al.* MHF1-MHF2, a Histone-Fold-Containing Protein Complex, Participates in the Fanconi Anemia Pathway via FANCM. *Molecular cell* **37,** 879–886 (2010).

419. Ciccia, A. *et al.* Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Molecular Cell* **25,** 331–343 (2007).

420. Kim, J. M., Kee, Y., Gurtan, A. & D'Andrea, A. D. Cell cycle–dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24. *Blood* **111,** 5215–5222 (2008).

421. Huang, Y. *et al.* Modularized Functions of the Fanconi Anemia Core Complex. *Cell reports* **7,** 1849–1857 (2014).

422. Rickman, K. A. *et al.* Deficiency of UBE2T, the E2 ubiquitin ligase necessary for FANCD2 and FANCI ubiquitination, causes FA-T subtype of Fanconi anemia. *Cell reports* **12,** 35–41 (2015).

423. Chaugule, V. K. *et al.* Allosteric mechanism for site-specific ubiquitination of FANCD2. *Nature Chemical Biology* **16,** 291–301 (2020).

424. Chaugule, V. K., Arkinson, C., Toth, R. & Walden, H. Enzymatic preparation of monoubiquitinated FANCD2 and FANCI proteins. *Methods in Enzymology* **618,** 73–104 (2019).

425. Alpi, A. *et al.* UBE2T, the Fanconi anemia core complex, and FANCD2 are recruited independently to chromatin: a basis for the regulation of FANCD2 monoubiquitination. *Molecular and Cellular Biology* **27,** 8421–8430 (2007).

426. Machida, Y. J. *et al.* UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation. *Molecular Cell* **23,** 589–596 (2006).

427. Meetei, A. R. *et al.* A novel ubiquitin ligase is deficient in Fanconi anemia. *Nature Genetics* **35,** 165–170 (2003).

428. Van Twest, S. *et al.* Mechanism of Ubiquitination and Deubiquitination in the Fanconi Anemia Pathway. *Molecular Cell* **65,** 247–259 (2017).

429. Smogorzewska, A. *et al.* Identification of the FANCI Protein, a Monoubiquitinated FANCD2 Paralog Required for DNA Repair. *Cell* **129.** Publisher: Elsevier, 289–301 (2007).

430. Rennie, M. L., Arkinson, C., Chaugule, V. K., Toth, R. & Walden, H. Structural basis of FANCD2 deubiquitination by USP1-UAF1. *Nature Structural & Molecular Biology* **28,** 356–364 (2021).

431. Kim, J. M. *et al.* Inactivation of murine Usp1 results in genomic instability and a Fanconi anemia phenotype. *Developmental Cell* **16,** 314–320 (2009).

432. Nijman, S. M. B. *et al.* The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. *Molecular Cell* **17,** 331–339 (2005).

433. Sijacki, T. *et al.* The DNA-damage kinase ATR activates the FANCD2-FANCI clamp by priming it for ubiquitination. *Nature Structural & Molecular Biology* **29.** Publisher: Nature Publishing Group, 881–890 (2022).

434. Garcia-Higuera, I. *et al.* Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Molecular Cell* **7,** 249–262 (2001).

435. Klein Douwel, D. *et al.* XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Molecular Cell* **54,** 460–471 (2014).

436. Joo, W. *et al.* Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway. *Science (New York, N.Y.)* **333,** 312–316 (2011).

437. Wang, S., Wang, R., Peralta, C., Yaseen, A. & Pavletich, N. P. Structure of the FA core ubiquitin ligase closing the ID clamp on DNA. *Nature Structural & Molecular Biology* **28,** 300–309 (2021).

438. Alcón, P. *et al.* FANCD2–FANCI surveys DNA and recognizes double- to single-stranded junctions. *Nature* **632.** Publisher: Nature Publishing Group, 1165–1173 (2024).

439. Li, L., Tan, W. & Deans, A. J. Structural insight into FANCI–FANCD2 monoubiquitination. *Essays in Biochemistry* **64,** 807–817 (2020).

440. Tian, Y. *et al.* UHRF1 Contributes to DNA Damage Repair as a Lesion Recognition Factor and Nuclease Scaffold. *Cell reports* **10,** 1957–1966 (2015).

441. De Silva, I. U., McHugh, P. J., Clingen, P. H. & Hartley, J. A. Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells. *Molecular and Cellular Biology* **20,** 7980–7990 (2000).

442. Yamamoto, K. N. *et al.* Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. *Proceedings of the National Academy of Sciences* **108.** Publisher: Proceedings of the National Academy of Sciences, 6492–6496 (2011).

443. Lachaud, C. *et al.* Distinct functional roles for the two SLX4 ubiquitin-binding UBZ domains mutated in Fanconi anemia. *Journal of Cell Science* **127,** 2811–2817 (2014).

444. Abdullah, U. B. *et al.* RPA activates the XPF-ERCC1 endonuclease to initiate processing of DNA interstrand crosslinks. *The EMBO journal* **36,** 2047–2060 (2017).

445. Wang, A. T. *et al.* Human SNM1A and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes & Development* **25,** 1859–1870 (2011).

446. Olivieri, M. *et al.* A Genetic Map of the Response to DNA Damage in Human Cells. *Cell* **182,** 481–496.e21 (2020).

447. Fontebasso, Y., Etheridge, T. J., Oliver, A. W., Murray, J. M. & Carr, A. M. The conserved Fanconi anemia nuclease Fan1 and the SUMO E3 ligase Pli1 act in two novel Pso2-independent pathways of DNA interstrand crosslink repair in yeast. *DNA Repair* **12,** 1011–1023 (2013).

448. Dronkert, M. L. *et al.* Disruption of mouse SNM1 causes increased sensitivity to the DNA interstrand cross-linking agent mitomycin C. *Molecular and Cellular Biology* **20,** 4553–4561 (2000).

449. Räschle, M. *et al.* Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell* **134,** 969–980 (2008).

450. Budzowska, M., Graham, T. G. W., Sobeck, A., Waga, S. & Walter, J. C. Regulation of the Rev1-pol complex during bypass of a DNA interstrand cross-link. *The EMBO journal* **34,** 1971–1985 (2015).

451. Roy, U. & Schärer, O. D. Involvement of translesion synthesis DNA polymerases in DNA interstrand crosslink repair. *DNA repair* **44,** 33–41 (2016).

452. Chun, J., Buechelmaier, E. S. & Powell, S. N. Rad51 Paralog Complexes BCDX2 and CX3 Act at Different Stages in the BRCA1-BRCA2-Dependent Homologous Recombination Pathway. *Molecular and Cellular Biology* **33,** 387–395 (2013).

453. Kim, H. & D'Andrea, A. D. Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes & Development* **26,** 1393–1408 (2012).

454. Elia, A. E. H. *et al.* RFWD3-Dependent Ubiquitination of RPA Regulates Repair at Stalled Replication Forks. *Molecular Cell* **60.** Publisher: Elsevier, 280–293 (2015).

455. Peng, M., Litman, R., Jin, Z., Fong, G. & Cantor, S. B. BACH1 is a DNA repair protein supporting BRCA1 damage response. *Oncogene* **25.** Publisher: Nature Publishing Group, 2245–2253 (2006).

456. Wilson, D. M. & Seidman, M. M. A Novel Link to Base Excision Repair. *Trends in biochemical sciences* **35,** 247–252 (2010).

457. Muniandy, P. A., Liu, J., Majumdar, A., Liu, S.-t. & Seidman, M. M. DNA interstrand crosslink repair in mammalian cells: step by step. *Critical Reviews in Biochemistry and Molecular Biology* **45,** 23–49 (2010).

458. Huang, J. *et al.* The DNA Translocase FANCM/MHF Promotes Replication Traverse of DNA Interstrand Crosslinks. *Molecular Cell* **52.** Publisher: Elsevier, 434–446 (2013).

459. Huang, J. *et al.* Remodeling of Interstrand Crosslink Proximal Replisomes Is Dependent on ATR, FANCM, and FANCD2. *Cell Reports* **27.** Publisher: Elsevier, 1794–1808.e5 (2019).

460. Ling, C. *et al.* Bloom syndrome complex promotes FANCM recruitment to stalled replication forks and facilitates both repair and traverse of DNA interstrand crosslinks. *Cell Discovery* **2.** Publisher: Nature Publishing Group, 1–18 (2016).

461. Sparks, J. L. *et al.* The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* **176,** 167–181.e21 (2019).

462. Mutreja, K. *et al.* ATR-Mediated Global Fork Slowing and Reversal Assist Fork Traverse and Prevent Chromosomal Breakage at DNA Interstrand Cross-Links. *Cell Reports* **24.** Publisher: Elsevier, 2629–2642.e5 (2018).

463. Kühbacher, U. & Duxin, J. P. How to fix DNA-protein crosslinks. *DNA repair* **94,** 102924 (2020).

464. Lu, K. *et al.* Structural Characterization of Formaldehyde-induced Cross-links Between Amino Acids and Deoxynucleosides and Their Oligomers. *Journal of the American Chemical Society* **132,** 3388–3399 (2010).

465. Hoffman, E. A., Frey, B. L., Smith, L. M. & Auble, D. T. Formaldehyde crosslinking: a tool for the study of chromatin complexes. *The Journal of Biological Chemistry* **290,** 26404–26411 (2015).

466. Mosbech, A. *et al.* DVC1 (C1orf124) is a DNA damage-targeting p97 adaptor that promotes ubiquitin-dependent responses to replication blocks. *Nature Structural & Molecular Biology* **19,** 1084–1092 (2012).

467. Stingele, J. *et al.* Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. *Molecular Cell* **64,** 688–703 (2016).

468. Gujar, H., Weisenberger, D. J. & Liang, G. The Roles of Human DNA Methyltransferases and Their Isoforms in Shaping the Epigenome. *Genes* **10,** 172 (2019).

469. Wu, J. C. & Santi, D. V. Kinetic and catalytic mechanism of HhaI methyltransferase. *The Journal of Biological Chemistry* **262,** 4778–4786 (1987).

470. Meyn, R. E., vanAnkeren, S. C. & Jenkins, W. T. The induction of DNA-protein crosslinks in hypoxic cells and their possible contribution to cell lethality. *Radiation Research* **109,** 419–429 (1987).

471. Mohni, K. N. *et al.* HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell* **176,** 144–153.e13 (2019).

472. Thompson, P. S., Amidon, K. M., Mohni, K. N., Cortez, D. & Eichman, B. F. Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. *Nature Structural & Molecular Biology* **26,** 613–618 (2019).

473. Pommier, Y. DNA topoisomerase I inhibitors: chemistry, biology, and interfacial inhibition. *Chemical Reviews* **109,** 2894–2902 (2009).

474. Bandele, O. J. & Osheroff, N. Cleavage of plasmid DNA by eukaryotic topoisomerase II. *Methods in Molecular Biology (Clifton, N.J.)* **582,** 39–47 (2009).

475. Staker, B. L. *et al.* The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 15387–15392 (2002).

476. Takashima, H. *et al.* Mutation of TDP1, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy. *Nature Genetics* **32,** 267–272 (2002).

477. Thompson, P. S. & Cortez, D. New Insights into Abasic Site Repair and Tolerance. *DNA repair* **90,** 102866 (2020).

478. Kanev, P.-B. *et al.* A unified mechanism for PARP inhibitor-induced PARP1 chromatin retention at DNA damage sites in living cells. *Cell Reports* **43,** 114234 (2024).

479. Riccio, A. A., Schellenberg, M. J. & Williams, R. S. Molecular mechanisms of topoisomerase 2 DNA-protein crosslink resolution. *Cellular and molecular life sciences: CMLS* **77,** 81–91 (2020).

480. Nitiss, J. L. Targeting DNA topoisomerase II in cancer chemotherapy. *Nature reviews. Cancer* **9,** 338–350 (2009).

481. Nitiss, J. L. & Nitiss, K. C. Tdp2: A Means to Fixing the Ends. *PLOS Genetics* **9.** Publisher: Public Library of Science, e1003370 (2013).

482. Paiano, J. *et al.* ATM and PRDM9 regulate SPO11-bound recombination intermediates during meiosis. *Nature Communications* **11,** 857 (2020).

483. Neale, M. J., Pan, J. & Keeney, S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436,** 1053–1057 (2005).

484. Duxin, J. P., Dewar, J. M., Yardimci, H. & Walter, J. C. Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell* **159,** 346–357 (2014).

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

486. Lopez-Mosqueda, J. *et al.* SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. *eLife* **5,** e21491 (2016).

487. Perry, M. & Ghosal, G. Mechanisms and Regulation of DNA-Protein Crosslink Repair During DNA Replication by SPRTN Protease. *Frontiers in Molecular Biosciences* **9,** 916697 (2022).

488. Priego Moreno, S., Jones, R. M., Poovathumkadavil, D., Scaramuzza, S. & Gambus, A. Mitotic replisome disassembly depends on TRAIP ubiquitin ligase activity. *Life Science Alliance* **2,** e201900390 (2019).

489. Sonneville, R. *et al.* TRAIP drives replisome disassembly and mitotic DNA repair synthesis at sites of incomplete DNA replication. *eLife* **8,** e48686 (2019).

490. Larsen, N. B. *et al.* Replication-Coupled DNA-Protein Crosslink Repair by SPRTN and the Proteasome in *Xenopus* Egg Extracts. *Molecular Cell* **73,** 574–588.e7 (2019).

491. Centore, R. C., Yazinski, S. A., Tse, A. & Zou, L. Spartan/C1orf124, a reader of PCNA ubiquitylation and a regulator of UV-induced DNA damage response. *Molecular Cell* **46,** 625–635 (2012).

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

493. Fabijan, A. *et al.* Domains in Action: Understanding Ddi1's Diverse Functions in the Ubiquitin-Proteasome System. *International Journal of Molecular Sciences* **25,** 4080 (2024).

494. Kojima, Y. *et al.* FAM111A protects replication forks from protein obstacles via its trypsin-like domain. *Nature Communications* **11,** 1318 (2020).

495. Quievryn, G. & Zhitkovich, A. Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function. *Carcinogenesis* **21,** 1573–1580 (2000).

496. Rosado, I. V., Langevin, F., Crossan, G. P., Takata, M. & Patel, K. J. Formaldehyde catabolism is essential in cells deficient for the Fanconi anemia DNA-repair pathway. *Nature Structural & Molecular Biology* **18,** 1432–1434 (2011).

497. Reardon, J. T. & Sancar, A. Repair of DNA-polypeptide crosslinks by human excision nuclease. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 4056–4061 (2006).

498. Baker, D. J. *et al.* Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells. *The Journal of Biological Chemistry* **282,** 22592–22604 (2007).

499. Zhao, Y. *et al.* Applying genome-wide CRISPR to identify known and novel genes and pathways that modulate formaldehyde toxicity. *Chemosphere* **269,** 128701 (2021).

500. Gao, Y. *et al.* A CRISPR-Cas9 screen identifies EXO1 as a formaldehyde resistance gene. *Nature Communications* **14.** Publisher: Nature Publishing Group, 381 (2023).

501. Carnie, C. J. *et al.* Transcription-coupled repair of DNA–protein cross-links depends on CSA and CSB. *Nature Cell Biology* **26.** Publisher: Nature Publishing Group, 797–810 (2024).

502. Oka, Y., Nakazawa, Y., Shimada, M. & Ogi, T. Endogenous aldehyde-induced DNA–protein crosslinks are resolved by transcription-coupled repair. *Nature Cell Biology* **26.** Publisher: Nature Publishing Group, 784–796 (2024).

503. Van Sluis, M. *et al.* Transcription-coupled DNA–protein crosslink repair by CSB and CRL4CSA-mediated degradation. *Nature Cell Biology* **26.** Publisher: Nature Publishing Group, 770–783 (2024).

504. Burgos-Morón, E. *et al.* The Cockayne syndrome protein B is involved in the repair of 5-AZA-2'-deoxycytidine-induced DNA lesions. *Oncotarget* **9,** 35069–35084 (2018).

505. Sun, Y., Nitiss, J. L. & Pommier, Y. SUMO: A Swiss Army Knife for Eukaryotic Topoisomerases. *Frontiers in Molecular Biosciences* **9,** 871161 (2022).

506. Horton, J. K. *et al.* XRCC1-mediated repair of strand breaks independent of PNKP binding. *DNA repair* **60,** 52–63 (2017).

507. Zhang, Y.-W. *et al.* 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 Research* **39,** 3607–3620 (2011).

508. Schellenberg, M. J. *et al.* ZATT (ZNF451)-mediated resolution of topoisomerase 2 DNA-protein cross-links. *Science (New York, N.Y.)* **357,** 1412–1416 (2017).

509. Deshpande, R. A., Lee, J.-H., Arora, S. & Paull, T. T. Nbs1 Converts the Human Mre11/Rad50 Nuclease Complex into an Endo/Exonuclease Machine Specific for Protein-DNA Adducts. *Molecular Cell* **64,** 593–606 (2016).

510. Mullen, J. R., Kaliraman, V., Ibrahim, S. S. & Brill, S. J. Requirement for three novel protein complexes in the absence of the Sgs1 DNA helicase in Saccharomyces cerevisiae. *Genetics* **157,** 103–118 (2001).

511. Coulon, S. *et al.* Slx1-Slx4 are subunits of a structure-specific endonuclease that maintains ribosomal DNA in fission yeast. *Molecular Biology of the Cell* **15,** 71–80 (2004).

512. Fricke, W. M. & Brill, S. J. Slx1-Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1-Top3. *Genes & Development* **17,** 1768–1778 (2003).

513. Fekairi, S. *et al.* Human SLX4 Is a Holliday Junction Resolvase Subunit that Binds Multiple DNA Repair/Recombination Endonucleases. *Cell* **138,** 78–89 (2009).

514. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences* **98.** Publisher: Proceedings of the National Academy of Sciences, 4569–4574 (2001).

515. Flott, S. *et al.* Phosphorylation of Slx4 by Mec1 and Tel1 regulates the single-strand annealing mode of DNA repair in budding yeast. *Molecular and Cellular Biology* **27,** 6433–6445 (2007).

516. Andersen, S. L. *et al.* Drosophila MUS312 and the vertebrate ortholog BTBD12 interact with DNA structure-specific endonucleases in DNA repair and recombination. *Molecular Cell* **35,** 128–135 (2009).

517. Muñoz, I. M. *et al.* Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. *Molecular Cell* **35,** 116–127 (2009).

518. Nishino, T., Komori, K., Ishino, Y. & Morikawa, K. X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes. *Structure (London, England: 1993)* **11,** 445–457 (2003).

519. Dunin-Horkawicz, S., Feder, M. & Bujnicki, J. M. Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC genomics* **7,** 98 (2006).

520. Li, F. *et al.* Role of Saw1 in Rad1/Rad10 complex assembly at recombination intermediates in budding yeast. *The EMBO journal* **32,** 461–472 (2013).

521. Xu, X. *et al.* Structure specific DNA recognition by the SLX1–SLX4 endonuclease complex. *Nucleic Acids Research* **49,** 7740–7752 (2021).

522. Payliss, B. J., Patel, A., Sheppard, A. C. & Wyatt, H. D. M. Exploring the Structures and Functions of Macromolecular SLX4-Nuclease Complexes in Genome Stability. *Frontiers in Genetics* **12.** Publisher: Frontiers (2021).

523. Yin, J. *et al.* Dimerization of SLX4 contributes to functioning of the SLX4-nuclease complex. *Nucleic Acids Research* **44,** 4871–4880 (2016).

524. Kim, Y. *et al.* Regulation of multiple DNA repair pathways by the Fanconi anemia protein SLX4. *Blood* **121,** 54–63 (2013).

525. Hoogenboom, W. S., Boonen, R. A. C. M. & Knipscheer, P. The role of SLX4 and its associated nucleases in DNA interstrand crosslink repair. *Nucleic Acids Research* **47,** 2377–2388 (2019).

526. Landwehr, R. *et al.* Mutation analysis of the SLX4/FANCP gene in hereditary breast cancer. *Breast Cancer Research and Treatment* **130,** 1021–1028 (2011).

527. Guervilly, J.-H. & Gaillard, P. H. SLX4: multitasking to maintain genome stability. *Critical Reviews in Biochemistry and Molecular Biology* **53.** Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10409238.2018.1488803, 475–514 (2018).

528. González-Prieto, R., Cuijpers, S. A. G., Luijsterburg, M. S., van Attikum, H. & Vertegaal, A. C. O. SUMOylation and PARylation cooperate to recruit and stabilize SLX4 at DNA damage sites. *EMBO reports* **16,** 512–519 (2015).

529. Ouyang, J. *et al.* Noncovalent interactions with SUMO and ubiquitin orchestrate distinct functions of the SLX4 complex in genome maintenance. *Molecular Cell* **57,** 108–122 (2015).

530. Guervilly, J.-H. & Gaillard, P.-H. L. SLX4 gains weight with SUMO in genome maintenance. *Molecular & Cellular Oncology* **3,** e1008297 (2016).

531. Guervilly, J.-H. *et al.* The SLX4 complex is a SUMO E3 ligase that impacts on replication stress outcome and genome stability. *Molecular Cell* **57,** 123–137 (2015).

532. Hertz, E. P. T. *et al.* The SUMO-NIP45 pathway processes toxic DNA catenanes to prevent mitotic failure. *Nature Structural & Molecular Biology* **30,** 1303–1313 (2023).

533. Lian, F.-M., Xie, S. & Qian, C. Crystal structure and SUMO binding of Slx1-Slx4 complex. *Scientific Reports* **6,** 19331 (2016).

534. Sarangi, P. *et al.* A versatile scaffold contributes to damage survival via sumoylation and nuclease interactions. *Cell Reports* **9,** 143–152 (2014).

535. Cussiol, J. R., Dibitetto, D., Pellicioli, A. & Smolka, M. B. Slx4 Scaffolding in Homologous Recombination and Checkpoint Control: Lessons From Yeast. *Chromosoma* **126,** 45 (2016).

536. Dibitetto, D. *et al.* Slx4 and Rtt107 control checkpoint signalling and DNA resection at double-strand breaks. *Nucleic Acids Research* **44,** 669–682 (2016).

537. Liu, Y. *et al.* TOPBP1Dpb11 plays a conserved role in homologous recombination DNA repair through the coordinated recruitment of 53BP1Rad9. *The Journal of Cell Biology* **216,** 623–639 (2017).

538. Crossan, G. P. *et al.* Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia. *Nature Genetics* **43,** 147–152 (2011).

539. Holloway, J. K. *et al.* Mammalian BTBD12 (SLX4) protects against genomic instability during mammalian spermatogenesis. *PLoS genetics* **7,** e1002094 (2011).

540. Castor, D. *et al.* Cooperative control of holliday junction resolution and DNA repair by the SLX1 and MUS81-EME1 nucleases. *Molecular Cell* **52,** 221–233 (2013).

541. Hodskinson, M. R. G. *et al.* Mouse SLX4 is a tumor suppressor that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair. *Molecular Cell* **54,** 472–484 (2014).

542. Schuster, B. *et al.* Whole exome sequencing reveals uncommon mutations in the recently identified Fanconi anemia gene SLX4/FANCP. *Human Mutation* **34,** 93–96 (2013).

543. Kikuchi, K. *et al.* Structure-specific endonucleases xpf and mus81 play overlapping but essential roles in DNA repair by homologous recombination. *Cancer Research* **73,** 4362–4371 (2013).

544. Elango, R. *et al.* The structure-specific endonuclease complex SLX4–XPF regulates Tus–Ter-induced homologous recombination. *Nature Structural & Molecular Biology* **29.** Publisher: Nature Publishing Group, 801–812 (2022).

545. Toh, G. W.-L. *et al.* Mec1/Tel1-dependent phosphorylation of Slx4 stimulates Rad1-Rad10-dependent cleavage of non-homologous DNA tails. *DNA repair* **9,** 718–726 (2010).

546. Sugawara, N., Pâques, F., Colaiácovo, M. & Haber, J. E. Role of Saccharomyces cerevisiae Msh2 and Msh3 repair proteins in double-strand break-induced recombination. *Proceedings of the National Academy of Sciences of the United States of America* **94,** 9214–9219 (1997).

547. Zhang, H. *et al.* SLX4IP acts with SLX4 and XPF-ERCC1 to promote interstrand crosslink repair. *Nucleic Acids Research* **47,** 10181–10201 (2019).

548. Ait Saada, A. *et al.* Widely spaced and divergent inverted repeats become a potent source of chromosomal rearrangements in long single-stranded DNA regions. *Nucleic Acids Research* **51,** 3722–3734 (2023).

549. Young, S. J. & West, S. C. Coordinated roles of SLX4 and MutS in DNA repair and the maintenance of genome stability. *Critical Reviews in Biochemistry and Molecular Biology* **56,** 157–177 (2021).

550. Guervilly, J.-H. *et al.* SLX4 dampens MutS-dependent mismatch repair. *Nucleic Acids Research* **50,** 2667–2680 (2022).

551. Young, S. J. *et al.* MutS Stimulates Holliday Junction Resolution by the SMX Complex. *Cell Reports* **33,** 108289 (2020).

552. Wu, L. & Hickson, I. D. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* **426,** 870–874 (2003).

553. Rass, U. *et al.* Mechanism of Holliday junction resolution by the human GEN1 protein. *Genes & Development* **24,** 1559–1569 (2010).

554. Wyatt, H. D. M., Sarbajna, S., Matos, J. & West, S. C. Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells. *Molecular Cell* **52,** 234–247 (2013).

555. Yildiz, O., Majumder, S., Kramer, B. & Sekelsky, J. J. Drosophila MUS312 interacts with the nucleotide excision repair endonuclease MEI-9 to generate meiotic crossovers. *Molecular Cell* **10,** 1503–1509 (2002).

556. Trowbridge, K., McKim, K., Brill, S. J. & Sekelsky, J. Synthetic lethality of Drosophila in the absence of the MUS81 endonuclease and the DmBlm helicase is associated with elevated apoptosis. *Genetics* **176,** 1993–2001 (2007).

557. Saito, T. T., Youds, J. L., Boulton, S. J. & Colaiácovo, M. P. Caenorhabditis elegans HIM-18/SLX-4 interacts with SLX-1 and XPF-1 and maintains genomic integrity in the germline by processing recombination intermediates. *PLoS genetics* **5,** e1000735 (2009).

558. Agostinho, A. *et al.* Combinatorial Regulation of Meiotic Holliday Junction Resolution in C. elegans by HIM-6 (BLM) Helicase, SLX-4, and the SLX-1, MUS-81 and XPF-1 Nucleases. *PLOS Genetics* **9.** Publisher: Public Library of Science, e1003591 (2013).

559. Saito, T. T., Lui, D. Y., Kim, H.-M., Meyer, K. & Colaiácovo, M. P. Interplay between structure-specific endonucleases for crossover control during Caenorhabditis elegans meiosis. *PLoS genetics* **9,** e1003586 (2013).

560. Saito, T. T. & Colaiácovo, M. P. Crossover recombination mediated by HIM-18/SLX4-associated nucleases. *Worm* **3,** e28233 (2014).

561. De Muyt, A. *et al.* BLM helicase ortholog Sgs1 is a central regulator of meiotic recombination intermediate metabolism. *Molecular Cell* **46,** 43–53 (2012).

562. Zakharyevich, K., Tang, S., Ma, Y. & Hunter, N. Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase. *Cell* **149,** 334–347 (2012).

563. Ranjha, L., Anand, R. & Cejka, P. The Saccharomyces cerevisiae Mlh1-Mlh3 heterodimer is an endonuclease that preferentially binds to Holliday junctions. *The Journal of Biological Chemistry* **289,** 5674–5686 (2014).

564. Manhart, C. M. *et al.* The mismatch repair and meiotic recombination endonuclease Mlh1-Mlh3 is activated by polymer formation and can cleave DNA substrates in trans. *PLoS biology* **15,** e2001164 (2017).

565. Garner, E., Kim, Y., Lach, F. P., Kottemann, M. C. & Smogorzewska, A. Human GEN1 and the SLX4-associated nucleases MUS81 and SLX1 are essential for the resolution of replication-induced Holliday junctions. *Cell Reports* **5,** 207–215 (2013).

566. Boddy, M. N. *et al.* Mus81-Eme1 are essential components of a Holliday junction resolvase. *Cell* **107,** 537–548 (2001).

567. Smith, G. R., Boddy, M. N., Shanahan, P. & Russell, P. Fission yeast Mus81.Eme1 Holliday junction resolvase is required for meiotic crossing over but not for gene conversion. *Genetics* **165,** 2289–2293 (2003).

568. Ho, C. K., Mazón, G., Lam, A. F. & Symington, L. S. Mus81 and Yen1 promote reciprocal exchange during mitotic recombination to maintain genome integrity in budding yeast. *Molecular Cell* **40,** 988–1000 (2010).

569. Tay, Y. D. & Wu, L. Overlapping roles for Yen1 and Mus81 in cellular Holliday junction processing. *The Journal of Biological Chemistry* **285,** 11427–11432 (2010).

570. Matos, J., Blanco, M. G., Maslen, S., Skehel, J. M. & West, S. C. Regulatory control of the resolution of DNA recombination intermediates during meiosis and mitosis. *Cell* **147,** 158–172 (2011).

571. Patel, A. *Investigating the temporal dynamics of SMX assembly and disassembly throughout the cell cycle* PhD thesis (University of Toronto, 2022).

572. Matos, J., Blanco, M. G. & West, S. C. Cell-cycle kinases coordinate the resolution of recombination intermediates with chromosome segregation. *Cell Reports* **4,** 76–86 (2013).

573. Szakal, B. & Branzei, D. Premature Cdk1/Cdc5/Mus81 pathway activation induces aberrant replication and deleterious crossover. *The EMBO journal* **32,** 1155–1167 (2013).

574. Gallo-Fernández, M., Saugar, I., Ortiz-Bazán, M. Á., Vázquez, M. V. & Tercero, J. A. Cell cycle-dependent regulation of the nuclease activity of Mus81-Eme1/Mms4. *Nucleic Acids Research* **40,** 8325–8335 (2012).

575. Dehé, P.-M. *et al.* Regulation of Mus81-Eme1 Holliday junction resolvase in response to DNA damage. *Nature Structural & Molecular Biology* **20,** 598–603 (2013).

576. Duda, H. *et al.* A Mechanism for Controlled Breakage of Under-replicated Chromosomes during Mitosis. *Developmental Cell* **39,** 740–755 (2016).

577. Payliss, B. J. *et al.* Phosphorylation of the DNA repair scaffold SLX4 drives folding of the SAP domain and activation of the MUS81-EME1 endonuclease. *Cell Reports* **41,** 111537 (2022).

578. Payliss, B. J. & Wyatt, H. D. M. Protocol for in vitro phosphorylation of the MUS81-binding region of SLX4 using CDK1-cyclin B. *STAR protocols* **4,** 102152 (2023).

579. Palma, A. *et al.* Phosphorylation by CK2 regulates MUS81/EME1 in mitosis and after replication stress. *Nucleic Acids Research* **46,** 5109–5124 (2018).

580. Seol, J.-H. *et al.* Distinct roles of XPF-ERCC1 and Rad1-Rad10-Saw1 in replication-coupled and uncoupled inter-strand crosslink repair. *Nature Communications* **9,** 2025 (2018).

581. Knipscheer, P. *et al.* The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science (New York, N.Y.)* **326,** 1698–1701 (2009).

582. Stoepker, C. *et al.* SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. *Nature Genetics* **43,** 138–141 (2011).

583. Katsuki, Y. *et al.* RNF168 E3 ligase participates in ubiquitin signaling and recruitment of SLX4 during DNA crosslink repair. *Cell Reports* **37,** 109879 (2021).

584. Mattiroli, F. *et al.* RNF168 Ubiquitinates K13-15 on H2A/H2AX to Drive DNA Damage Signaling. *Cell* **150.** Publisher: Elsevier, 1182–1195 (2012).

585. Dehé, P.-M. & Gaillard, P.-H. L. Control of structure-specific endonucleases to maintain genome stability. *Nature Reviews. Molecular Cell Biology* **18,** 315–330 (2017).

586. Hashimoto, K., Wada, K., Matsumoto, K. & Moriya, M. Physical interaction between SLX4 (FANCP) and XPF (FANCQ) proteins and biological consequences of interaction-defective missense mutations. *DNA repair* **35,** 48–54 (2015).

587. Klein Douwel, D., Hoogenboom, W. S., Boonen, R. A. & Knipscheer, P. Recruitment and positioning determine the specific role of the XPF-ERCC1 endonuclease in interstrand crosslink repair. *The EMBO journal* **36,** 2034–2046 (2017).

588. Kuraoka, I. *et al.* Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease. *The Journal of Biological Chemistry* **275,** 26632–26636 (2000).

589. Fisher, L. A., Bessho, M. & Bessho, T. Processing of a psoralen DNA interstrand cross-link by XPF-ERCC1 complex in vitro. *The Journal of Biological Chemistry* **283,** 1275–1281 (2008).

590. Sengerová, B. *et al.* Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. *The Journal of Biological Chemistry* **287,** 26254–26267 (2012).

591. Salewsky, B., Schmiester, M., Schindler, D., Digweed, M. & Demuth, I. The nuclease hSNM1B/Apollo is linked to the Fanconi anemia pathway via its interaction with FANCP/SLX4. *Human Molecular Genetics* **21,** 4948–4956 (2012).

592. Abraham, J. *et al.* Eme1 is involved in DNA damage processing and maintenance of genomic stability in mammalian cells. *The EMBO journal* **22,** 6137–6147 (2003).

593. McPherson, J. P. *et al.* Involvement of mammalian Mus81 in genome integrity and tumor suppression. *Science (New York, N.Y.)* **304,** 1822–1826 (2004).

594. Hanada, K. *et al.* The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks. *The EMBO journal* **25,** 4921–4932 (2006).

595. Hiyama, T. *et al.* Haploinsufficiency of the Mus81-Eme1 endonuclease activates the intra-S-phase and G2/M checkpoints and promotes rereplication in human cells. *Nucleic Acids Research* **34,** 880–892 (2006).

596. Nair, N., Castor, D., Macartney, T. & Rouse, J. Identification and characterization of MUS81 point mutations that abolish interaction with the SLX4 scaffold protein. *DNA repair* **24,** 131–137 (2014).

597. Jung, Y. & Lippard, S. J. Direct cellular responses to platinum-induced DNA damage. *Chemical Reviews* **107,** 1387–1407 (2007).

598. Pommier, Y. Topoisomerase I inhibitors: camptothecins and beyond. *Nature Reviews. Cancer* **6,** 789–802 (2006).

599. Koster, D. A., Palle, K., Bot, E. S. M., Bjornsti, M.-A. & Dekker, N. H. Antitumour drugs impede DNA uncoiling by topoisomerase I. *Nature* **448,** 213–217 (2007).

600. Regairaz, M. *et al.* Mus81-mediated DNA cleavage resolves replication forks stalled by topoisomerase I-DNA complexes. *The Journal of Cell Biology* **195,** 739–749 (2011).

601. Wu, X. & Wang, B. Abraxas suppresses DNA end resection and limits break-induced replication by controlling SLX4/MUS81 chromatin loading in response to TOP1 inhibitor-induced DNA damage. *Nature Communications* **12,** 4373 (2021).

602. Ishimoto, R. *et al.* SLX4-XPF mediates DNA damage responses to replication stress induced by DNA-protein interactions. *The Journal of Cell Biology* **220,** e202003148 (2021).

603. Dungrawala, H. *et al.* The Replication Checkpoint Prevents Two Types of Fork Collapse without Regulating Replisome Stability. *Molecular Cell* **59,** 998–1010 (2015).

604. Alghoul, E. *et al.* Compartmentalization of the SUMO/RNF4 pathway by SLX4 drives DNA repair. *Molecular Cell* **83.** Publisher: Elsevier, 1640–1658.e9 (2023).

605. Chen, S., Pan, C., Huang, J. & Liu, T. ATR limits Rad18-mediated PCNA monoubiquitination to preserve replication fork and telomerase-independent telomere stability. *The EMBO journal* **43,** 1301–1324 (2024).

606. Hanada, K. *et al.* The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. *Nature Structural & Molecular Biology* **14,** 1096–1104 (2007).

607. Pepe, A. & West, S. C. MUS81-EME2 promotes replication fork restart. *Cell Reports* **7,** 1048–1055 (2014).

608. Malacaria, E., Franchitto, A. & Pichierri, P. SLX4 Prevents GEN1-Dependent DSBs During DNA Replication Arrest Under Pathological Conditions in Human Cells. *Scientific Reports* **7,** 44464 (2017).

609. Xu, Y. *et al.* 53BP1 and BRCA1 control pathway choice for stalled replication restart. *eLife* **6,** e30523 (2017).

610. Tsukada, K. *et al.* BLM and BRCA1-BARD1 coordinate complementary mechanisms of joint DNA molecule resolution. *Molecular Cell* **84.** Publisher: Elsevier, 640–658.e10 (2024).

611. Fugger, K. *et al.* FBH1 co-operates with MUS81 in inducing DNA double-strand breaks and cell death following replication stress. *Nature Communications* **4,** 1423 (2013).

612. Couch, F. B. *et al.* ATR phosphorylates SMARCAL1 to prevent replication fork collapse. *Genes & Development* **27,** 1610–1623 (2013).

613. Wang, Y. *et al.* Homolog of Saccharomyces cerevisiae SLX4 is required for cell recovery from MMS-induced DNA damage in Candida albicans. *FEMS yeast research* **21,** foab010 (2021).

614. Ragland, R. L. *et al.* RNF4 and PLK1 are required for replication fork collapse in ATR-deficient cells. *Genes & Development* **27,** 2259–2273 (2013).

615. Somyajit, K., Saxena, S., Babu, S., Mishra, A. & Nagaraju, G. Mammalian RAD51 paralogs protect nascent DNA at stalled forks and mediate replication restart. *Nucleic Acids Research* **43,** 9835–9855 (2015).

616. Panichnantakul, P., Patel, A., Tse, E. Y. W. & Wyatt, H. D. M. An open-source platform to quantify subnuclear foci and protein colocalization in response to replication stress. *DNA repair* **105,** 103156 (2021).

617. Bétous, R. *et al.* SMARCAL1 catalyzes fork regression and Holliday junction migration to maintain genome stability during DNA replication. *Genes & Development* **26,** 151–162 (2012).

618. Neelsen, K. J., Zanini, I. M. Y., Herrador, R. & Lopes, M. Oncogenes induce genotoxic stress by mitotic processing of unusual replication intermediates. *The Journal of Cell Biology* **200,** 699–708 (2013).

619. Thompson, R., Montano, R. & Eastman, A. The Mre11 nuclease is critical for the sensitivity of cells to Chk1 inhibition. *PloS One* **7,** e44021 (2012).

620. Lai, X. *et al.* MUS81 nuclease activity is essential for replication stress tolerance and chromosome segregation in BRCA2-deficient cells. *Nature Communications* **8,** 15983 (2017).

621. Rondinelli, B. *et al.* EZH2 promotes degradation of stalled replication forks by recruiting MUS81 through histone H3 trimethylation. *Nature Cell Biology* **19,** 1371–1378 (2017).

622. Beck, H. *et al.* Cyclin-dependent kinase suppression by WEE1 kinase protects the genome through control of replication initiation and nucleotide consumption. *Molecular and Cellular Biology* **32,** 4226–4236 (2012).

623. Ohouo, P. Y., Bastos de Oliveira, F. M., Almeida, B. S. & Smolka, M. B. DNA damage signaling recruits the Rtt107-Slx4 scaffolds via Dpb11 to mediate replication stress response. *Molecular Cell* **39,** 300–306 (2010).

624. Ohouo, P. Y., Bastos de Oliveira, F. M., Liu, Y., Ma, C. J. & Smolka, M. B. DNA-repair scaffolds dampen checkpoint signalling by counteracting the adaptor Rad9. *Nature* **493,** 120–124 (2013).

625. Balint, A. *et al.* Assembly of Slx4 signaling complexes behind DNA replication forks. *The EMBO journal* **34,** 2182–2197 (2015).

626. Majka, J. & Burgers, P. M. J. Yeast Rad17/Mec3/Ddc1: A sliding clamp for the DNA damage checkpoint. *Proceedings of the National Academy of Sciences* **100.** Publisher: Proceedings of the National Academy of Sciences, 2249–2254 (2003).

627. Cussiol, J. R., Jablonowski, C. M., Yimit, A., Brown, G. W. & Smolka, M. B. Dampening DNA damage checkpoint signalling via coordinated BRCT domain interactions. *The EMBO journal* **34,** 1704–1717 (2015).

628. Gritenaite, D. *et al.* A cell cycle-regulated Slx4-Dpb11 complex promotes the resolution of DNA repair intermediates linked to stalled replication. *Genes & Development* **28,** 1604–1619 (2014).

629. Zhao, X., Wan, B., Guan, D., Li, S. & Chwat-Edelstein, T. The Mms22-Rtt107 axis dampens the DNA damage checkpoint by reducing the stability of the Rad9 checkpoint mediator. *Research Square,* rs.3.rs–4417144 (2024).

630. Princz, L. N. *et al.* Dbf4-dependent kinase and the Rtt107 scaffold promote Mus81-Mms4 resolvase activation during mitosis. *The EMBO journal* **36,** 664–678 (2017).

631. Martin, P. R. *et al. The mitotic CIP2A-TOPBP1 axis facilitates mitotic pathway choice between MiDAS and MMEJ* Pages: 2024.11.12.621593 Section: New Results. 2024.

632. Minocherhomji, S. *et al.* Replication stress activates DNA repair synthesis in mitosis. *Nature* **528,** 286–290 (2015).

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

634. Min, J., Wright, W. E. & Shay, J. W. Alternative Lengthening of Telomeres Mediated by Mitotic DNA Synthesis Engages Break-Induced Replication Processes. *Molecular and Cellular Biology* **37,** e00226–17 (2017).

635. Özer, Ö., Bhowmick, R., Liu, Y. & Hickson, I. D. Human cancer cells utilize mitotic DNA synthesis to resist replication stress at telomeres regardless of their telomere maintenance mechanism. *Oncotarget* **9,** 15836–15846 (2018).

636. Pedersen, R. T., Kruse, T., Nilsson, J., Oestergaard, V. H. & Lisby, M. TopBP1 is required at mitosis to reduce transmission of DNA damage to G1 daughter cells. *The Journal of Cell Biology* **210,** 565–582 (2015).

637. De Marco Zompit, M. *et al.* The CIP2A-TOPBP1 complex safeguards chromosomal stability during mitosis. *Nature Communications* **13.** Publisher: Nature Publishing Group, 4143 (2022).

638. Mengoli, V. *et al.* WRN helicase and mismatch repair complexes independently and synergistically disrupt cruciform DNA structures. *The EMBO journal* **42,** e111998 (2023).

639. Garribba, L. *et al.* Folate stress induces SLX1- and RAD51-dependent mitotic DNA synthesis at the fragile X locus in human cells. *Proceedings of the National Academy of Sciences of the United States of America* **117,** 16527–16536 (2020).

640. Lange, T. d. Shelterin-Mediated Telomere Protection. *Annual Review of Genetics* **52.** Publisher: Annual Reviews, 223–247 (Volume 52, 2018 2018).

641. Maestroni, L., Matmati, S. & Coulon, S. Solving the Telomere Replication Problem. *Genes* **8.** Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, 55 (2017).

642. Doksani, Y., Wu, J. Y., de Lange, T. & Zhuang, X. Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. *Cell* **155,** 345–356 (2013).

643. Saint-Léger, A. *et al.* The basic N-terminal domain of TRF2 limits recombination endonuclease action at human telomeres. *Cell Cycle (Georgetown, Tex.)* **13,** 2469–2474 (2014).

644. Wan, B. *et al.* SLX4 assembles a telomere maintenance toolkit by bridging multiple endonucleases with telomeres. *Cell Reports* **4,** 861–869 (2013).

645. Sarkar, J. *et al.* SLX4 contributes to telomere preservation and regulated processing of telomeric joint molecule intermediates. *Nucleic Acids Research* **43,** 5912–5923 (2015).

646. Wilson, J. S. J. *et al.* Localization-dependent and -independent roles of SLX4 in regulating telomeres. *Cell Reports* **4,** 853–860 (2013).

647. Sobinoff, A. P. *et al.* Irreversible inhibition of TRF2TRFH recruiting functions by a covalent cyclic peptide induces telomeric replication stress in cancer cells. *Cell Chemical Biology* **30,** 1652–1665.e6 (2023).

648. Calado, R. T. & Dumitriu, B. Telomere dynamics in mice and humans. *Seminars in hematology* **50,** 165–174 (2013).

649. Poulet, A. *et al.* TRF2 promotes, remodels and protects telomeric Holliday junctions. *The EMBO journal* **28,** 641–651 (2009).

650. Schmutz, I., Timashev, L., Xie, W., Patel, D. J. & de Lange, T. TRF2 binds branched DNA to safeguard telomere integrity. *Nature Structural & Molecular Biology* **24,** 734–742 (2017).

651. Sadhukhan, R. & Ghosh, U. PARP1 modulates telomere sister chromatid exchange and telomere length homeostasis by regulating telomere localization of SLX4 in U2OS cells. *Life Sciences* **277,** 119556 (2021).

652. Rai, R., Chen, Y., Lei, M. & Chang, S. TRF2-RAP1 is required to protect telomeres from engaging in homologous recombination-mediated deletions and fusions. *Nature Communications* **7,** 10881 (2016).

653. Vannier, J.-B., Pavicic-Kaltenbrunner, V., Petalcorin, M. I. R., Ding, H. & Boulton, S. J. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell* **149,** 795–806 (2012).

654. Sarek, G., Vannier, J.-B., Panier, S., Petrini, J. H. J. & Boulton, S. J. TRF2 recruits RTEL1 to telomeres in S phase to promote t-loop unwinding. *Molecular Cell* **57,** 622–635 (2015).

655. Sobinoff, A. P. *et al.* BLM and SLX4 play opposing roles in recombination-dependent replication at human telomeres. *The EMBO journal* **36,** 2907–2919 (2017).

656. Poole, L. A. *et al.* SMARCAL1 maintains telomere integrity during DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 14864–14869 (2015).

657. Rivera, T., Haggblom, C., Cosconati, S. & Karlseder, J. A balance between elongation and trimming regulates telomere stability in stem cells. *Nature Structural & Molecular Biology* **24,** 30–39 (2017).

658. Guh, C.-Y. *et al.* XPF activates break-induced telomere synthesis. *Nature Communications* **13,** 5781 (2022).

659. Yang, Z., Takai, K. K., Lovejoy, C. A. & de Lange, T. Break-induced replication promotes fragile telomere formation. *Genes & Development* **34,** 1392–1405 (2020).

660. Crossley, M. P. *et al.* R-loop-derived cytoplasmic RNA-DNA hybrids activate an immune response. *Nature* **613,** 187–194 (2023).

661. Takedachi, A. *et al.* SLX4 interacts with RTEL1 to prevent transcription-mediated DNA replication perturbations. *Nature Structural & Molecular Biology* **27,** 438–449 (2020).

662. Laguette, N. *et al.* Premature activation of the SLX4 complex by Vpr promotes G2/M arrest and escape from innate immune sensing. *Cell* **156,** 134–145 (2014).

663. Brégnard, C. *et al.* Upregulated LINE-1 Activity in the Fanconi Anemia Cancer Susceptibility Syndrome Leads to Spontaneous Pro-inflammatory Cytokine Production. *EBioMedicine* **8,** 184–194 (2016).

664. Bona, N. & Crossan, G. P. Fanconi anemia DNA crosslink repair factors protect against LINE-1 retrotransposition during mouse development. *Nature Structural & Molecular Biology* **30,** 1434–1445 (2023).

665. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1804,** 996–1010 (2010).

666. Mészáros, B., Erdos, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research* **46,** W329–W337 (W1 2018).

667. Robinson, N. J. *et al.* SLX4IP and telomere dynamics dictate breast cancer metastasis and therapeutic responsiveness. *Life Science Alliance* **3,** e201900427 (2020).

668. Panier, S. *et al.* SLX4IP Antagonizes Promiscuous BLM Activity during ALT Maintenance. *Molecular Cell* **76,** 27–43.e11 (2019).

669. Robinson, N. J. *et al.* SLX4IP promotes RAP1 SUMOylation by PIAS1 to coordinate telomere maintenance through NF-B and Notch signaling. *Science Signaling* **14,** eabe9613 (2021).

670. Teo, H. *et al.* Telomere-independent Rap1 is an IKK adaptor and regulates NF-kappaB-dependent gene expression. *Nature Cell Biology* **12,** 758–767 (2010).

671. Greenhough, L. A. *et al.* Structure and function of the RAD51B-RAD51C-RAD51D-XRCC2 tumour suppressor. *Nature* **619,** 650–657 (2023).

672. Przanowski, P., Przanowska, R. K. & Guertin, M. J. ANKLE1 cleaves mitochondrial DNA and contributes to cancer risk by promoting apoptosis resistance and metabolic dysregulation. *Communications Biology* **6.** Publisher: Nature Publishing Group, 1–13 (2023).

673. Hong, Y. *et al.* The conserved LEM-3/Ankle1 nuclease is involved in the combinatorial regulation of meiotic recombination repair and chromosome segregation in Caenorhabditis elegans. *PLoS genetics* **14,** e1007453 (2018).

674. Verma, P., Kumari, P., Negi, S., Yadav, G. & Gaur, V. Holliday junction resolution by At-HIGLE: an SLX1 lineage endonuclease from Arabidopsis thaliana with a novel in-built regulatory mechanism. *Nucleic Acids Research* **50,** 4630–4646 (2022).

675. Gaur, V. *et al.* Structural and Mechanistic Analysis of the Slx1-Slx4 Endonuclease. *Cell Reports* **10,** 1467–1476 (2015).

676. Gaur, V. *et al.* Recognition and processing of branched DNA substrates by Slx1-Slx4 nuclease. *Nucleic Acids Research* **47,** 11681–11690 (2019).

677. Sokolowska, M., Czapinska, H. & Bochtler, M. Hpy188I-DNA pre- and post-cleavage complexes–snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Research* **39,** 1554–1564 (2011).

678. Lehmann, J. *et al.* Splice variants of the endonucleases XPF and XPG contain residual DNA repair capabilities and could be a valuable tool for personalized medicine. *Oncotarget* **9.** Publisher: Impact Journals, 1012–1027 (2017).

679. Pingoud, A., Fuxreiter, M., Pingoud, V. & Wende, W. Type II restriction endonucleases: structure and mechanism. *Cellular and molecular life sciences: CMLS* **62,** 685–707 (2005).

680. Ciccia, A., McDonald, N. & West, S. C. Structural and Functional Relationships of the XPF/MUS81 Family of Proteins. *Annual Review of Biochemistry* **77.** Publisher: Annual Reviews, 259–287 (Volume 77, 2008 2008).

681. Chen, S. *et al.* Human MUS81: A Fence-Sitter in Cancer. *Frontiers in Cell and Developmental Biology* **9.** Publisher: Frontiers (2021).

682. Xue, X., Sung, P. & Zhao, X. Functions and regulation of the multitasking FANCM family of DNA motor proteins. *Genes & Development* **29,** 1777–1788 (2015).

683. Wu, Y. Unwinding and Rewinding: Double Faces of Helicase? *Journal of Nucleic Acids* **2012,** 140601 (2012).

684. Komori, K., Fujikane, R., Shinagawa, H. & Ishino, Y. Novel endonuclease in Archaea cleaving DNA with various branched structure. *Genes & Genetic Systems* **77,** 227–241 (2002).

685. Newman, M. *et al.* Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. *The EMBO Journal* **24.** Num Pages: 905 Publisher: John Wiley & Sons, Ltd, 895–905 (2005).

686. Roberts, J. A. & White, M. F. An Archaeal Endonuclease Displays Key Properties of Both Eukaryal XPF-ERCC1 and Mus81 *. *Journal of Biological Chemistry* **280.** Publisher: Elsevier, 5924–5928 (2005).

687. Yu, J. *et al.* Molecular architecture and functional dynamics of the pre-incision complex in nucleotide excision repair. *Nature Communications* **15.** Publisher: Nature Publishing Group, 8511 (2024).

688. Zheng, X.-F. *et al.* Processing of DNA structures via DNA unwinding and branch migration by the S. cerevisiae Mph1 protein. *DNA repair* **10,** 1034–1043 (2011).

689. Sgouros, J., Gaillard, P.-H. L. & Wood, R. D. A relationship betweena DNA-repair/recombination nuclease family and archaeal helicases. *Trends in Biochemical Sciences* **24,** 95–97 (1999).

690. Aravind, L., Makarova, K. S. & Koonin, E. V. SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Research* **28,** 3417–3432 (2000).

691. Komori, K. *et al.* Cooperation of the N-terminal Helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks. *The Journal of Biological Chemistry* **279,** 53175–53185 (2004).

692. Nishino, T., Komori, K., Ishino, Y. & Morikawa, K. Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms. *Structure (London, England: 1993)* **13,** 1183–1192 (2005).

693. Tsodikov, O. V., Enzlin, J. H., Schärer, O. D. & Ellenberger, T. Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF–ERCC1. *Proceedings of the National Academy of Sciences* **102.** Publisher: Proceedings of the National Academy of Sciences, 11236–11241 (2005).

694. Tripsianes, K. *et al.* The Structure of the Human ERCC1/XPF Interaction Domains Reveals a Complementary Role for the Two Proteins in Nucleotide Excision Repair. *Structure* **13.** Publisher: Elsevier, 1849–1858 (2005).

695. Das, D. *et al.* The HhH domain of the human DNA repair protein XPF forms stable homodimers. *Proteins* **70,** 1551–1563 (2008).

696. Jones, M. *et al.* Cryo-EM structures of the XPF-ERCC1 endonuclease reveal how DNA-junction engagement disrupts an auto-inhibited conformation. *Nature Communications* **11.** Publisher: Nature Publishing Group, 1120 (2020).

697. Gwon, G. H. *et al.* Crystal structures of the structure-selective nuclease Mus81-Eme1 bound to flap DNA substrates. *The EMBO journal* **33,** 1061–1072 (2014).

698. Interthal, H. & Heyer, W.-D. MUS81 encodes a novel Helix-hairpin-Helix protein involved in the response to UV- and methylation-induced DNA damage in Saccharomyces cerevisiae. *Molecular and General Genetics MGG* **263,** 812–827 (2000).

699. Boddy, M. N. *et al.* Damage tolerance protein Mus81 associates with the FHA1 domain of checkpoint kinase Cds1. *Molecular and Cellular Biology* **20,** 8758–8766 (2000).

700. Chen, X. B. *et al.* Human Mus81-associated endonuclease cleaves Holliday junctions in vitro. *Molecular Cell* **8,** 1117–1127 (2001).

701. Ciccia, A., Constantinou, A. & West, S. C. Identification and characterization of the human mus81-eme1 endonuclease. *The Journal of Biological Chemistry* **278,** 25172–25178 (2003).

702. Kaliraman, V., Mullen, J. R., Fricke, W. M., Bastin-Shanower, S. A. & Brill, S. J. Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease. *Genes & Development* **15,** 2730–2740 (2001).

703. Fadden, A. J. *et al.* A winged helix domain in human MUS81 binds DNA and modulates the endonuclease activity of MUS81 complexes. *Nucleic Acids Research* **41,** 9741–9752 (2013).

704. Amangyeld, T., Shin, Y.-K., Lee, M., Kwon, B. & Seo, Y.-S. Human MUS81-EME2 can cleave a variety of DNA structures including intact Holliday junction and nicked duplex. *Nucleic Acids Research* **42,** 5846–5862 (2014).

705. Hua, Z. *et al.* Crystal structure of the human MUS81-EME2 complex. *Structure* **30,** 743–752.e3 (2022).

706. Awwad, S. W., Serrano-Benitez, A., Thomas, J. C., Gupta, V. & Jackson, S. P. Revolutionizing DNA repair research and cancer therapy with CRISPR–Cas screens. *Nature Reviews Molecular Cell Biology* **24.** Publisher: Nature Publishing Group, 477–494 (2023).

707. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).

708. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578,** 94–101 (2020).

709. Adler, N. *et al.* Mutational processes of tobacco smoking and APOBEC activity generate protein-truncating mutations in cancer genomes. *Science Advances* **9,** eadh3083 (2023).

710. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nature reviews. Genetics* **15,** 585–598 (2014).

711. Collins, A. *et al.* Measuring DNA modifications with the comet assay: a compendium of protocols. *Nature protocols* **18,** 929–989 (2023).

712. Chen, S., Lai, W. & Wang, H. Recent advances in high-performance liquid chromatography tandem mass spectrometry techniques for analysis of DNA damage and epigenetic modifications. *Mutation Research. Genetic Toxicology and Environmental Mutagenesis* **896,** 503755 (2024).

713. Mah, L.-J., El-Osta, A. & Karagiannis, T. C. H2AX: a sensitive molecular marker of DNA damage and repair. *Leukemia* **24.** Publisher: Nature Publishing Group, 679–686 (2010).

714. Zhang, L. *et al.* 53BP1 regulates heterochromatin through liquid phase separation. *Nature Communications* **13.** Publisher: Nature Publishing Group, 360 (2022).

715. Shi, W. *et al.* The role of RPA2 phosphorylation in homologous recombination in response to replication arrest. *Carcinogenesis* **31,** 994–1002 (2010).

716. Collins, P. L. *et al.* DNA double-strand breaks induce H2Ax phosphorylation domains in a contact-dependent manner. *Nature Communications* **11.** Publisher: Nature Publishing Group, 3158 (2020).

717. Rochette, P. J. & Brash, D. E. Human Telomeres Are Hypersensitive to UV-Induced DNA Damage and Refractory to Repair. *PLOS Genetics* **6.** Publisher: Public Library of Science, e1000926 (2010).

718. Wellinger, R. E. & Thoma, F. Nucleosome structure and positioning modulate nucleotide excision repair in the non-transcribed strand of an active gene. *The EMBO Journal* **16,** 5046–5056 (1997).

719. Mao, P., Smerdon, M. J., Roberts, S. A. & Wyrick, J. J. Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proceedings of the National Academy of Sciences* **113.** Publisher: Proceedings of the National Academy of Sciences, 9057–9062 (2016).

720. Mao, P. & Wyrick, J. J. Genome-Wide Mapping of UV-Induced DNA Damage with CPD-Seq. *Methods in Molecular Biology (Clifton, N.J.)* **2175,** 79–94 (2020).

721. Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand breaks mapping by next-generation sequencing. *Nature methods* **10,** 361–365 (2013).

722. Canela, A. *et al.* DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Molecular Cell* **63,** 898–911 (2016).

723. Wong, N., John, S., Nussenzweig, A. & Canela, A. END-seq: An Unbiased, High-Resolution, and Genome-Wide Approach to Map DNA Double-Strand Breaks and Resection in Human Cells. *Methods in Molecular Biology (Clifton, N.J.)* **2153,** 9–31 (2021).

724. Matos-Rodrigues, G. *et al.* S1-END-seq reveals DNA secondary structures in human cells. *Molecular Cell* **82,** 3538–3552.e5 (2022).

725. Chehrehasa, F., Meedeniya, A. C. B., Dwyer, P., Abrahamsen, G. & Mackay-Sim, A. EdU, a new thymidine analogue for labelling proliferating cells in the nervous system. *Journal of Neuroscience Methods* **177,** 122–130 (2009).

726. Macheret, M. & Halazonetis, T. D. Monitoring early S-phase origin firing and replication fork movement by sequencing nascent DNA from synchronized cells. *Nature Protocols* **14.** Publisher: Nature Publishing Group, 51–67 (2019).

727. Macheret, M. *et al.* High-resolution mapping of mitotic DNA synthesis regions and common fragile sites in the human genome through direct sequencing. *Cell Research* **30.** Publisher: Nature Publishing Group, 997–1008 (2020).

728. Wu, W. *et al.* Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature* **593.** Number: 7859 Publisher: Nature Publishing Group, 440–444 (2021).

729. Ding, J., Taylor, M. S., Jackson, A. P. & Reijns, M. A. M. Genome-wide mapping of embedded ribonucleotides and other noncanonical nucleotides using emRiboSeq and EndoSeq. *Nature Protocols* **10.** Publisher: Nature Publishing Group, 1433–1444 (2015).

730. Ali Khan, A., Raess, M. & de Angelis, M. H. Moving forward with forward genetics: A summary of the INFRAFRONTIER Forward Genetics Panel Discussion. *F1000Research* **10,** 456 (2021).

731. Adhikari, L. *et al.* A high-throughput skim-sequencing approach for genotyping, dosage estimation and identifying translocations. *Scientific Reports* **12.** Publisher: Nature Publishing Group, 17583 (2022).

732. Messner, C. B. *et al.* The proteomic landscape of genome-wide genetic perturbations. *Cell* **186,** 2018–2034.e21 (2023).

733. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)* **339,** 819–823 (2013).

734. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)* **337,** 816–821 (2012).

735. Gaj, T., Gersbach, C. A. & Barbas, C. F. ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology* **31,** 397–405 (2013).

736. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8,** 2281–2308 (2013).

737. Trevino, A. E. & Zhang, F. Genome editing using Cas9 nickases. *Methods in Enzymology* **546,** 161–174 (2014).

738. Brezgin, S., Kostyusheva, A., Kostyushev, D. & Chulanov, V. Dead Cas Systems: Types, Principles, and Applications. *International Journal of Molecular Sciences* **20,** 6041 (2019).

739. Porto, E. M., Komor, A. C., Slaymaker, I. M. & Yeo, G. W. Base editing: advances and therapeutic opportunities. *Nature Reviews Drug Discovery* **19.** Publisher: Nature Publishing Group, 839–859 (2020).

740. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods* **10,** 973–976 (2013).

741. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nature Methods* **10,** 977–979 (2013).

742. Farzadfard, F., Perli, S. D. & Lu, T. K. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS synthetic biology* **2,** 604–613 (2013).

743. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–451 (2013).

744. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155,** 1479–1491 (2013).

745. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576.** Publisher: Nature Publishing Group, 149–157 (2019).

746. Bartz, S. R. *et al.* Small interfering RNA screens reveal enhanced cisplatin cytotoxicity in tumor cells having both BRCA network and TP53 disruptions. *Molecular and Cellular Biology* **26,** 9377–9386 (2006).

747. Smogorzewska, A. *et al.* A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Molecular Cell* **39,** 36–47 (2010).

748. Mohr, S., Bakal, C. & Perrimon, N. Genomic screening with RNAi: results and challenges. *Annual Review of Biochemistry* **79,** 37–64 (2010).

749. Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163,** 1515–1526 (2015).

750. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170,** 564–576.e16 (2017).

751. Hussmann, J. A. *et al.* Mapping the genetic landscape of DNA double-strand break repair. *Cell* **184,** 5653–5669.e25 (2021).

752. Zimmermann, M. *et al.* CRISPR screens identify genomic ribonucleotides as a source of PARP-trapping lesions. *Nature* **559,** 285–289 (2018).

753. Enache, O. M. *et al.* Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nature Genetics* **52.** Publisher: Nature Publishing Group, 662–668 (2020).

754. Mair, B. *et al.* Essential Gene Profiles for Human Pluripotent Stem Cells Identify Uncharacterized Genes and Substrate Dependencies. *Cell Reports* **27,** 599–615.e12 (2019).

755. Hart, T. *et al.* Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3 (Bethesda, Md.)* **7,** 2719–2727 (2017).

756. Enoiu, M., Jiricny, J. & Schärer, O. D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Research* **40,** 8953–8964 (2012).

757. Wang, M., Dingler, F. A. & Patel, K. J. Genotoxic aldehydes in the hematopoietic system. *Blood* **139,** 2119–2129 (2022).

758. Nayak, M. *et al.* Epigenetic signature in neural plasticity: the journey so far and journey ahead. *Heliyon* **8,** e12292 (2022).

759. Wang, C. *et al.* Scalable Production of iPSC-Derived Human Neurons to Identify Tau-Lowering Compounds by High-Content Screening. *Stem Cell Reports* **9,** 1221–1233 (2017).

760. Fernandopulle, M. S. *et al.* Transcription Factor–Mediated Differentiation of Human iPSCs into Neurons. *Current Protocols in Cell Biology* **79,** e51 (2018).

761. Rothbauer, U. *et al.* A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Molecular & cellular proteomics: MCP* **7,** 282–289 (2008).

762. Wang, Z. *et al.* Structural insights into the binding of nanobodies LaM2 and LaM4 to the red fluorescent protein mCherry. *Protein Science* **30.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pro.4194, 2298–2309 (2021).

763. Sun, D. *et al.* A functional genetic toolbox for human tissue-derived organoids. *eLife* **10,** e67886 (2021).

764. Williams, R. M. *et al.* Genome and epigenome engineering CRISPR toolkit for in vivo modulation of cis-regulatory interactions and gene expression in the chicken embryo. *Development (Cambridge, England)* **145,** dev160333 (2018).

765. Weinrich, S. L. *et al.* Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT. *Nature Genetics* **17,** 498–502 (1997).

766. Hindul, N. L. *et al.* Construction of a human hTERT RPE-1 cell line with inducible Cre for editing of endogenous genes. *Biology Open* **11,** bio059056 (2022).

767. Zinder, J. C. *et al.* Shelterin is a dimeric complex with extensive structural heterogeneity. *Proceedings of the National Academy of Sciences* **119.** Publisher: Proceedings of the National Academy of Sciences, e2201662119 (2022).

768. Olivieri, M. & Durocher, D. Genome-scale chemogenomic CRISPR screens in human cells using the TKOv3 library. *STAR protocols* **2,** 100321 (2021).

769. Gallagher, J. R., Kim, A. J., Gulati, N. M. & Harris, A. K. Negative-Stain Transmission Electron Microscopy of Molecular Complexes for Image Analysis by 2D Class Averaging. *Current Protocols in Microbiology* **54.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpmc.90, e90 (2019).

770. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7** (eds Egelman, E. H. & Kuriyan, J.) Publisher: eLife Sciences Publications, Ltd, e42166 (2018).

771. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pro.3943, 70–82 (2021).

772. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10,** R25 (2009).

773. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

774. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26,** 841–842 (2010).

775. Schmid, E. W. & Walter, J. C. Predictomes: A classifier-curated database of AlphaFold-modeled protein-protein interactions. *bioRxiv: The Preprint Server for Biology,* 2024.04.09.588596 (2024).

776. Weeda, G. *et al.* Disruption of mouse *ERCC1* results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. *Current Biology* **7,** 427–439 (1997).

777. Dollé, M. E. *et al.* Broad segmental progeroid changes in short-lived Ercc1 /7 mice. *Pathobiology of Aging & Age-related Diseases* **1.** Publisher: Taylor & Francis _eprint: https://doi.org/10.3402/pba.v1i0.7219, 7219 (2011).

778. Oka, Y. *et al.* Digenic mutations in ALDH2 and ADH5 impair formaldehyde clearance and cause a multisystem disorder, AMeD syndrome. *Science Advances* **6,** eabd7197 (2020).

779. Matsumoto, M. *et al.* Characteristic phenotypes of ADH5/ALDH2 deficiency during childhood. *European Journal of Medical Genetics* **69,** 104939 (2024).

780. Bakker, S. T., de Winter, J. P. & te Riele, H. Learning from a paradox: recent insights into Fanconi anaemia through studying mouse models. *Disease Models & Mechanisms* **6,** 40–47 (2013).

781. Whitney, M. A. *et al.* Germ Cell Defects and Hematopoietic Hypersensitivity to y-Interferon in Mice With a Targeted Disruption of the Fanconi Anemia C Gene. *Blood* **88,** 49–58 (1996).

782. Chen, M. *et al.* Inactivation of Fac in mice produces inducible chromosomal instability and reduced fertility reminiscent of Fanconi anaemia. *Nature Genetics* **12.** Publisher: Nature Publishing Group, 448–451 (1996).

783. Carreau, M. *et al.* Bone Marrow Failure in the Fanconi Anemia Group C Mouse Model After DNA Damage. *Blood* **91,** 2737–2744 (1998).

784. Kelly, P. F. *et al.* Stem cell collection and gene transfer in Fanconi anemia. *Molecular Therapy: The Journal of the American Society of Gene Therapy* **15,** 211–219 (2007).

785. Langevin, F., Crossan, G. P., Rosado, I. V., Arends, M. J. & Patel, K. J. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature* **475.** Publisher: Nature Publishing Group, 53–58 (2011).

786. Howlett, N. G., Taniguchi, T., Durkin, S. G., D'Andrea, A. D. & Glover, T. W. The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. *Human Molecular Genetics* **14,** 693–701 (2005).

787. Okamoto, Y. *et al.* FANCD2 protects genome stability by recruiting RNA processing enzymes to resolve R-loops during mild replication stress. *The FEBS journal* **286,** 139–150 (2019).

788. Raya, A. *et al.* Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature* **460,** 53–59 (2009).

789. Perillo, B., Tramontano, A., Pezone, A. & Migliaccio, A. LSD1: more than demethylation of histone lysine residues. *Experimental & Molecular Medicine* **52.** Publisher: Nature Publishing Group, 1936–1947 (2020).

790. Schwab, R. A. *et al.* The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Molecular Cell* **60,** 351–361 (2015).

791. Bell, S. *et al.* Differentiation of Human Induced Pluripotent Stem Cells (iPSCs) into an Effective Model of Forebrain Neural Progenitor Cells and Mature Neurons. *Bio-protocol* **9,** e3188 (2019).

792. Yuan, F. *et al.* Efficient generation of region-specific forebrain neurons from human pluripotent stem cells under highly defined condition. *Scientific Reports* **5,** 18550 (2015).

793. Kang, S. *et al.* Characteristic analyses of a neural differentiation model from iPSC-derived neuron according to morphology, physiology, and global gene expression pattern. *Scientific Reports* **7.** Publisher: Nature Publishing Group, 12233 (2017).

794. Gregson, S. J. *et al.* Design, synthesis, and evaluation of a novel pyrrolobenzodiazepine DNA-interactive agent with highly efficient cross-linking ability and potent cytotoxicity. *Journal of Medicinal Chemistry* **44,** 737–748 (2001).

795. Cohen, E., Ophir, I. & Shaul, Y. B. Induced differentiation in HT29, a human colon adenocarcinoma cell line. *Journal of Cell Science* **112 (Pt 16),** 2657–2666 (1999).

796. Fedi, A. *et al. In vitro* models replicating the human intestinal epithelium for absorption and metabolism studies: A systematic review. *Journal of Controlled Release* **335,** 247–268 (2021).

797. Park, J.-H., Lee, J.-M., Lee, E.-J., Kim, D.-J. & Hwang, W.-B. Kynurenine promotes the goblet cell differentiation of HT-29 colon carcinoma cells by modulating Wnt, Notch and AhR signals. *Oncology Reports* **39.** Publisher: Spandidos Publications, 1930–1938 (2018).

798. Kim, Y. S. & Ho, S. B. Intestinal goblet cells and mucins in health and disease: recent insights and progress. *Current Gastroenterology Reports* **12,** 319–330 (2010).

799. Walport, L. J., Hopkinson, R. J. & Schofield, C. J. Mechanisms of human histone and nucleic acid demethylases. *Current Opinion in Chemical Biology. Mechanisms• Aesthetics• Molecular imaging* **16,** 525–534 (2012).

800. Culhane, J. C. & Cole, P. A. LSD1 and The Chemistry of Histone Demethylation. *Current opinion in chemical biology* **11,** 561–568 (2007).

801. Cuthbert, G. L. *et al.* Histone deimination antagonizes arginine methylation. *Cell* **118,** 545–553 (2004).

802. Shi, Y. *et al.* Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **119,** 941–953 (2004).

803. Metzger, E. *et al.* LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437.** Publisher: Nature Publishing Group, 436–439 (2005).

804. Huang, J. *et al.* p53 is regulated by the lysine demethylase LSD1. *Nature* **449.** Publisher: Nature Publishing Group, 105–108 (2007).

805. Bochtler, M., Kolano, A. & Xu, G.-L. DNA demethylation pathways: Additional players and regulators. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* **39,** 1–13 (2017).

806. De Miranda, A. S., Milagre, C. D. F. & Hollmann, F. Alcohol Dehydrogenases as Catalysts in Organic Synthesis. *Frontiers in Catalysis* **2.** Publisher: Frontiers (2022).

807. Edenberg, H. J. The Genetics of Alcohol Metabolism: Role of Alcohol Dehydrogenase and Aldehyde Dehydrogenase Variants. *Alcohol Research & Health* **30,** 5–13 (2007).

808. Oota, H. *et al.* Conservative evolution in duplicated genes of the primate Class I ADH cluster. *Gene* **392,** 64–76 (2007).

809. Plapp, B. V., Green, D. W., Sun, H.-W., Park, D.-H. & Kim, K. in *Enzymology and Molecular Biology of Carbonyl Metabolism 4* (eds Weiner, H., Crabb, D. W. & Flynn, T. G.) 391–400 (Springer US, Boston, MA, 1993).

810. Kumar, S., Sandell, L. L., Trainor, P. A., Koentgen, F. & Duester, G. Alcohol and Aldehyde Dehydrogenases: Retinoid Metabolic Effects in Mouse Knockout Models. *Biochimica et biophysica acta* **1821,** 198–205 (2012).

811. Pohanka, M. Toxicology and the biological role of methanol and ethanol: Current view. *Biomedical Papers of the Medical Faculty of the University Palacky, Olomouc, Czechoslovakia* **160,** 54–63 (2016).

812. Tulpule, K. & Dringen, R. Formaldehyde in brain: an overlooked player in neurodegeneration? *Journal of Neurochemistry* **127,** 7–21 (2013).

813. Kim, K. & Plapp, B. V. Inversion of substrate stereoselectivity of horse liver alcohol dehydrogenase by substitutions of Ser-48 and Phe-93. *Chemico-Biological Interactions* **276,** 77–87 (2017).

814. Plapp, B. V. *et al.* Horse Liver Alcohol Dehydrogenase: Zinc Coordination and Catalysis. *Biochemistry* **56.** Publisher: American Chemical Society, 3632–3646 (2017).

815. Pham, D. H., Moretti, P. A., Goodall, G. J. & Pitson, S. M. Attenuation of Leakiness in Doxycycline-Inducible Expression Via Incorporation of 3 AU-rich mRNA Destabilizing Elements. *BioTechniques* **45.** Publisher: Taylor & Francis _eprint: https://doi.org/10.2144/000112896, 155–162 (2008).

816. Costello, A. *et al.* Leaky Expression of the TET-On System Hinders Control of Endogenous miRNA Abundance. *Biotechnology Journal* **14.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/biot.201800219, 1800219 (2019).

817. Nakanishi, K., Cavallo, F., Brunet, E. & Jasin, M. Homologous Recombination Assay for Interstrand Cross-Link Repair. *Methods in Molecular Biology (Clifton, N.j.)* **745,** 283–291 (2011).

818. Manandhar, M., Boulware, K. S. & Wood, R. D. The ERCC1 and ERCC4 (XPF) genes and gene products. *Gene* **569,** 153–161 (2015).

819. Mori, T. *et al.* ERCC4 Variants Identified in a Cohort of Patients with Segmental Progeroid Syndromes. *Human mutation* **39,** 255–265 (2018).

820. Niedernhofer, L. J. *et al.* A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature* **444,** 1038–1043 (2006).

821. Chatzinikolaou, G. *et al.* XPF interacts with TOP2B for R-loop processing and DNA looping on actively transcribed genes. *Science Advances* **9,** eadi2095 (2023).

822. Collie, G. W. *et al.* Fragment-Based Discovery of Novel MUS81 Inhibitors. *ACS medicinal chemistry letters* **15,** 1151–1158 (2024).

823. Arora, S. *et al.* Identification of small molecule inhibitors of ERCC1-XPF that inhibit DNA repair and potentiate cisplatin efficacy in cancer cells. *Oncotarget* **7,** 75104–75117 (2016).

824. McNeil, E. M. *et al.* Inhibition of the ERCC1-XPF structure-specific endonuclease to overcome cancer chemoresistance. *DNA repair* **31,** 19–28 (2015).

825. Gentile, F. *et al.* Computer-aided drug design of small molecule inhibitors of the ERCC1-XPF protein-protein interaction. *Chemical Biology & Drug Design* **95,** 460–471 (2020).

826. Sabatella, M. *et al.* Repair protein persistence at DNA lesions characterizes XPF defect with Cockayne syndrome features. *Nucleic Acids Research* **46,** 9563–9577 (2018).

827. Muller, P. a. J. *et al.* Mutant p53 enhances MET trafficking and signalling to drive cell scattering and invasion. *Oncogene* **32.** Publisher: Nature Publishing Group, 1252–1265 (2013).

828. Larsen, A. K., Galmarini, C. M. & D'Incalci, M. Unique features of trabectedin mechanism of action. *Cancer Chemotherapy and Pharmacology* **77,** 663–671 (2016).

829. Wang, J. *et al.* Trabectedin in Cancers: Mechanisms and Clinical Applications. *Current Pharmaceutical Design* **28,** 1949–1965 (2022).

830. Takebayashi, Y. *et al.* Antiproliferative activity of ecteinascidin 743 is dependent upon transcription-coupled nucleotide-excision repair. *Nature Medicine* **7,** 961–966 (2001).

831. Zewail-Foote, M. & Hurley, L. H. Differential rates of reversibility of ecteinascidin 743-DNA covalent adducts from different sequences lead to migration to favored bonding sites. *Journal of the American Chemical Society* **123,** 6485–6495 (2001).

832. Son, K. *et al.* Trabectedin derails transcription-coupled nucleotide excision repair to induce DNA breaks in highly transcribed genes. *Nature Communications* **15.** Publisher: Nature Publishing Group, 1388 (2024).

833. Chan, Y. W., Fugger, K. & West, S. C. Unresolved recombination intermediates lead to ultra-fine anaphase bridges, chromosome breaks and aberrations. *Nature Cell Biology* **20.** Publisher: Nature Publishing Group, 92–103 (2018).

834. Binz, R. L. *et al.* Identification of novel breakpoints for locus- and region-specific translocations in 293 cells by molecular cytogenetics before and after irradiation. *Scientific Reports* **9.** Publisher: Nature Publishing Group, 10554 (2019).

835. Giannuzzi, G. *et al.* The Human-Specific BOLA2 Duplication Modifies Iron Homeostasis and Anemia Predisposition in Chromosome 16p11.2 Autism Individuals. *The American Journal of Human Genetics* **105.** Publisher: Elsevier, 947–958 (2019).

836. Nabet, B. *et al.* The dTAG system for immediate and target-specific protein degradation. *Nature chemical biology* **14,** 431–441 (2018).

837. Mulderrig, L. & Garaycoechea, J. I. XPF-ERCC1 protects liver, kidney and blood homeostasis outside the canonical excision repair pathways. *PLOS Genetics* **16** (ed Maizels, N.) Publisher: Public Library of Science (PLoS), e1008555 (2020).

838. Caine, E. A. *et al.* Targeted Protein Degradation Phenotypic Studies Using HaloTag CRISPR/Cas9 Endogenous Tagging Coupled with HaloPROTAC3. *Current Protocols in Pharmacology* **91.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpph.81, e81 (2020).

839. Pelosse, M. *et al.* MultiBac: from protein complex structures to synthetic viral nanosystems. *BMC Biology* **15,** 99 (2017).

840. Nie, Y., Bellon-Echeverria, I., Trowitzsch, S., Bieniossek, C. & Berger, I. Multiprotein complex production in insect cells by using polyproteins. *Methods in Molecular Biology (Clifton, N.J.)* **1091,** 131–141 (2014).

841. Shang, H., Garretson, T. A., Kumar, C. M. S., Dieter, R. F. & Cheng, X.-W. Improved pFastBac™ donor plasmid vectors for higher protein production using the Bac-to-Bac® baculovirus expression vector system. *Journal of Biotechnology* **255,** 37–46 (2017).

842. Verma, R., Boleti, E. & George, A. J. T. Antibody engineering: Comparison of bacterial, yeast, insect and mammalian expression systems. *Journal of Immunological Methods* **216,** 165–181 (1998).

843. Kallas, A., Pook, M., Maimets, M., Zimmermann, K. & Maimets, T. Nocodazole Treatment Decreases Expression of Pluripotency Markers Nanog and Oct4 in Human Embryonic Stem Cells. *PLoS ONE* **6,** e19114 (2011).

844. Barger, C. J., Branick, C., Chee, L. & Karpf, A. R. Pan-Cancer Analyses Reveal Genomic Features of FOXM1 Overexpression in Cancer. *Cancers* **11,** 251 (2019).

845. Li, A. *et al. Structures of the human adult muscle-type nicotinic receptor in resting and desensitised states* Pages: 2024.12.13.624530 Section: New Results. 2024.

846. Zhang, Y. H. *et al.* Effect of Cell Cycle Phase on Sf9 Cell Activity and Autographa Californica Multiple Nucleopolyhedrovirus Infection. *Advanced Materials Research* **610-613.** Publisher: Trans Tech Publications Ltd, 115–119 (2013).

847. Kim, S. *et al.* Polyubiquitinated PCNA triggers SLX4-mediated break-induced replication in alternative lengthening of telomeres (ALT) cancer cells. *Nucleic Acids Research* **52,** 11785–11805 (2024).

848. Bruning, J. B. & Shamoo, Y. Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-delta p66 subunit and flap endonuclease-1. *Structure (London, England: 1993)* **12,** 2209–2219 (2004).

849. Swift, L. P. *et al.* SNM1A is crucial for efficient repair of complex DNA breaks in human cells. *Nature Communications* **15.** Publisher: Nature Publishing Group, 5392 (2024).

850. Álvarez-Salmoral, D. *et al. AlphaBridge: tools for the analysis of predicted macromolecular complexes* Pages: 2024.10.23.619601 Section: New Results. 2024.

851. Bridge, W. L., Vandenberg, C. J., Franklin, R. J. & Hiom, K. The BRIP1 helicase functions independently of BRCA1 in the Fanconi anemia pathway for DNA crosslink repair. *Nature Genetics* **37,** 953–957 (2005).

852. Levran, O. *et al.* The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. *Nature Genetics* **37,** 931–933 (2005).

853. Peng, M. *et al.* The FANCJ/MutLalpha interaction is required for correction of the cross-link response in FA-J cells. *The EMBO journal* **26,** 3238–3249 (2007).

854. Sommers, J. A. *et al.* Novel function of the Fanconi anemia group J or RECQ1 helicase to disrupt protein-DNA complexes in a replication protein A-stimulated manner. *The Journal of Biological Chemistry* **289,** 19928–19941 (2014).

855. Suhasini, A. N. *et al.* Interaction between the helicases genetically linked to Fanconi anemia group J and Bloom's syndrome. *The EMBO journal* **30,** 692–705 (2011).

856. Gong, Z., Kim, J.-E., Leung, C. C. Y., Glover, J. N. M. & Chen, J. BACH1/FANCJ acts with TopBP1 and participates early in DNA replication checkpoint control. *Molecular Cell* **37,** 438–446 (2010).

857. Nath, S. & Nagaraju, G. FANCJ helicase promotes DNA end resection by facilitating CtIP recruitment to DNA double-strand breaks. *PLoS genetics* **16,** e1008701 (2020).

858. Yaneva, D. *et al.* The FANCJ helicase unfolds DNA-protein crosslinks to promote their repair. *Molecular Cell* **83,** 43–56.e10 (2023).

859. Aprosoff, C. M. *et al.* Comprehensive Interactome Mapping of the DNA Repair Scaffold SLX4 Using Proximity Labeling and Affinity Purification. *Journal of Proteome Research* **22,** 1660–1681 (2023).

860. Xie, J. *et al.* Targeting the FANCJ-BRCA1 interaction promotes a switch from recombination to poleta-dependent bypass. *Oncogene* **29,** 2499–2508 (2010).

861. Hopton, S. R. & Thompson, A. S. Nuclear magnetic resonance solution structures of inter- and intrastrand adducts of DNA cross-linker SJG-136. *Biochemistry* **50,** 4720–4732 (2011).

862. Cole, R. S. Repair of DNA containing interstrand crosslinks in Escherichia coli: sequential excision and recombination. *Proceedings of the National Academy of Sciences of the United States of America* **70,** 1064–1068 (1973).

863. Umansky, C. *et al.* Endogenous formaldehyde scavenges cellular glutathione resulting in redox disruption and cytotoxicity. *Nature Communications* **13.** Publisher: Nature Publishing Group, 745 (2022).

864. Savage, S. A. & Walsh, M. F. Myelodysplastic Syndrome, Acute Myeloid Leukemia, and Cancer Surveillance in Fanconi Anemia. *Hematology/Oncology Clinics of North America* **32,** 657–668 (2018).

865. Zhang, H., Xiong, Y. & Chen, J. DNA–protein cross-link repair: what do we know now? *Cell & Bioscience* **10,** 3 (2020).

866. Frey, A. G., Palenchar, D. J., Wildemann, J. D. & Philpott, C. C. A Glutaredoxin·BolA Complex Serves as an Iron-Sulfur Cluster Chaperone for the Cytosolic Cluster Assembly Machinery. *The Journal of Biological Chemistry* **291,** 22344–22356 (2016).

867. Tsai, H.-J. *et al.* Hypo-osmotic-like stress underlies general cellular defects of aneuploidy. *Nature* **570,** 117–121 (2019).

868. Weissmann, F. *et al.* biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* **113,** E2564–2569 (2016).

869. Qiao, R. *et al.* Mechanism of APC/CCDC20 activation by mitotic phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* **113,** E2570–2578 (2016).

870. Brosh, R. M. & Cantor, S. B. Molecular and cellular functions of the FANCJ DNA helicase defective in cancer and in Fanconi anemia. *Frontiers in Genetics* **5,** 372 (2014).

871. Stark, J. M., Pierce, A. J., Oh, J., Pastink, A. & Jasin, M. Genetic Steps of Mammalian Homologous Repair with Distinct Mutagenic Consequences. *Molecular and Cellular Biology* **24.** Publisher: Taylor & Francis _eprint: https://doi.org/10.1128/MCB.24.21.9305-9316.2004, 9305–9316 (2004).

872. Certo, M. T. *et al.* Tracking genome engineering outcome at individual DNA breakpoints. *Nature Methods* **8,** 671–676 (2011).

873. Olive, P. L. & Banáth, J. P. The comet assay: a method to measure DNA damage in individual cells. *Nature Protocols* **1.** Publisher: Nature Publishing Group, 23–29 (2006).

874. Wu, J. H. & Jones, N. J. Assessment of DNA interstrand crosslinks using the modified alkaline comet assay. *Methods in Molecular Biology (Clifton, N.J.)* **817,** 165–181 (2012).

875. Pavani, R. *et al.* Structure and repair of replication-coupled DNA breaks. *Science* **385.** Publisher: American Association for the Advancement of Science, eado3867 (2024).

876. Xu, Y. *et al.* DNA nicks in both leading and lagging strand templates can trigger break-induced replication. *Molecular Cell* **85,** 91–106.e5 (2025).

877. Rozbeh, R. & Forchhammer, K. Split NanoLuc technology allows quantitation of interactions between PII protein and its receptors with unprecedented sensitivity and reveals transient interactions. *Scientific Reports* **11.** Publisher: Nature Publishing Group, 12535 (2021).