SAN1 – a novel nuclease involved in DNA Interstrand Crosslink Repair

By

Heather Joan McCartney

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Cell and Developmental Biology

September 30th, 2018

Nashville, TN

Approved:

William Tansey, Ph.D.
Ian Macara, Ph.D.
Neil Osheroff, Ph.D.
Katherine Friedman, Ph.D.
Jason MacGurn, Ph.D.

ACKNOWLEDGEMENTS

This work would not have been possible without the help and support of many other people over the years. First I would like to thank my mentor Dr. Ian Macara for his scientific support and guidance. Ian is a brilliant scientist who provided endless hours of advice, discussion, and mentoring to me as a young scientist in the lab.

Thank you also to my all of the Macara Lab members, past and present. Although my thesis work did not always align with their areas of expertise they were all always happy and willing to provide any direction they could. I also want to thank my thesis committee for their support and constructive criticism of my project as it developed over the years. Committee meetings were invaluable to me throughout my time in the lab and provided a space for me to get feedback on my work and progression as a scientist.

I also want to thank my friends at Vanderbilt for being the most amazing support system and for making this experience enjoyable even when aspects of my thesis project were challenging. In particular, Kristy Walsh, Richard Guyer, Amanda Duran, Leslie Roteta Sedgeman, and Tracy Fetterly. Being surrounded by such smart and talented people who are also kind-hearted and hilarious is rare and I appreciate them greatly.

Lastly, and most importantly, I want to thank my parents, my brother Andrew, my boyfriend Bradley Reuter, and my friends Melissa Wiel and Ally Mallya. I am thankful to them for facing this chapter in my life with unlimited patience, support, love, and encouragement. They have all been unwavering in their support of me during this time and for that I am indescribably grateful. They are the best people I know.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ii

LIST OF FIGURES v

LIST OF ABBREVIATIONS vii

1 Introduction 1

1.1 DNA Damage Response 1

1.2 Mismatch Repair 5

1.3 Base Excision Repair 6

1.4 Nucleotide Excision Repair 7

1.5 Double Strand Break Repair 9

1.6 Interstrand Crosslink Repair 12

1.7 Fanconi anemia 19

1.8 FEN1 Structure-specific Nucleases 21

1.9 Research Objectives 26

2 SAN1 Nuclease Activity and Substrate Preferences 28

2.1 Abstract 28

2.2 Introduction 28

2.3 Results and Discussion 29

3 SAN1 Involvement in Interstrand Crosslink Repair 57

3.1 Abstract 57

3.2 Introduction 57

3.3 Results and Discussion 59

4 Discussion and Future Directions 77

4.1 Discussion 77

4.2 Future Directions 85

APPENDIX I: Additional Data 89

APPENDIX II: Oligonucleotides Used in Nuclease Assays 93

APPENDIX III: Protocols 94

REFERENCES 100

LIST OF FIGURES

Figure Page

1. DNA damage response signaling pathway components. 2

2. DNA damage response overview and interstrand crosslink repair 4

3. Fanconi anemia-dependent interstrand crosslink repair. 15

4. Pathways of interstrand crosslink repair 18

5. Fanconi anemia phenotypes. 20

6. FEN1 family member substrate preferences and nuclease domain structure 23

7. SAN1 homology to the FEN1 family of structure specific nucleases 31

8. SAN1 acts as an exonuclease on 5' ssDNA 34

9. Recombinant mouse WT SAN1 shows similar nuclease activity to human WT SAN1 36

10. Two-step affinity purified human WT SAN1 shows nuclease activity in spin assay 38

11. SAN1 purification from mammalian cells and cofactor characterization. 41

12. SAN1 cleaves 5' ssDNA substrates including splayed arm and flap structures 45

13. SAN1 acts as an exonuclease on 5' ssDNA structures 50

14. SAN1 C-terminus is required for nuclease activity 53

15. Schematic of a model for SAN1 nuclease activity on DNA structures 56

16. Loss of SAN1 leads to sensitization of cells to ICL agents 60

17. SAN1$^{-/-}$ HeLa cells are not sensitized to hydroxyurea or camptothecin 63

18. SAN$^{-/-}$ cells display increased levels of radial chromosomes in response to MMC. 65

19. SAN1 functions independently of the FA pathway. 68

20. SAN1 interacts with Senataxin and interaction is required for ICL resistance 70

21. SAN1$^{-/-}$ cells show increased R-loop formation with MMC treatment 73

22. Speculative model for SAN1 function with SETX in processing of ICLs 88

23. Raw blot data from Figures 16, 17, 21 89

24. Genome-wide yeast two-hybrid results from Hybrigenics 92

LIST OF ABBREVIATIONS

ALT-EJ/SSA……Alternative end-joining/Single Strand Annealing
APE1…….....…….AP endonuclease 1
AP site……...Apurinic site
ATM……...Ataxia telangiectasia mutated
BER…….Base Excision Repair
BRCA1……...Breast Cancer type 1 susceptibility protein
BRCA2……...Breast Cancer type 2 susceptibility protein
CSA……...……Colony Survival Assay
CtIP…….CtBP-interacting protein
C-NHEJ……..Classical Non-homologous End-Joining
DNA……Deoxyribonucleic Acid
DNA-PKcs……...DNA-dependent protein kinase, catalytic subunit
DSB……Double stranded break
dsDNA……double-stranded DNA
EME1…….Essential Meiotic Structure-Specific Endonuclease 1
ERCC4/ERCC1……DNA repair endonuclease XPF
EXO1……..Exonuclease 1
FA……Fanconi anemia
FAN1…….Fanconi-associated nuclease 1
FANCL…….E3 ubiquitin-protein ligase FANCL
FANCI…….Fanconi anemia complementation group I
FANCD2……..Fanconi anemia complementation group D2
FEN1…….Flap Endonuclease 1
GEN1…….Holliday Junction resolvase
GG-NER…….Global Genome-Nucleotide Excision Repair
γH2AX……...H2A Histone Family Member X
HJ……...Holliday Junction
HR……Homologous Recombination
ICL……Interstrand Crosslink Repair
IDL……Insertion Deletion
LIG 1……Ligase 1
MMR……..Mismatch Repair
MRE11……..…MRE11 Homolog, Double Strand Break Repair Nuclease
MRN……...Mre11/Rad50/Nbs1 Complex
MUS81……...MUS81 structure-specific endonuclease
NBS1…….Nijmegen Breakage Syndrome 1 Protein
NER……...Nucleotide Excision Repair
NHEJ……..Non-homologous End Joining
MMC……Mitomycin C
MSI…….Microsatellite Instability
PCNA……Proliferating Core Nuclear Antigen
RAD50……..Rad50 Double Strand Break Repair Protein
REV1……...DNA repair protein REV1
RNA POL II……...RNA Polymerase II

ROS……...Reactive Oxygen Species
RPA…….Replication Protein A
SLX1……Structure-specific endonuclease subunit SLX1
SLX4……Structure-specific endonuclease subunit SLX4
ssDNA……..single-stranded DNA
TC-NER……...Transcription-Coupled Nucleotide Excision Repair
UV……Ultraviolet
XPC……Xeroderma pigmentosum complementation group C
XPE…….Xeroderma pigmentosum complementation group E
XPG……Xeroderma Pigmentosum complementation group G

CHAPTER 1: Introduction

In order for cells to survive, they must coordinate complicated cellular signaling responses to many different types of stimuli including DNA damage. Preserving genomic integrity is key to maintaining normal cellular functions along with preventing the accumulation of mutations that can lead to cell death, disease, or aberrant cell growth. Although the pathways involved in repairing various DNA lesions have been studied for years, there remain fundamental gaps in understanding the overlap of these pathways and the machinery involved in certain cases. This dissertation characterizes a previously unknown component of a Fanconi anemia-independent interstrand crosslink (ICL) DNA repair pathway and explores its importance in protecting cells against ICL damage.

1.1 DNA Damage Response

Every day, cells encounter many different types of DNA damage that if left unrepaired can result in cell death or diseases such as cancer. Detection and repair of DNA lesions is tightly regulated and DNA damage checkpoint pathways are interconnected[1]. DNA damage checkpoints are signaling cascades that delay cell cycle progression to allow for repair of DNA lesions[2]. Eukaryotic cells have four main checkpoints throughout the cell cycle including the G1/S checkpoint, the intra-S phase checkpoint, the G2/M checkpoint and the S/M checkpoint[3]. In addition to arresting cell cycle progression, DNA damage checkpoint factors initiate DNA damage response repair pathways, delivery of repair proteins to sites of damage, and activation of transcriptional responses[4]. There are several key components of DNA damage checkpoints including sensors, mediators, transducers,

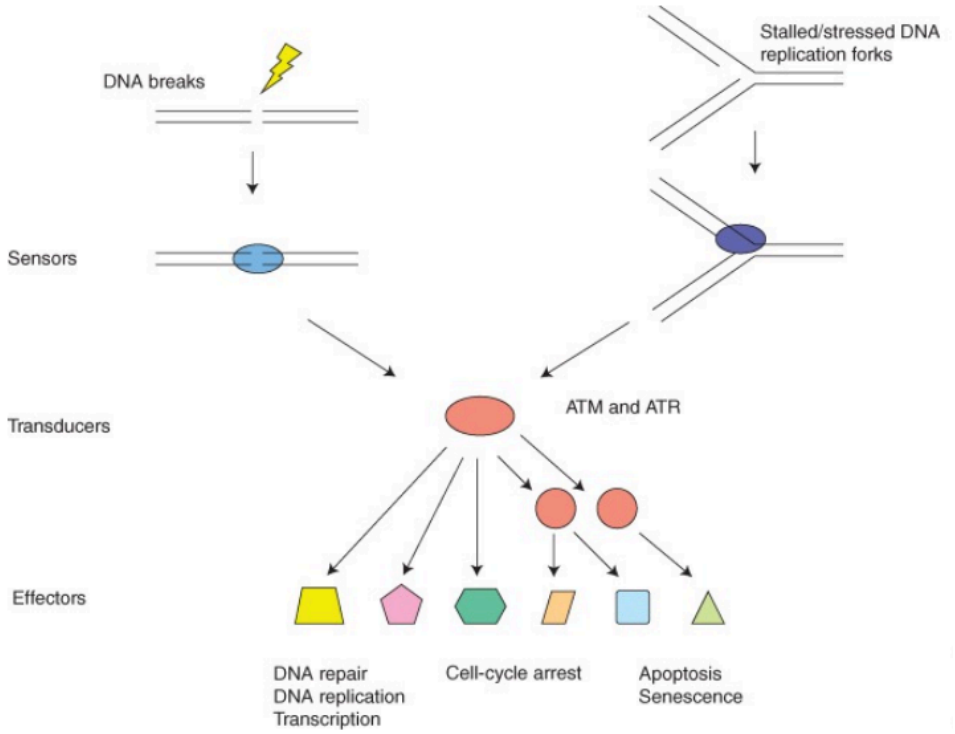


Figure from Marechal, *et al.* 2013 (3).

Figure 1. DNA Damage Response signaling pathway components. Like other signal transduction pathways, the DDR signaling pathway consists of signal sensors, transducers, and effectors. The sensors of this pathway are proteins that recognize DNA structures induced by DNA damage and DNA replication stress. The transducers of this pathway are kinases, including ATM, ATR, and their downstream kinases. The effectors of this pathway are substrates of ATM, ATR, and their downstream kinases. These effectors of ATM and ATR are involved in a broad spectrum of cellular processes that are important for maintenance of genomic stability of organisms.

.

and effectors[5] (Figure 1). Sensor proteins are required to detect DNA lesions and initiate DNA damage checkpoint activation[6]. Meditator proteins interact with transducer proteins as an additional layer of specificity to activate the proper downstream protein factors depending on the phase of the cell cycle[6]. Kinases transduce the checkpoint activation signal by phosphorylating additional protein factors, called effectors, that act to regulate other cellular processes[7]. Proteins can have more than one role in DNA damage checkpoint activation allowing for multiple levels of regulation and complexity in DNA damage checkpoint activation and signal propagation[5]. If DNA damage is too extensive, and cannot be repaired, the cell will initiate apoptosis, or programmed cell death[7].

DNA damage can result from exogenous sources including ultraviolet radiation or chemotherapeutics and can also result from endogenous sources including, but not limited to, reactive aldehydes[8–10]. Each lesion is repaired by a specific pathway that must act to detect the lesion, excise the lesion, and repair the DNA (Figure 2). In the absence of DNA repair, cells accumulate damage resulting in inhibition of basic cellular processes, including replication and transcription, or the formation of chromosomal aberrations[11] (Figure 2). The main DNA repair pathways include: mismatch repair (MMR), base-excision repair (BER), nucleotide excision repair (NER), non-homologous end joining (NHEJ), homologous recombination (HR), and interstrand crosslink (ICL) repair[4]. However, there are still gaps in our understanding of which proteins are involved in each step of repair and how the various DNA damage response pathways are interconnected. Interestingly, there is significant overlap in the protein machinery responsible for detecting, excising, and repairing different DNA lesions[4] as well as in the

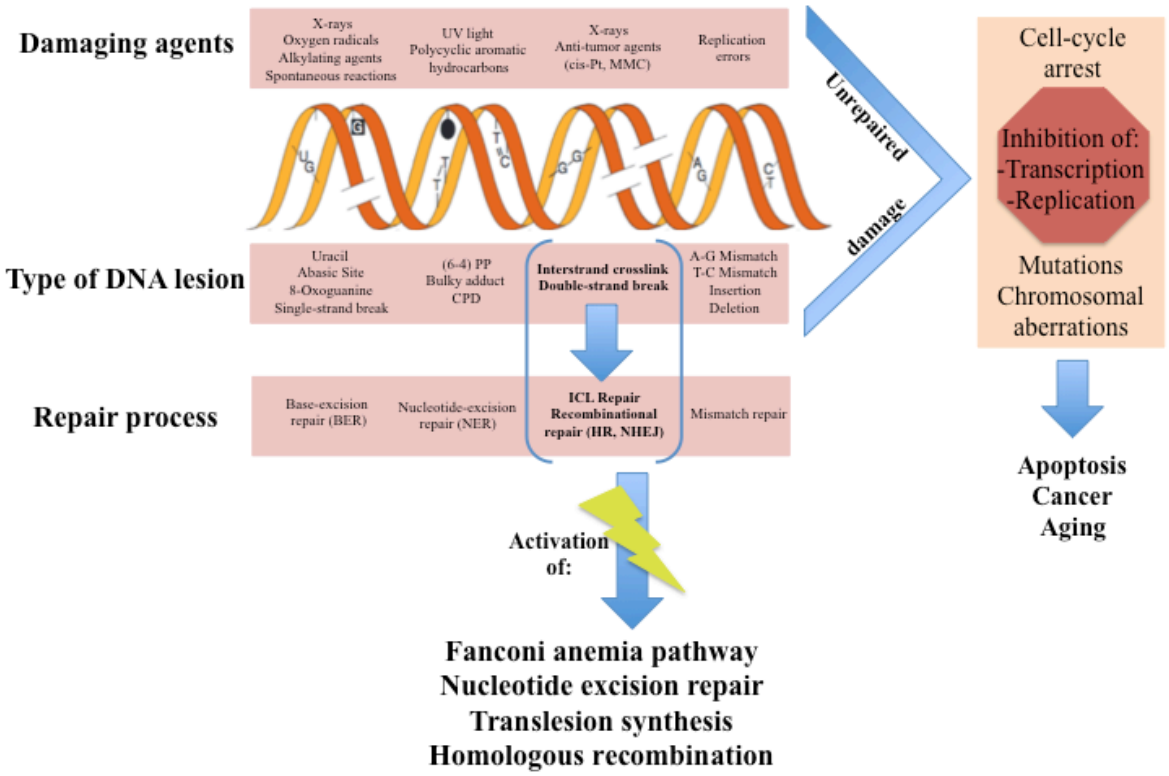


Figure 2. DNA damage response overview and Interstrand Crosslink Repair. Cells encounter many different types of endogenous and exogenous DNA damaging agents and require an extensive network of proteins to detect the lesion, remove the damage, and repair the DNA to prevent inhibition of basic cellular processes such as transcription and replication. The inability to repair DNA lesions results in persistent damage that can cause aging and various diseases such a cancer. Interstrand crosslinks require the coordination of the Fanconi anemia pathway, nucleotide excision repair, translesion synthesis, and homologous recombination and are thus complicated lesions to repair.

contribution of various protein components to each DNA repair pathway. For example, MMR and HR are linked during replication where MMR replaces mismatched bases and HR provides replication fork stability[12]. The coordination of these two distinct DNA repair pathways is required for maintaining genomic stability and preventing the accumulation of DNA damage. The importance of DDR crosstalk is also exemplified in ICL repair where the repair of crosslinks requires a unique combination of nucleotide excision repair, translesion synthesis, and homologous recombination[13]. Thus, the identification and characterization of additional protein components of the DNA damage response is crucial in developing a deeper understanding of how cells handle genomic stress.

1.2 Mismatch Repair

The mismatch repair pathway is responsible for detecting, removing, and repairing insertions and deletions (IDLs) that result from microsatellite regions in the genome along with repairing base-base mismatches[14]. Microsatellite regions are regions of short repetitive sequences that can be unstable during DNA replication resulting in the DNA replication primer dissociating and re-annealing at a different location in the microsatellite region[14]. Microsatellite instability (MSI) refers to the hypermutability of these regions including the possibility of insertions and deletions that can lead to accumulated mutations and cancer[15]. In addition to preventing MSI, the MMR pathway also replaces base-base mismatches that are not detected or repaired by the proofreading activity of the DNA polymerase, a critical component for maintaining genetic integrity

during DNA replication[14]. Combined, MMR is responsible for increasing genomic fidelity by orders of magnitude[16]. An important step in MMR is the cleavage of the heteroduplex structures resulting from IDLs or base-base mismatches and resection of the region by 3'-5' or 5'-3' exonucleases[12]. Exonuclease 1 (EXO1), a 5'-3' exonuclease from the FEN1 family of structure specific nucleases, has been implicated in this process in *in vitro* and *in vivo* studies and is, surprisingly, the only required exonuclease for MMR in a reconstituted system[17]. In addition to functioning in MMR, EXO1 has exonuclease activity on dsDNA and also flap endonuclease activity[18] and, as described below, is involved in repair of double-stranded DNA breaks (DSBs).

1.3 Base Excision Repair

In addition to being able to repair base-base mismatches, cells also need to be able to replace bases that have been damaged by endogenous damaging agents such as reactive oxygen species or by exogenous damaging agents such as ultraviolet light. These changes result in chemical instability and structural alterations to the DNA[19]. If oxidized, alkylated, or deaminated bases are left unrepaired by the BER pathway, the cell can accumulate damage resulting in DNA breaks during replication[20]. BER removes these lesions by using a toolkit of DNA glycosylases to generate apurinic (AP) sites, followed by additional processing of the AP site by AP endonuclease 1 (APE1), and replacement of the damaged nucleotide with a new nucleotide[21]. There are several factors that contribute to the selection of short-patch BER or long-patch BER, including cell cycle stage and the type of damage present on the DNA[22]. Short-patch BER replaces only the

damaged nucleotide while long-patch BER replaces a small region of nucleotides surrounding the damage site[22]. In long-patch BER, flap endonuclease 1 (FEN1) is responsible for cleaving the 5' flap that forms as a result of DNA polymerase ß/Ø/∑ polymerizing 2-15 nucleotides to replace the damaged base resulting in the damaged region being displaced on the flap[23]. FEN1 is a member of a structure-specific family of nucleases that act in DNA replication and repair. In addition to long-patch BER, FEN1 also functions in processing Okazaki fragments during DNA replication[24].

1.4 Nucleotide Excision Repair

Another DNA repair pathway that functions to remove damaged DNA is nucleotide excision repair (NER), which removes DNA damaged by ultraviolet (UV) light, various environmental damaging agents, bulky lesions, and some chemotherapeutics[25]. Importantly, this pathway is also required during the processing of interstrand crosslinks, as is discussed below.

Nucleotide excision repair has two main repair mechanisms: global genome NER (GG-NER) and transcription-coupled NER (TC-NER). GG-NER is responsible for scanning the genome for helical distortions caused by any of the previously listed DNA mutagens. TC-NER, by comparison, functions with transcription machinery to remove DNA lesions on genes that are transcribed for normal cellular function[26]. The set of enzymes that function in repair must be able to recognize many different lesions that represent NER substrates including cyclopurines generated by reactive oxygen species (ROS),

cyclobutane-pyrimidine dimers (CPDs), 6-4 pyrimidine-pyrimidone photoproducts (6-4PPs) from UV radiation, and additional bulky adducts from various sources[25]. The ability to recognize a wide range of lesions is the result of the NER system developing the ability to group together bulky lesions based on their structural commonalities[27]. Although the sub-pathways utilize the same core NER machinery to remove and repair the lesions, there are different initiation events for the two processes and the factors involved in the initiation event must be highly sensitive to recognizing NER-specific damage. GG-NER is initiated by complexes of factors XPC and XPE while TC-NER is initiated when RNA pol II is inhibited by damage encountered during transcription[28].

The steps of NER can be described as initiation or damage recognition, assembly of the multi-subunit repair complex, excision of the damaged DNA, and DNA repair synthesis followed by strand ligation[29]. In the case of GG-NER, initiation results from XPC coming in contact with damaged DNA as it scans the genome, an event that signals for assembly of the repair complex. In TC-NER, initiation results from collision of RNA polymerase II with damaged DNA causing a block in transcription[30]. After TFIIH complex assembly, including binding of factors XPA, RPA, and XPG, TFIIH opens the helix to allow the endonucleases XPF/ERCC1 and XPG to access and excise the damaged DNA[31]. DNA synthesis and ligation is then completed by replication proteins DNA polymerase σ/ε, PCNA and Lig1[32]. The ability to repair lesions that block the transcription complex in a timely manner is critical to maintaining normal gene expression and protein levels in the cell, making TC-NER a particularly important repair process. People who inherit NER protein deficiencies (XPA-XPG) display the UV

intolerance disease Xeroderma Pigementosum (XP)[33]. XP is characterized by an extreme sensitivity to sunlight with any exposure resulting in severe blistering and burns along with progressive mental decline, cancer and eventual death[34]. Although XP is not the only NER-associated disease, XP is a chronic disease with no known treatments that leaves patients with a terrible quality of life.

1.5 Double Strand Break Repair

The two main DNA repair pathways that detect and repair double-strand breaks (DSBs) are homologous recombination (HR) and non-homologous end joining (NHEJ). Although both pathways function to repair DSB lesions, they act at different points in the cell cycle, and are regulated by different effector enzymes. End-resection, and thus HR only, drives the commitment step of end-resection[35]. DSBs can form as a result of DNA damaging agents such as IR or as a result of genomic stress such as the persistent stalling of a replication fork[36]. DSBs are highly cytotoxic and when left unrepaired can result in insertions and deletions in the DNA, chromosomal aberrations, and cell death[37].

Homologous recombination is a high-fidelity repair process that functions in the S and G2 phases of the cell cycle to repair DSBs, rescue stalled replication forks, and reinitiate collapsed replication forks[38]. It is also essential to the processing of interstrand cross-links (ICLs), described in the next section. HR relies on the sister chromatid of the damaged DNA to serve as a template for repair[39]. The lesion is typically detected by the Ataxia telangiectasia mutated (ATM) sensor kinase resulting in the phosphorylation of

histone variant H2AX to generate the damage signal γH2AX[40]. This signaling event recruits the MRN complex consisting of MRE11, RAB50, and NBS1 along with BRCA1 which initiate processing of the end of the DNA of the DSB[41]. Although MRN is responsible for initial processing of the DSB, other nucleases are required to generate the 3' single-stranded DNA overhang created by resecting the 5' strand. CtIP is an important nuclease required for end-resection and in certain cases where more extensive end-resection is required other nucleases may be involved including EXO1 and Dna2[42,43]. As stated previously, 5'-3' end resection of the DSB is the determinant step of whether the HR or NHEJ pathway acts to repair the damage and thus is highly regulated[44]. There are also helicases including BLM and WRN that act to stimulate nuclease activity to expedite repair[45].

After generation of the 3' ssDNA overhang, the single-stranded DNA binding protein replication protein A (RPA) binds the ssDNA overhang while recruiting the HR protein RAD51[43]. RPA acts to protect ssDNA from additional degradation or damage by other damaging sources or additional nuclease processing[46]. Aided by a mediator protein, RAD51 then displaces RPA and initiates strand invasion and formation of a D-loop, in which the protein/DNA complex invades the homologous chromosome searching for a region of homology to the ssDNA overhang[47]. After identification of the homologous region, a DNA polymerase is able to synthesize new DNA at the location of the DSB. The resulting repair intermediate can be resolved by synthesis-dependent strand annealing or by the action of resolvase enzymes (required for repair of a Holliday Junction)[48]. GEN1 is a resolvase enzyme that is a member of the FEN1 family of

nucleases and acts to resolve Holliday Junctions[49]. Holliday Junctions are four-stranded DNA structures that require two cleavage events to process the Holliday Junctions to downstream repair steps[48].

Although HR is the preferred method of repairing DSBs, due to HR being an extremely high fidelity process, there are three other more error-prone pathways by which cells can repair DSBs including classical non-homologous end-joining (C-NHEJ), alternative end-joining (Alt-EJ), and single-strand annealing (SSA)[44]. Cell cycle is a major determinant of which pathway is chosen, as SSA and Alt-EJ both require activation of end resection nucleases MRN and CtIP to repair the DSB and these nucleases are only active during S and G2 phases[50]. C-NHEJ on the other hand, does not require any end resection and thus functions in G1 and M phases of the cell cycle[11]. C-NHEJ is the main pathway for repairing DSBs in adult mammalian cells and is initiated when a DSB is recognized by the Ku80/Ku70 DNA binding proteins[51]. This binding event recruits DNA-PKs and activates the catalytic component of the complex[52]. In certain cases other proteins are involved including Artemis and DNA ligase IV[53]. Although C-NHEJ is a relatively quick process, it is fundamentally error-prone in that the blunt ligation of the ends of the DSB can result in insertions and deletions[53]. Alt-EJ/SSA is a mutagenic repair pathway that utilizes a mix of end resection and homology-based repair, therefore competing with RAD51 and HR[50]. Even though NHEJ pathways are error-prone, they are vitally important outside of S and G2 phases of the cell cycle and when cells are HR deficient[54].

1.6 Interstrand Crosslink Repair

Interstrand crosslinks (ICLs) are a highly cytotoxic form of DNA damage that, when left unrepaired, result in the inhibition of cellular processes including transcription and replication[13]. ICLs are repaired by interstrand crosslink repair which is a complex repair pathway involving the coordination of several previously discussed DDR pathways including nucleotide excision repair, translesion synthesis, and homologous recombination[13] (Figure 3). The requirement of proteins from several different DNA repair pathways to complete ICL repair makes the coordination and crosstalk between transducer and effector proteins particularly important. Persistent ICLs can result in DSBs that are processed aberrantly by NHEJ. This results in the formation of chromosomal aberrations and radial chromosomes[55]. The cytotoxicity of ICLs is due to the fact that they covalently link opposing sides of the DNA helix and result in various levels of helix distortion depending on the chemistry of the ICL compound[56]. ICLs can result from exogenous sources such as the chemotherapeutic drugs mitomycin C (MMC) or cisplatin or from endogenous sources including reactive aldehydes and oxidized lipids that are a product of normal metabolism[11]. ICL-inducing agents represent a wide range of distortion severity in the DNA double helix and this characteristic has been shown to be one of the ways by which cells detect ICL lesions[56]. Although ICLs are commonly detected in S phase when replication forks collide with the ICL, cells are also able to repair ICLs by a replication-independent process that is less well-defined[57] (Figure 4). Although many components of these pathways are conserved between yeast and higher organisms, vertebrates have evolved an additional network of >20 proteins specialized for

ICL repair, called the Fanconi anemia (FA) pathway[11,58].

Replication-dependent ICL repair at the molecular level was described in a study using *Xenopus* egg extracts[59]. Replication-dependent repair is the most prevalent method by which ICLs are resolved and is restricted to the S and G2 phases of the cell cycle[60]. Repair is initiated when a replication fork collides with an ICL, resulting in the eviction of the CMG complex from the DNA[60]. The replication fork then proceeds to the -1 nucleotides of the crosslinked bases and initiates removal of the ICL by NER proteins and the Fanconi anemia Pathway proteins FANCI and FANCD2, which are mono-ubiquitylated by FANCL of the FA core complex[60] (Figure 3). Activation of the E3 ubiquitin ligase activity of FANCL is dependent on assembly of the entire core complex[61]. Without any one member of the core complex, ubiquitylation will not occur and the pathway will not be activated, resulting in the disease FA[11]. FA is a cancer predisposition disease in which the inability of the patient's cells to repair ICL lesions results in progressive bone marrow failure along with the development of solid tumors and blood cancers[62]. Inactivation of any of the 20 plus proteins involved in the FA pathway results in development of the disease with varying degrees of severity depending on the FA protein affected[11,62]. Patients with FA have a short life expectancy, as there are no effective treatment options[63]. The activation of the FA complex recruits a toolkit of nucleases that act to "unhook" or remove the lesion from the DNA, resulting in a DSB. The nucleases involved in unhooking include SLX1, MUS81/EME1, and XPF/ERCCI, and these nucleases are tethered to SLX4, a critical scaffolding protein[64]. FAN1 has also been shown to act under certain conditions[65]. Although XPF is known to be the 3'

nuclease that acts to unhook the ICL lesion (and has been shown to be able to act on the 5' side of the ICL), there still remains the possibility that there are other unidentified nucleases that could function on the 5' side of the lesion[64]. After the ICL is unhooked, translesion synthesis polymerases REV1 or polymerase γ come in to bypass the remaining lesion and the damage is repaired by homologous recombination[57].

After its identification and initial characterization, the FANCD2 associated nuclease, FAN1, was thought to be the potential nuclease that could act on the 5' side of ICL structures in the FA pathway thus acting with XPF in the unhooking step[65–67]. Although FAN1 was found to bind directly to FANCD2 and there was a decrease in cell survival when FAN1 depleted cells were treated with crosslinking agents, patients with deficient FAN1 develop a kidney disease called karyomegalic interstitital nephritis and not FA[68]. This could be due to the fact that FAN1 acts on an ICL resulting from a particular type of endogenous bifunctional reactive metabolite[69,70]. It is also possible that FAN1 has redundant function with other FA nucleases in addition to another role outside of FA repair. It is very important to note that although FAN1 was found in 2010, and has been shown to be able to unhook ICL structures in in vitro assays, the exact substrate of FAN1 has likely not been identified. Thus, additional experimentation is required.

Replication-independent repair is crucial for repairing ICL damage in post mitotic or rarely diving cells including neurons or stem cells[60,71]. Although replication-independent ICL repair has specific differences when compared to replication-dependent (classical) ICL repair, the main steps of the pathways are the same. Cells must act to recognize the

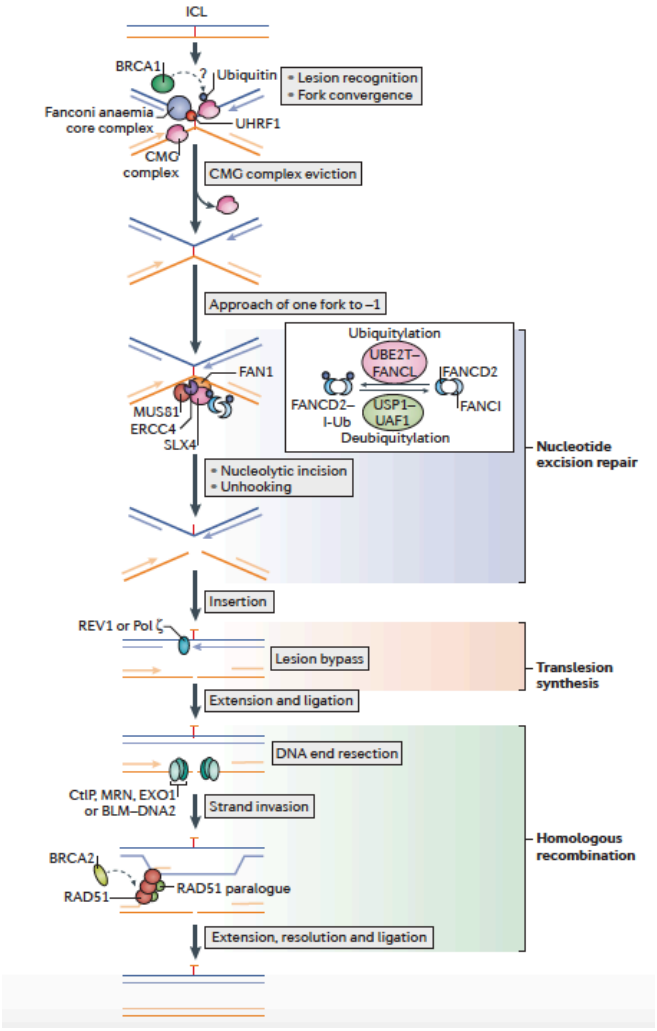


Figure from Ceccaldi, *et al.* 2016 (11).

Figure 3. Fanconi anemia-dependent interstrand crosslink repair. Cooperation of Fanconi anemia, nucleotide excision repair, translesion synthesis and homologous recombination proteins in a common interstrand crosslink repair pathway. Stalling of replication forks on DNA interstrand crosslinks (ICLs) induces lesion recognition by the FANCM-FAAP24-MHF1-MHF2 complex (not shown) and subsequent recruitment of the Fanconi anemia core complex. UHRF1 (ubiquitin-like with PHD and RING-finger domains 1) might also be involved in lesion sensing. FANCM promotes an ATR (ataxia telangiectasia and RAD3-related) kinase-dependent checkpoint response, which in turn phosphorylates and activates multiple Fanconi anemia proteins. A consequence of activation of the Fanconi anemia core complex is the monoubiquitylation of the FANCD2-FANCI (FAND2-I) heterodimer, which is promoted by the ubiquitin ligase FANCL and its partner ubiquitin-conjugating enzyme E2 T (UBE2T). ICLs in the S phase of the cell cycle impede replication fork progression, and leading strands pause 20-40 nucleotides away on either side of the ICL. Eviction of the replicative helicase CMG complex through action of breast cancer type 1 susceptibility protein (BRCA1) allows the approach of one replication fork to within one nucleotide of the ICL. Ubiquitylated

FANCD2 is directed to the ICL region, where it functions as a landing pad for the recruitment of several factors, including SLX4 and Fanconi-associated nuclease 1 (FAN1), and coordinates nucleolytic incisions that are probably mediated by ERCC4 (a structure-specific endonuclease that also functions in nucleotide exicison repair) and possibly MUS81. Unhooking the DNA leaves the complementary strand, which is bypassed by translesion synthesis polymerases such as REV1 or DNA polymerase (REV3-REV7). Ligation restores an intact DNA duplex, which functions as a template for homologous recombination-mediated repair of the double-strand break (DSB). The DNA incisions create a DSB, which is further processed by nucleases such as CtBP-interacting protein (CtIP), MRN (MRE11-RAD50-NBS1), exonuclease 1 (EXO1), and the helicase-nuclease complex BLM-DNA2 (Bloom syndrome protein-DNA replication ATP-dependent helicase/nuclease 2) that creates a single-stranded DNA (ssDNA) overhang. This ssDNA coated with replication protein a (RPA) is a substrate for RAD51-mediated strand invasion promoted by BRCA2 and subsequent homologous recombination. The USP1-UAF1 (ubiquitin carboxyl-terminal hydrolase 1-USP1-associated factor 1) complex deubiquitylates the FANCD2-I heterodimer and completes repair.

ICL damage, signal for recruitment of downstream repair proteins, and repair the lesion. Replication-independent ICL detection during transcription is limited to RNA polymerase II collision followed by recruitment of a collection of NER factors along with Cockayne Syndrome A and B (CSA and CSB)[27,60]. Lesions that result in more significant distortion of the DNA helix, such as cisplatin, require XPA, XPF, and XPG proteins while less distorting lesions such as MMC require action by XPC[56,57]. After lesion recognition and recruitment of additional factors through activation of the FA core complex, the lesion must be incised by nucleases. There is still much debate surrounding the ICL repair mechanisms that are highlighted in the pink box (left). Four steps are shown (left). Three distinct mechanisms can sense ICL lesions (indicated by a red line) (top): repair proteins recognizing distortion in the DNA helix (left column), a transcribing polymerase (middle column), or a replisome blocked by the lesion (right column). Recognition can occur through dedicated repair enzymes [nucleotide incision repair (NER) or mismatch repair (MMR)] or following collision during DNA transactions. Several nucleases are thought to participate in the incision steps (second row); however, their exact sites of action and which nucleases unhook the cross-linked base in replication-independent ICL repair are still debated (Figure 4). XPF/ERCCI is a likely candidate as it has been shown to cleave 5' and 3' of the ICL *in vitro*[57,72]. SNM1A has also been shown to act after the action of a 5' endonuclease to digest through the crosslinked base resulting in a better substrate for downstream repair[72]. Furthermore, data suggests that SNM1A has also been shown to interact with CSB, a known component of replication-independent ICL repair[27]. It will be of interest to further explore the coordination between various nucleases involved in replication-independent ICL repair.

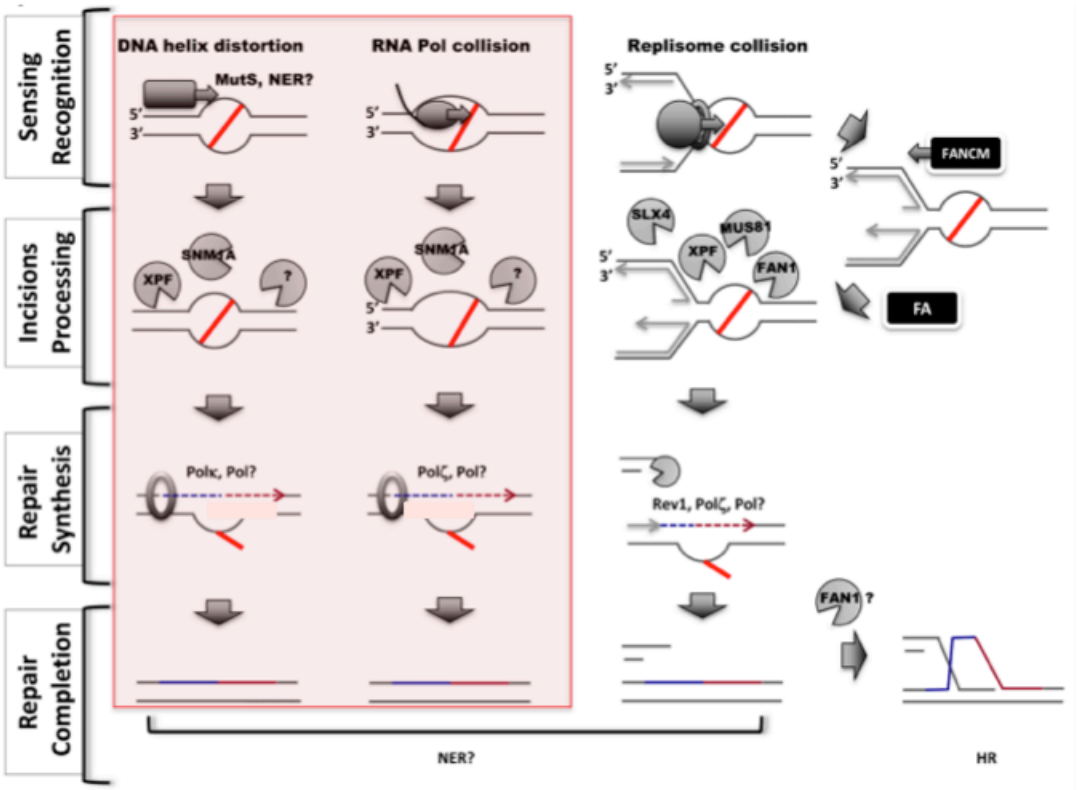


Figure from Williams *et al.* 2013 (71).

Figure 4. Pathways of interstrand crosslink (ICL) repair. Replication-independent not defined. XPR/ERCC1, by analogy with NER, is thought to act 50 of the ICLs, whereas the 30 nuclease acting in replication-independent ICL repair (RIR) is not known (?). The Fanconi anemia (FA) pathway regulates the activity of nucleases in the replication-dependent pathway. FANCM facilitates fork regression, whereas other FA proteins could recruit nucleases. Specific translesion synthesis nucleases participate in distinct repair pathways (third row): Polk and Polz in replication-independent repair and Rev1 and Polz in the replication-dependent pathway. More than one DNA polymerase could be involved in the repair of an ICL. Replication-dependent ICL repair generates a double-strand break (DSB) intermediate (bottom left), which is ultimately repaired via homologous recombination (HR).

ICL repair is also known to take place outside of replication or transcription and the proteins involved are predicted to be of MMR or NER pathways but this has not been definitively confirmed[71].

1.7 Fanconi anemia

FA is a cancer predisposition syndrome that has varied clinical phenotypes but the hallmark of the disease is bone marrow failure that initiates early in life[73] (Figure 5). FA patients have a much higher risk of solid tumors including squamous cell carcinomas and blood cancers such as myeloid leukemia[11]. The heterogeneity of physical manifestations of the disease is likely due to the fact that there are 20+ genes that can contribute to FA development[74]. Diagnosis commonly occurs after patients present with pancytopenia and are tested for sensitivity to crosslinking agents. The gold standard of FA diagnosis is to treat a sample of peripheral blood cells from the patient with crosslinking agents such as MMC or diepoxybutane (DEB) and analyze a metaphase spread for radials and other aberrations[75]. The inability to repair ICLs can result in replication fork stalling, which leads to replication fork collapse and subsequently single and double stranded breaks that create massive genomic instability in FA patients[76]. Although FA was first discovered over 90 years ago, there are still minimal effective treatment options beyond total bone marrow replacement and life expectancy for patients with FA hovers around 30-35 years[11]. One unforeseen benefit of FA research was the improvement of bone marrow transplants in general[68]. Also understanding the FA pathway, in the context of ICL repair, will be helpful in further understanding various types of breast and ovarian cancer as

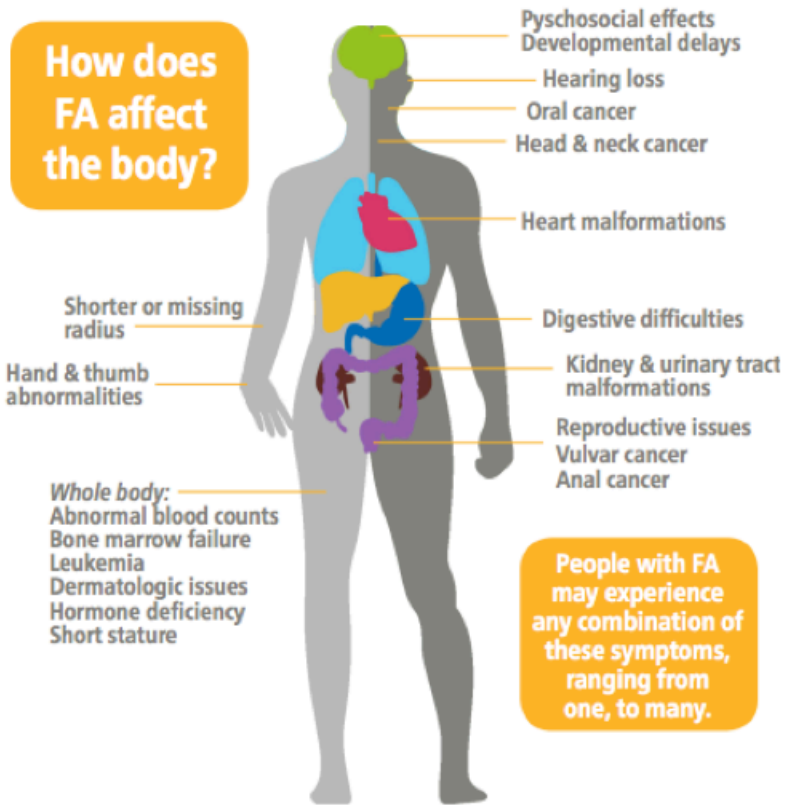


Figure from Fanconi anemia Research Fund.

Figure 5. Fanconi anemia phenotypes. Patients with FA can present with a variety of phenotypes, the most common being abnormal blood cell counts. The genetic heterogeneity of the disease contributes to the phenotypic heterogeneity. FA patients may have any combination of these possible symptoms.

some protein factors that play a role in those diseases are also FA proteins[77].

1.8 FEN1 Structure-specific Nucleases

Our discovery of SAN1, or Senataxin-associated nuclease 1, was based on a significant sequence similarity with the nuclease domains of a family of four other structure specific nucleases, including Flap endonuclease 1 (FEN1), Exonuclease 1 (EXO1), Holliday Junction 5' flap endonuclease (GEN1), and Xeroderma Pigmentosum G (XPG)[78–80]. Each member of the family has a highly conserved nuclease domain containing "N" (N-terminal) and "I" (internal) regions that constitute the active site of these proteins[81]. The N-terminal region or NH_2 region binds DNA, while the Internal region coordinates two magnesium ions via its several cysteine and glutamate residues, ultimately helping tether the enzyme to the DNA through interactions with phosphate groups on the backbone[81,82]. These two critical nuclease domain components are also separated by a spacer region. FEN1 family members also display specific substrate preferences that correspond with the various DNA replication or repair pathways they act in (Figure 6). These substrate preferences can be largely attributed to various pockets on the surface of the enzyme adjacent to the active site that probe the environment surrounding the potential substrate and provide additional recognition[80]. The crystal structure of FEN1 bound to DNA and EXO1 bound to DNA have identified previously unknown commonalities between how members of this family of proteins bind and cleave specific DNA structures[80,83] (Figure 6). FEN1 family members are regulated by post-translational modifications and through activating or inhibitory structural changes

caused by protein-protein interactions[79,84]. FEN1 family nucleases function in various aspects of DNA repair and metabolism and are critical for genomic stability.

FEN1 functions in several key replication and repair processes including Okazaki fragment processing, long-patch base excision repair, stalled replication fork rescue and telomere maintenance[78]. In addition to DNA replication and repair, FEN1 has been shown to degrade DNA during apoptotic DNA fragmentation. Apoptotic DNA fragmentation by endonucleases results in the generation of 150-180 base pair fragments of DNA[85]. Although the idea that a nuclease like FEN1 can play a role in DNA replication and also in the destruction of DNA through apoptotic DNA fragmentation may seem contradictory, these seemingly opposing actions of FEN1 can be attributed to the extensive network of FEN1 interacting proteins, cellular localization of the protein, and post-translational modifications[79]. As a nuclease on flap structures, FEN1 acts using a threading mechanism in which it first recognizes the free 5' end of single-stranded DNA of a flap. The enzyme then moves down the single-stranded DNA region to the junction of single-stranded and double-stranded DNA and after binding, threads the DNA back through the enzyme for cleavage[78]. This interesting binding mechanism is the reason that FEN1 does not act on the single-stranded DNA resulting from processing Okazaki fragments and also explains the ability for FEN1 to act as an exonuclease[86]. Although FEN1 appears to have flap endonuclease activity, exonuclease activity, and gap endonuclease activity, it has been shown that the kcat/Km of flap endonuclease activity is significantly higher than of the other two types of nuclease activity exhibited by FEN1[83]. Throughout other studies, FEN1 has been shown to interact with over 30 proteins, and

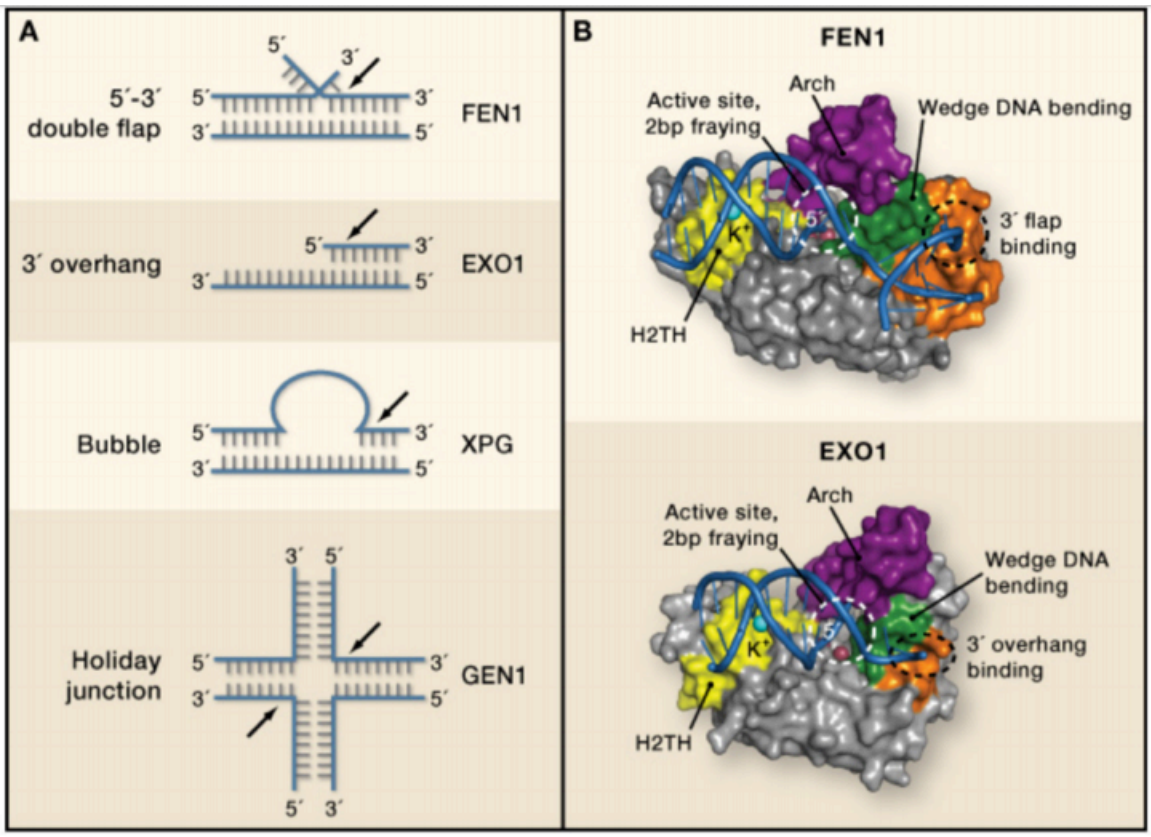


Figure from Williams, et al. 2011 (80).

Figure 6. FEN1 family member substrate preferences and nuclease domain structure. (a) Structures of preferred substrates for FEN family members. Arrows indicate cleavage. FEN1 cleaves flap structures and is active during replication, EXO1 acts on dsDNA to generate overhangs for HR, XPG acts on bubbles and other structures formed by UV and other damaging agents that are then repaired by NER, and GEN1 acts during HR to resolve Holliday Junctions (b) Structural comparison of FEN1 bound to DNA as compared to EXO1 bound to DNA. Purple: Arch region. Yellow: H2TH DNA binding motif. Green: Helical wedge. Orange: Specificity binding pocket for double-stranded DNA bearing either a 3' flap (FEN1) or a 3' overhang (EXO1).

along with the post-translational modifications that regulate the function of FEN1, this nuclease has a significant scope of activity in global genome maintenance[79,87,88]. FEN1 is also regulated by subcellular localization and has been shown to move into the nucleus based on the cell cycle and stimuli from DNA damage signals[89].

EXO1 is an exonuclease that acts in several different DNA repair pathways including MMR, HR, and alt-EJ/SSA[83,88,90]. As the only nuclease to function during MMR, EXO1 degrades long stretches of DNA where mismatches occur[91]. During HR, EXO1 functions as an end resection nuclease in which the 5' end of double stranded DNA is resected in the 5'-3' direction, generating long 3' single-stranded DNA overhangs that are bound by RPA[45,90]. Although RPA is known to bind ssDNA to protect it from degradation and also serve as a docking point for other factors involved in DNA double-strand break repair, the relationship between EXO1 and RPA is not fully understood[92]. There have been reports of RPA inhibiting EXO1 activity and there are also reports of RPA stimulating EXO1 activity[93,94]. Further experimentation is needed to fully understand the complexities of the relationship between EXO1 and RPA.

EXO1 often acts after initial processing by another nuclease, CtIP, and is activated by phosphorylation by the sensor kinase ATM. To stimulate nuclease activity during end resection, EXO1 can also form complexes with BLM, WRN, and RECQL1 helicases[45,95]. Of interest, EXO1 is a processive nuclease, giving it the capability to resect long stretches of dsDNA for homologous recombination. After homology-searching the sister chromatid, these overhangs then invade the double-stranded DNA region of the sister

chromatid resulting in D-loop formation and synthesis of new DNA[96]. As mentioned previously, end resection is the critical commitment step of HR. Thus, EXO1 is a highly important factor that acts to commit to HR[44]. In addition to exonuclease activity, EXO1 also has flap endonuclease activity similar to FEN1 and can also act on RNA[84]. EXO1 has also been shown to have a connection to familial colorectal cancer with EXO-1 deficient patients having a poor prognosis[97]. In addition to the human cancer link, EXO-1 knockout mice develop tumors, have decreased lifespan, and also have reproductive issues[98].

GEN1 is a Holliday Junction resolvase endonuclease that acts with MUS81/EME1 to process Holliday Junctions, or four-way DNA structures resulting from HR[48]. GEN1 possesses the same nuclease domain as other members of the FEN1 family and regulates cleavage of HJ structures through an initial dimerization step[49]. After dimerization on the DNA junction, the GEN1 dimer makes two nicks in the joined DNA allowing for further downstream processing and repair[99]. Correct resolution of Holliday Junction structures is important for proper DNA separation and maintaining genomic integrity[100]. Interestingly, GEN1 has a largely disordered C-terminal domain making purification of the full-length protein challenging until recently[99]. GEN1 also possesses a unique region called a chromodomain, which acts as an additional DNA binding region[49]. Chromodomains have been well established in chromatin-modeling enzymes but had not been discovered in a nuclease until GEN1. The importance of the chromodomain in the function of GEN1 is to provide additional DNA binding support to the GEN1 active site to position the DNA in a way that allows for GEN1 to properly cleave the HJ substrate[49]. GEN1 is also the only

member of the FEN1 family to have such a region, making it a unique component of GEN1 substrate recognition and processing[78].

XPG is an NER nuclease that cleaves 3' to DNA damage commonly resulting from exposure to UV light[88]. In addition to XP, mutations in XPG can lead to Cockayne syndrome (CS) which manifests as skeletal, neurological, and developmental disorders. Studies have shown that in high salt conditions XPG exists as a monomer. In lower salt conditions, XPG can dimerize and through additional experimental evidence it has been shown that XPG dimers often acts in complex with other protein factors[101].

1.9. Research Objectives

Although there have been major advances in understanding the various ways by which cells detect and repair ICLs, there are still fundamental gaps in our understanding of the mechanism of ICL repair throughout various phases of the cell cycle (Figure 4). Replication-dependent ICL repair through activation of the FA pathway is the most well-characterized method by which ICLs are detected and excised from the genome. However, this pathway is only active during S and G2 phases of the cell cycle[63]. The steps and pathway components of replication-independent ICL repair, or Fanconi anemia-independent ICL repair, on the other hand, are much less clear. Recent advances in understanding the ways by which cells can detect ICLs outside of replication, either through MMR-dependent mechanisms or through transcription-coupled ICL repair, are significant but there are still many unknowns[56,60,72]. Identifying the molecular mechanisms that orchestrate the repair of every DNA damage event that occurs in cells

every day is critical in not only understanding how these pathways are regulated but also how we can modulate them to decrease disease burden of patients afflicted with ICL repair protein diseases including FA, XP, and CS. The lack of understanding of the crossover between pathways by which ICLs can be repaired during each phase of the cell cycle is reflected by the lack of treatment options for people with these diseases and thus there is an urgent need to progress this area of research.

This dissertation aims to characterize SAN1, which is a completely novel component of ICL repair and the first nuclease involved in interstrand crosslink repair to be discovered since FAN1 was identified in 2010[65,67]. Preliminary characterization of the protein as a nuclease raised many questions about the potential involvement of SAN1 in DNA repair. Although there are many known nucleases that act in the various DNA damage response pathways there are still unknown components of these pathways that may act under certain conditions or have redundancy with other known nucleases. There are also significant gaps in the understanding of the molecular mechanisms of replication-independent ICL repair pathways. This dissertation work builds upon the initial discovery of the previously uncharacterized protein, SAN1, and explores our understanding of the unique nuclease activity of the protein. This work also describes the complex relationship between the structure and function of a newly identified member of the FEN1 family.

CHAPTER 2: SAN1 Nuclease Activity and Substrate Preferences

2.1 Abstract

As the most recently discovered member of the FEN1 family of structure-specific nucleases, SAN1 has significant similarities and unique differences when compared to other family members. As an exonuclease, SAN1 is able to digest 2 or 7 nucleotide fragments from ssDNA structures of larger than 25 nucleotides and requires a free 5' end of DNA for loading onto the DNA and subsequent cleavage. Also of note, SAN1 cleaves ssDNA structures with different efficiencies generating a distinct laddering pattern for each sequence. SAN1 has a conserved nuclease domain that is homologous to other FEN1 family members, a unique C-terminus that was previously uncharacterized but has been shown to be required for nuclease activity, and an internal repeat region that is not required for nuclease activity. SAN1 preferentially cleaves DNA with splayed arm structures but will also cleave 5' ssDNA flaps, albeit with less efficiency than single-stranded DNA or splayed arm structures. In contrast to other FEN1 family members, SAN1 does not cleave nicked or gapped regions of DNA efficiently. Taken together, these results describe the characterization of a protein that was previously completely uncharacterized and provide evidence for a unique member of an already diverse nuclease family.

2.2 Introduction

Members of the FEN1 family of structure specific nucleases have roles in many aspects of DNA replication and repair and are thus critical components of genome

integrity[102]. Although each member of the family acts on different substrates, the catalytic activity of these nucleases is very similar. FEN1 family members have a highly conserved nuclease domain with the N (NH2) and I (Internal) regions of the nuclease domain separated by a spacer region[88]. The N region is critical for DNA binding and the I region binds two Mg^{2+} ions that ultimately stimulates substrate binding and subsequent DNA cleavage[83]. EXO1 and GEN1, for example, have been shown to resect double-stranded DNA for strand invasion during HR and to resolve Holliday Junctions that result after HR, respectively[81,99]. XPG is involved in processing damage repaired by NER and FEN1 is involved long-patch BER[101]. In recent years, our understanding of the nuclease activity of each of these unique enzymes has increased exponentially. Until SAN1, the novel FEN1 family nuclease described in this study, there had been, to date, no evidence of involvement of any other FEN1 family members in ICL repair.

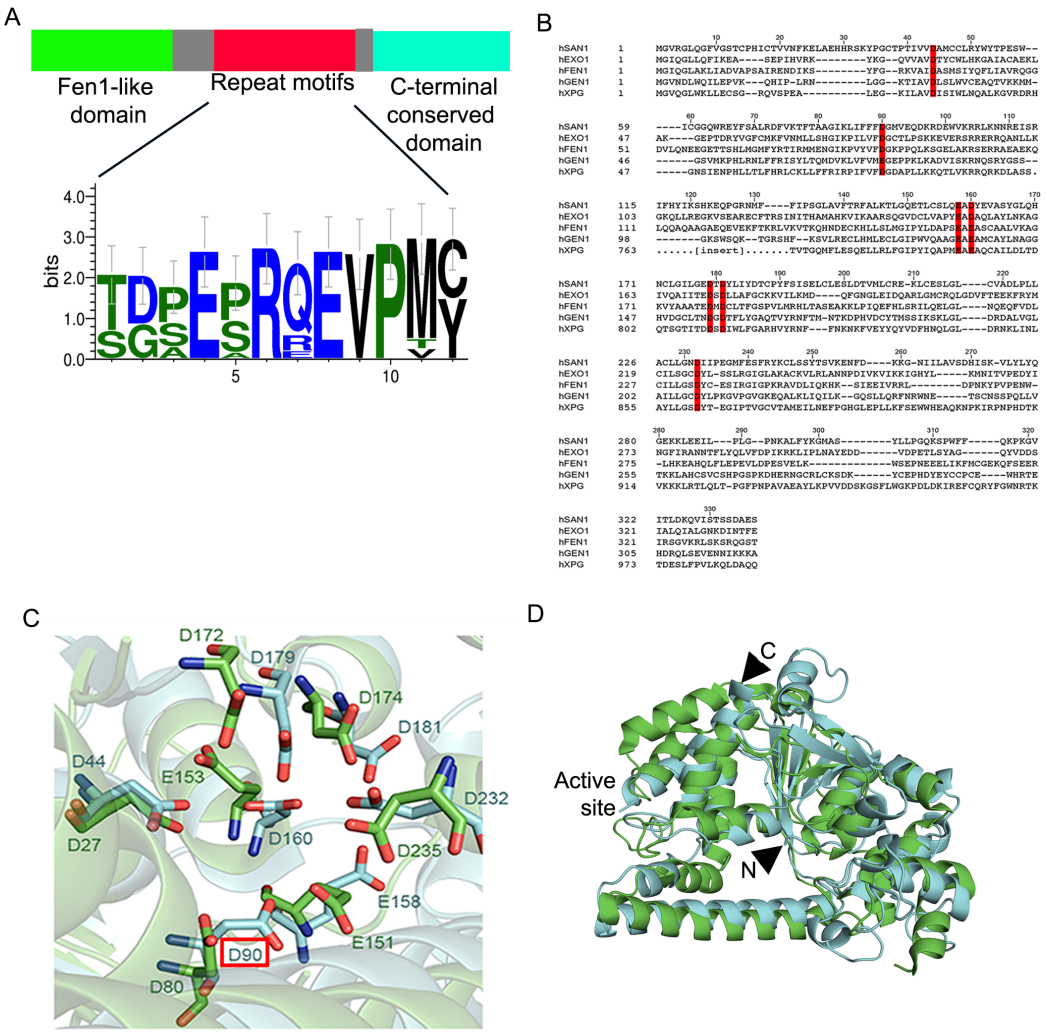
Our knowledge of the DNA repair machinery remains incomplete, and additional nucleases might exist for removal of specific lesions. In this study we identified an uncharacterized protein that contains an N-terminal domain closely related to the FEN1 family of structure-specific nucleases. This protein is not a known component of any DNA repair complex, but we now report that it is a 5'exonuclease for single-stranded DNA, and is required for the cellular response to ICLs.

2.3 Results and Discussion

Previous work identified an uncharacterized protein with sequence homology to the FEN1 family of structure-specific nucleases (Figure 7a), which we named SAN1, or

Senataxin-associated nuclease 1. The protein shares with the other members of the FEN1 family (FEN1, EXO1, XPG, GEN1) conserved acidic residues that constitute the active site of these proteins (Figure 7b). Using the Robetta server, these acidic residues were modeled to the crystal structure of FEN1 (Figures 7c, d). Acidic active site residues are present in the N and I regions of the nuclease domains of this family of proteins and are responsible for coordinating substrate binding and metal cofactor binding along with catalyzing the reaction. Using these modeling data the conserved aspartic acid 90 residue was mutated to alanine generating a D90A mutant of SAN1. The D90A point mutant of SAN1 was hypothesized to eliminate catalytic activity. In addition to the conserved FEN1 nuclease domain, SAN1 also has a unique internal repeat region, and a C-terminus that is conserved among SAN1 homologs but is of unknown function (Figure 7a). The internal repeat region consists of about twelve 15- amino acid repeat motifs with the general sequence of QEVPM that are predicted to be largely unstructured and are not found in any other protein (Figure 7a).

To determine if SAN1 was able to act as a nuclease, WT SAN1 or D90A SAN1 were tested for activity against 5' ^{32}P labeled ssDNA or 3' ^{32}P labeled ssDNA substrates after affinity purifying the FLAG-tagged protein from mammalian 293T cells (Figure 8a). Interestingly, when WT SAN1 is incubated with a 5' ^{32}P labeled ssDNA oligonucleotide and the results are visualized on a denaturing urea gel, it is clear that SAN1 cleaves ssDNA liberating a ~3 nucleotide and ~7 nucleotide ^{32}P labeled fragment from the DNA (Figure 8b). Importantly, the D90A SAN1 mutant showed no activity on the 5' or 3' ^{32}P ssDNA oligonucleotide confirming that nuclease activity can be completely abolished by a single point mutation. Incubating WT SAN1 with 3' ^{32}P ssDNA in a nuclease assay



Tim Errington

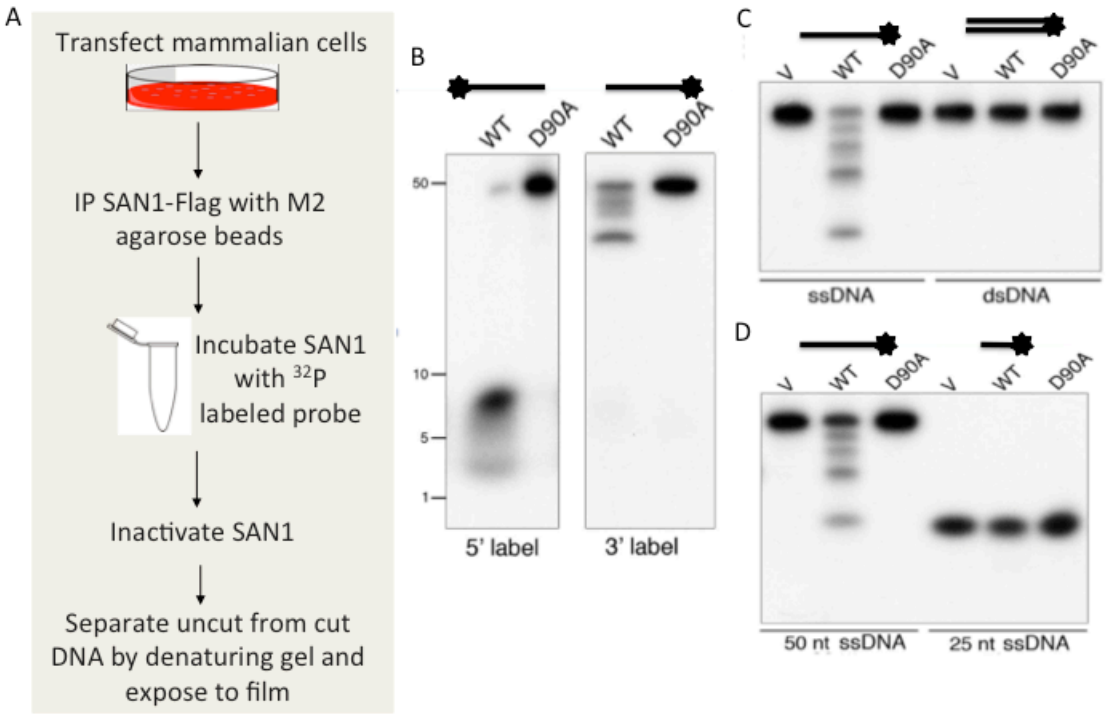
Figure 7. SAN1 homology to the FEN1 family of structure specific nucleases. (a) Schematic of the domain architecture of SAN1 including the N-terminal nuclease domain, central repeat region of around twelve 15-amino acid repeats, and conserved C-terminus. (b) Sequence alignment of nuclease domains of human SAN1, EXO1, FEN1, GEN1 and XPG. Identical and similar residues are in bold. Conserved acidic residues in the active site are highlighted in red**.** (c) Modeling of conserved carboxylates in active site of SAN1 (light blue), using the Robetta server (http://robetta.bakerlab.org) and the *A. fulgidus* FEN1 structure (PDB 1RXW) (green) as template. Residue highlighted by red box is the aspartate mutated to make D90A. (d) The nuclease domain sequence of SAN1 was submitted to the Robetta server (http://robetta.bakerlab.org) and the server found a confident match to the A. fulgidus FEN1 protein, whose structure was used as a template for comparative modeling. SAN1 model (light blue) and FEN1 template (PDB 1RXW)

(green) were aligned and visualized with the PyMol Molecular Graphics System, Version 1.3 Schrödinger, LLC.

generated a ladder pattern in which WT SAN1 cleaves the 50 nucleotide sequence down to about 25 nucleotides but will not cut further (Figure 8b). In contrast to EXO1 which readily resects dsDNA during HR, SAN1 is not active on a dsDNA structure (Figure 8c). SAN1 also showed activity against a 3' ^{32}P ssDNA oligonucleotide of 50 nucleotides, cleaving the sequence down to about 25 nucleotides but showed no activity against a 25 nucleotide sequence (Figure 8d). Together these results provide an initial understanding of the unique nuclease function of SAN1. SAN1 has homology to the FEN1 family of structure specific nucleases and WT SAN1 exhibits 5' exonuclease activity on ssDNA of greater than 25 nucleotides only. D90A SAN1 exhibits no detectable nuclease activity, providing evidence that the single-point mutation of WT SAN1 is sufficient to abolish nuclease activity and to confirm that the nuclease activity observed is attributable to SAN1.

After the initial discovery of SAN1 as a new member of the FEN1 family of structure-specific nucleases and that SAN1 is able to act as a nuclease on ssDNA substrates of larger than 25 nucleotides (Figure 8), a key goal was to further investigate the properties of this novel enzyme. A first step was to confirm that the nuclease activity is endogenous to SAN1 rather than to a contaminating cofactor, using recombinant protein expressed in bacteria.

A tetracycline-inducible WT or D90A SAN1 construct transformed into BL21 DE3 RIPL cells was used to express SAN1. These cells allow increased expression of mammalian proteins due to the expression of tRNAs of amino acids more commonly found in mammalian proteins. Mouse SAN1 (mSAN1) was selected instead of human



Tim Errington

FIGURE 8. SAN1 acts as an exonuclease on 5' ssDNA. (a) Schematic of WT SAN1-FLAG or D90A SAN1-FLAG purification from 293T cells. Nuclease assays with affinity purified WT SAN1-FLAG or D90A SAN1-FLAG and various substrates (b) 5' ^{32}P labeled 50 nucleotide ssDNA as compared to 3' ^{32}P labeled 50 nucleotide ssDNA (c) 3' ^{32}P labeled 50 nucleotide ssDNA as compared to 3' ^{32}P labeled 50 nucleotide dsDNA (d) 3' ^{32}P labeled 50 nucleotide ssDNA as compared to 3' ^{32}P labeled 25 nucleotide ssDNA.

SAN1 (hSAN1) because of its smaller size. Although the nuclease domains are highly similar mSAN1 possesses fewer internal repeat motifs than hSAN1 (Figure 9a).

Expression of the mSAN1 protein was induced with tetracycline at 18°C overnight followed by lysis using a French Press and purification of the cell lysate over a Strep-Tactin column with desthiobiotin elution (Figure 9b). To verify mSAN1 expression and purification, an SDS gel of WT mSAN1 and D90A mSAN1 were stained with Coomassie brilliant blue and the signals were compared to BSA standards (Figure 9c). WT mSAN1 and D90A mSAN1 were expressed, captured, and eluted from the Strep-Tactin beads (Figure 9c).

Activity of the recombinant mSAN1 was tested on 60 nucleotide ssDNA substrates X1 and X4 (shown previously in Figure 8 and listed in Appendix II). Recombinant WT mSAN1 or D90A mSAN1 was incubated with radiolabeled probe and a Start Buffer containing magnesium (a vital cofactor for FEN1 family proteins) for 2 hrs at 37°C. Samples were then analyzed on a denaturing urea gel and products were detected by autoradiography. As shown in Figure 9d, WT mSAN1 is able to cleave both X1 and X4 substrates and the D90A mSAN1 is unable to cleave either substrate. This result supports our proposal that SAN1 is indeed responsible for the nuclease activity observed on ssDNA substrates.

However, although bacterial expression was effective in providing additional support that the nuclease activity on ssDNA substrates was specific to SAN1, the method

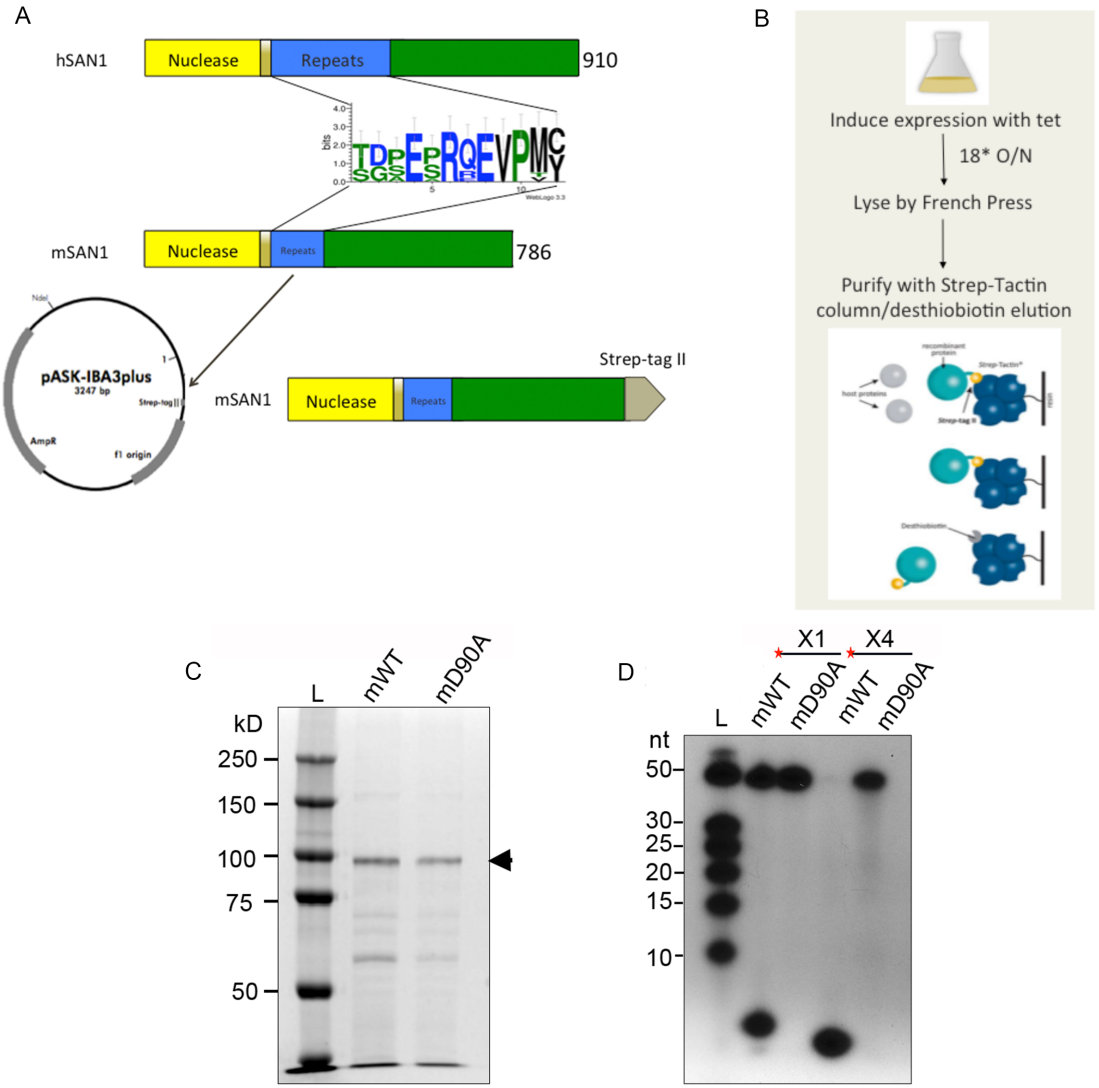


FIGURE 9. Recombinant mouse WT SAN1 shows similar nuclease activity to human WT SAN1. (a) Schematic of human SAN1 compared to mouse WT SAN1 (mSAN1). (b) Bacterial purification scheme of mSAN1 produced from tetracycline inducible BL31 DE3 (RIPL) cells. (c) Murine SAN1 was expressed with a C-terminal Strep tag in *E. coli* and purified over Strep-Tactin beads. Purified protein (0.2 ug) was analysed by PAGE and stained with Coomassie brilliant blue. Arrow shows mSAN1 expected size of 100 kD. (d) Synthetic 50-mer oligonucleotides were 5'-labeled with ^{32}P and incubated with mSAN1 for 120 min. Products were separated by PAGE and ^{32}P-fragments were detected by autoradiography (see Appendix II for sequences).

produced only a small quantity of protein that was unstable in solution, and thus was not ideal for further characterization of the nuclease profile of SAN1. To address this I utilized a two-step purification system of WT SAN1 or D90A SAN1 purified from 293T cells, described in Figure 10a. To further confirm that the human SAN1 purified from transfected 293T cells is an active nuclease, I developed a two-step purification scheme using the $SAN1^{-/-}$ HeLa cell line stably expressing SAN1 SSF (Strep-Strep-FLAG) tag. $SAN1^{-/-}$ HeLa + SAN1 SSF cells were lysed and centrifuged, and the soluble fraction of the cell lysate was harvested. The soluble fraction was then incubated with Strep-Tactin beads, washed three times, and eluted with desthiobiotin for the first round of purification. The elution from the Strep-Tactin beads was then incubated with FLAG beads, washed three times, and eluted with FLAG peptide (Figure 10a). Fractions from each wash and elution step were saved for silver stain analysis and western blot analysis (Figures 10b, c). The major protein band eluted from each affinity purification step corresponded to the expected molecular weight of hSAN1 (150 kD). FLAG antibodies readily detected a protein of 150 kD in the elution fractions, which corresponds to hSAN1, and the wash fractions had no detectable SAN1 (Figure 10c).

To quantify nuclease activity, I developed a spin assay in which the affinity purified SAN1 was incubated with a 5' ^{32}P labeled probe before spinning the reaction through a Qiagen Nucleotide Removal Column. Cleaved product, which consists of 2-7 nucleotide ^{32}P labeled fragments, is eluted from the column, while the substrate (50 nucleotides) is retained (Figure 10f). Elution fractions 1 and 2 (E1 and E2) showed a high concentration of radiolabeled product, indicating that the fractions that correspond to SAN1 by western

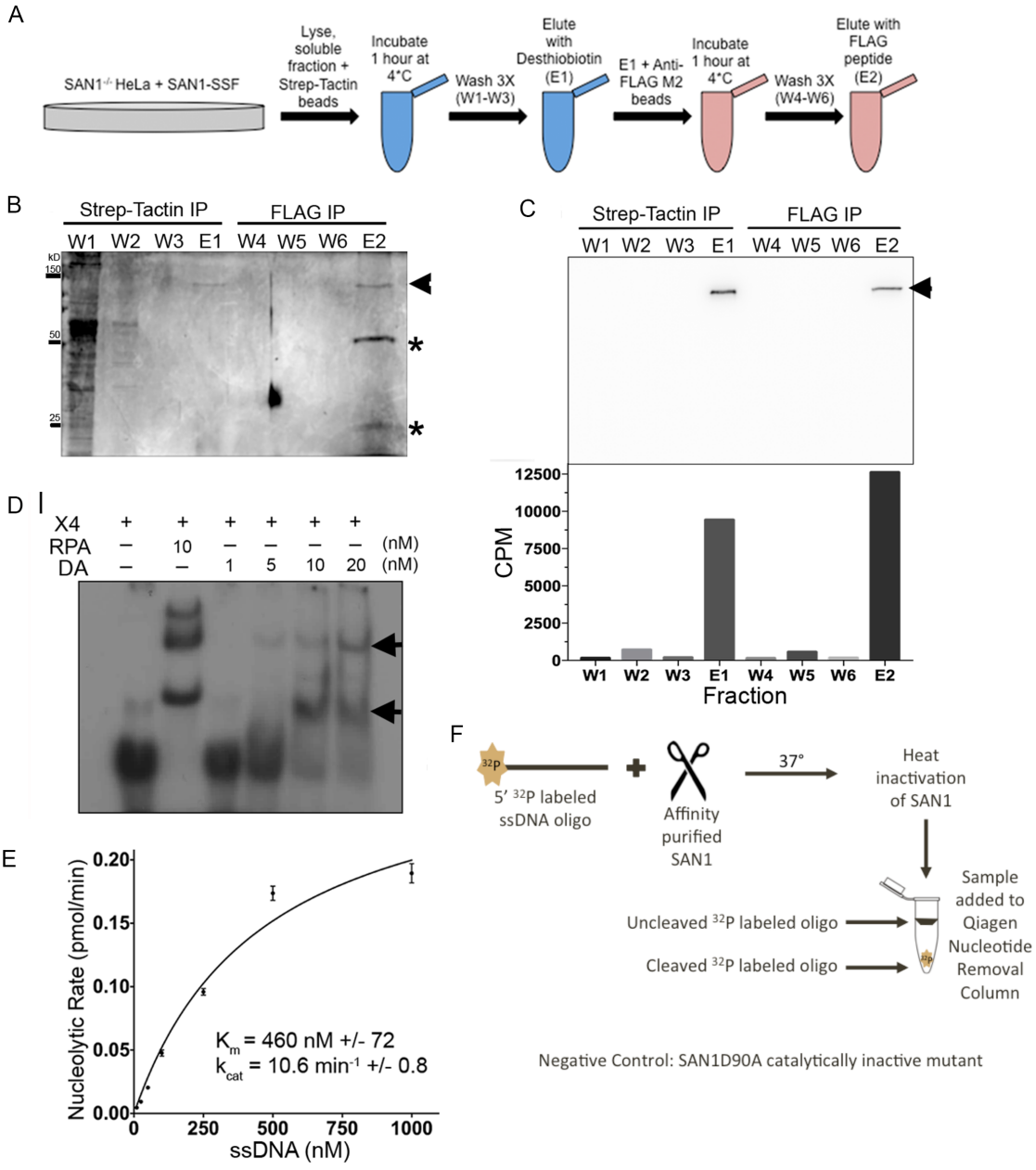


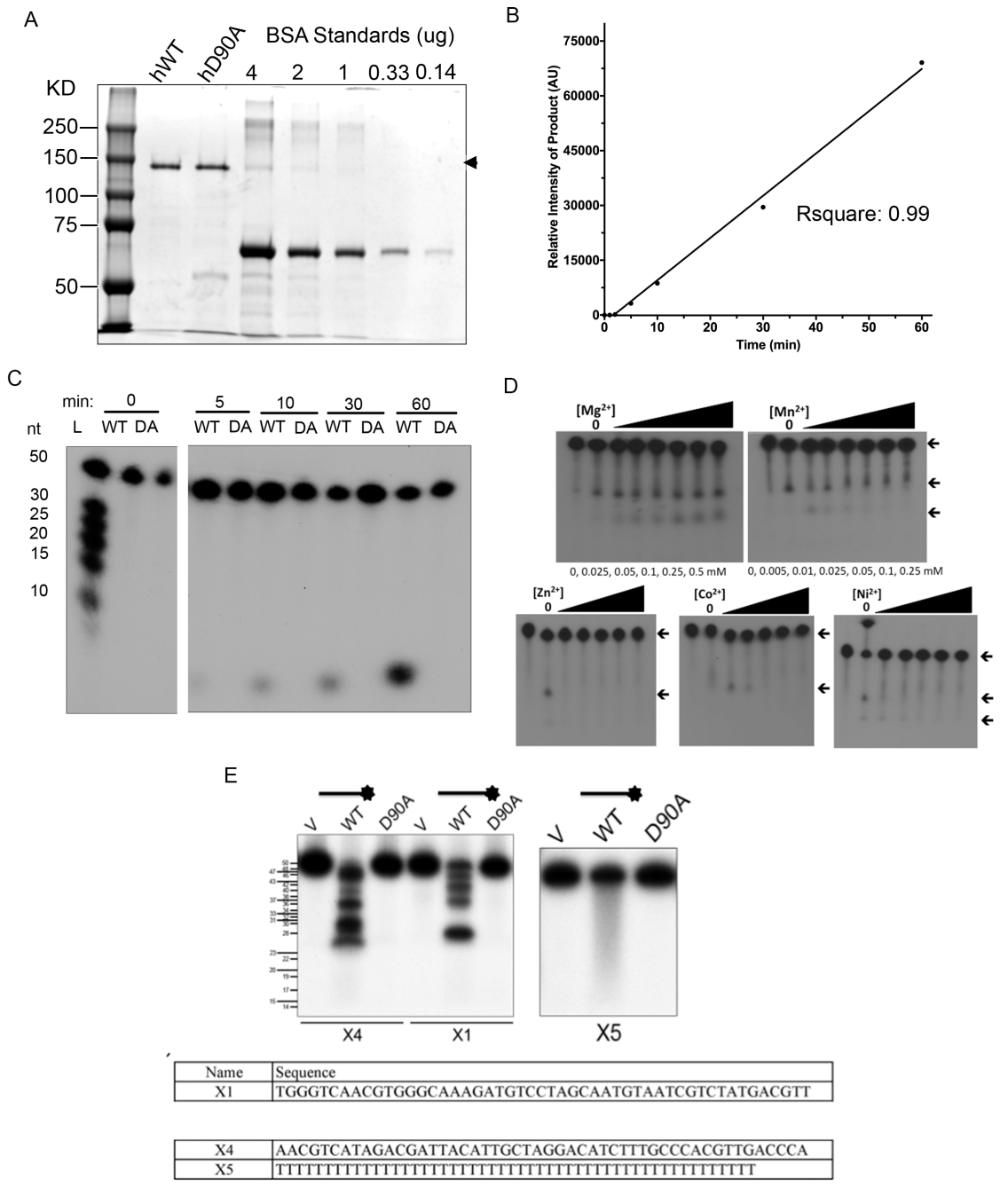
FIGURE 10. Two-step affinity purified hWT SAN1 shows nuclease activity in spin assay. (a) Schematic of double-affinity Strep-FLAG tag purification for human SAN1 WT and SAN1 DA. (b) Silver stained fractions from the purification where "W" denotes Wash steps and "E" denotes Elution steps for the Strep and FLAG IPs. Arrow shows human SAN1 WT (expected size 150 kD) and asterisks show FLAG antibody heavy and light chains. (c) Top panel shows immunoblot of fractions from two-step purification of SAN1 where arrow shows SAN1 (expected size 150 kD), detected using mouse M2 anti-FLAG-antibody. Bottom panel shows corresponding filter spin nuclease assay. (d) 50 nucleotide X4 was 5' ^{32}P labeled and incubated with RPA as a positive control or

increasing concentrations of SAN1 D90A (catalytically inactive). Samples were analyzed on a native gel and exposed to X-ray film. (e) Using the filter spin assay, initial rates of 5' ^{32}P-labeled X4 hydrolysis were measured at different substrate concentrations. Line was fitted using Prism software, assuming Michaelis-Menten kinetics. (f) Spin nuclease assay. Affinity purified WT SAN1 is added to a 5' radiolabeled probe (50nt) before being added to a Qiagen Nucleotide Removal Column. Cleaved substrates will be present in the elution while uncleaved substrates will remain bound to the filter.

blot also have nuclease activity against ssDNA (Figure 10c). These data, combined with the recombinant SAN1 protein purification and corresponding nuclease activity, confirm that the nuclease activity observed in our assays can be attributed specifically to the SAN1 nuclease, and not to any contaminating protein.

To determine if the D90A SAN1 point mutant, used as a negative control, was able to bind DNA, affinity purified D90A SAN1 was tested in a mobility shift assay (EMSA). The protein was labeled with ^{32}P-ssDNA in binding buffer and analyzed by native gel electrophoresis. As a positive control, the single-stranded DNA binding protein RPA was used (Figure 9d). As shown in Figure 10d, there is an upward shift up in the location of the visualized radiolabeled ssDNA substrate (similar to RPA) with increasing concentrations of D90A SAN1. These data demonstrate that the point mutation in Asp90 abolishes catalytic activity without disrupting the structure of the nuclease.

I next examined whether the kinetic properties of SAN1 nuclease activity were similar to those of other FEN1 family members. Using the filter spin assay and affinity purified WT SAN1, initial rates of 5' ^{32}P-labeled X4 hydrolysis were measured at different substrate concentrations. Affinity-purified hSAN1 was used for this assay and the protein concentration was quantified using Coomassie brilliant blue stain of samples as compared to BSA controls of known concentration (Figure 11a). Importantly, this assay was completed during the linear phase of the nuclease reaction as shown in Figure 11b. The line was fitted using Prism software assuming Michaelis-Menten kinetics. These data demonstrate that SAN1 has a similar kinetic profile to human EXO1 with a Km of 460



Name	Sequence
X1	TGGGTCAACGTGGGCAAAGATGTCCTAGCAATGTAATCGTCTATGACGTT

X4	AACGTCATAGACGATTACATTGCTAGGACATCTTTGCCCACGTTGACCCA
X5	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

Tim Errington – panel E

FIGURE 11. SAN1 purification from mammalian cells and substrate and cofactor characterization. (a) Affinity purified hWT SAN1 and hDA SAN1 from 293T cells as compared to BSA standards visualized by Coomassie brilliant blue stain. Arrow shows hSAN1 expected size of 150 kD. (b) Time course of SAN1 WT in linear range of the nuclease reaction with ssDNA X1 as a substrate. (c) Time course of 5' ^{32}P-labeled ssDNA X1 digest. (d) WT SAN1 was incubated with increasing concentrations of various metal ions and assayed for activity against a 5' ^{32}P ssDNA substrate (e) V, WT, or D90A

SAN1 was incubated with ssDNA 3’ ^{32}P ssDNA probes (sequences showed below). X4 and X1 are random sequences and X5 is a homopolymer of T’s of the same length.

nM +/- 72 and a kcat of 10.6 min^-1 +/- 0.8 [83] (Figure 10e). To test whether larger, intermediate products are formed during the SAN1 nuclease assay, I performed a time course experiment where radiolabeled ssDNA was incubated with WT SAN1 over a time course from 0 min to 60 min. From the results of this assay, it was clear that no intermediate products are produced at short incubation times (Figure 11c). To determine whether SAN1 utilizes Mg^{2+} or another cation as the preferred metal catalytic cofactor, various divalent metal ions were titrated into the nuclease reaction with affinity purified SAN1 and the nuclease activity on a ssDNA substrate was measured (Figure 11d). After separating the products on a denaturing urea gel and imaging the products formed using autoradiography, I observed the largest increase in radiolabeled product in the reaction where increasing concentrations of Mg^{2+} were added (Figure 11d). Other members of the FEN1 family also prefer Mg^{2+} as the metal cofactor, providing further evidence that although SAN1 has differences when compared to members of the family, catalytically the enzymes function very similarly.

When SAN1 is incubated with a 3' ^{32}P labeled ssDNA probe, it generated products that resemble a ladder. To probe this result deeper, SAN1 nuclease activity was tested against several additional ssDNA oligonucleotides with varying sequence (Figure 11e). X1 and X4 are 50 nucleotide ssDNA sequences and figure X5 is a homopolymer of T's. From these assays it was clear that SAN1 cleaves different sequences with different efficiencies and will not cut a homopolymer of T's. A homopolymer of T's was used over other sequences as other homopolymers form higher order structures that would block SAN1 nuclease activity. From these data, I propose that SAN1 may have some sequence preference in that it more easily cuts some sequences over others but to date I do not

know the full extent of what this varying substrate preference means biologically for the nuclease.

After confirming that the nuclease activity observed in previous assays was specific to SAN1, it was important to understand how SAN1 recognized and bound its preferred substrate. To address how SAN1 recognized substrates, and specifically, single-stranded DNA I designed a modified oligonucleotide. 3' ^{32}P labeled X4 (a known substrate of SAN1) and 3' ^{32}P labeled X4 modified with a 5' biotin covalently linked to the oligonucleotide (Figure 12a) were utilized in a nuclease assay. The ability of SAN1 to cleave 2-7 nucleotides from the 5' end of ssDNA left us with the question of whether SAN1 was truly acting as an exonuclease (requiring a free 5' end of DNA) or if it was acting as an endonuclease (able to bind and cleave in the middle of the sequence). Members of the FEN1 family have both endonuclease and exonuclease activity, so determining how SAN1 recognizes and cleaves substrates was important to understanding what types of structures it might recognize *in vivo*. The results from this assay show that WT SAN1 cleaves the 3' ^{32}P labeled un-biotinylated substrate but does not cleave the 3' ^{32}P labeled 5' biotinylated substrate (Figure 12a). The 5' biotin blocks SAN1 from recognizing the free 5' phosphate of the ssDNA and thus blocks all activity, confirming that SAN1 acts as an exonuclease and likely requires a free 5' end of DNA for binding.

After determining that SAN1 requires a free 5' end to recognize potential substrates, it was crucial to further characterize additional SAN1 substrates, including structures that could be regular substrates of nucleases during DNA replication or repair. The first structure tested was an overhang structure where 20 nucleotide of ssDNA was followed

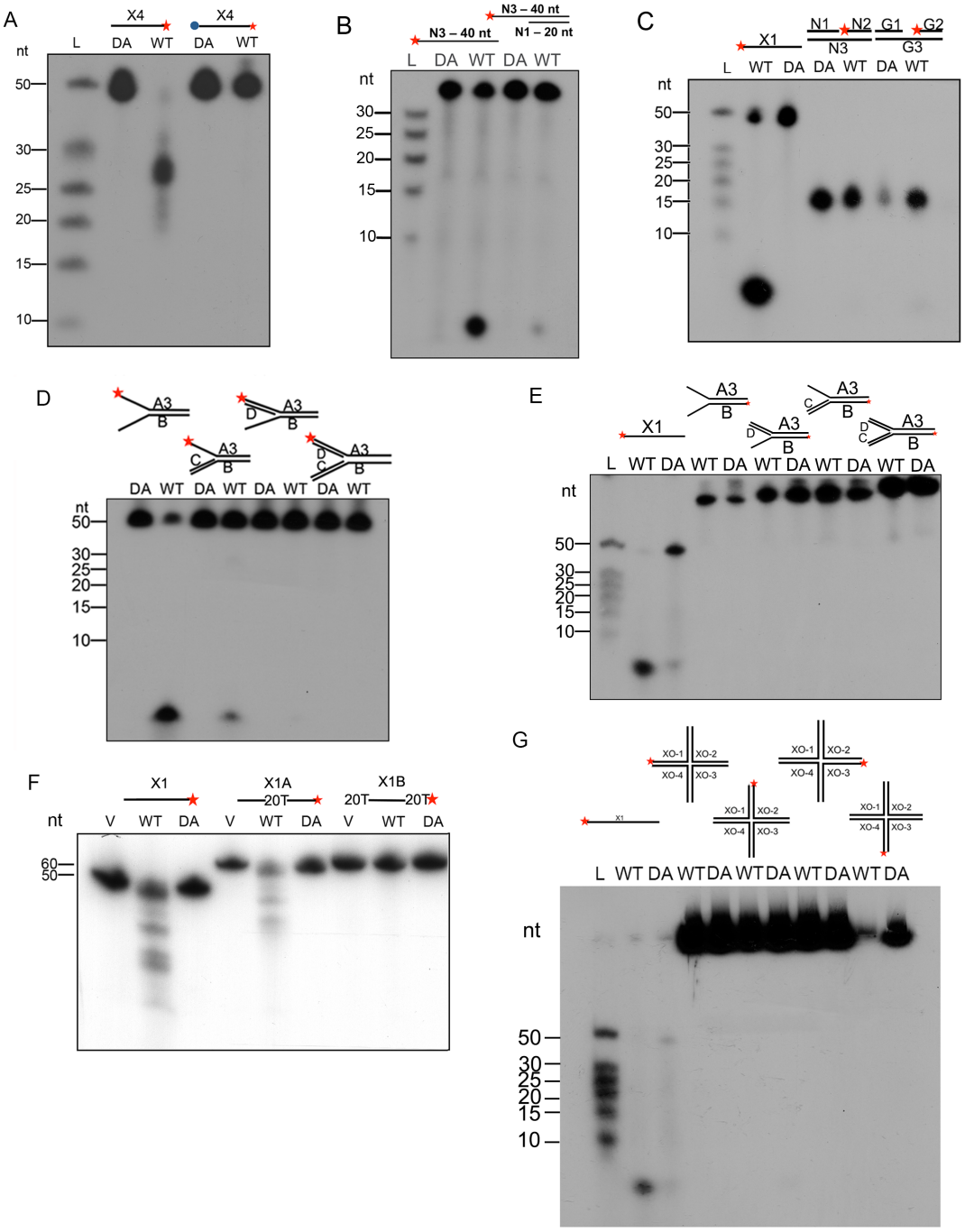


FIGURE 12. SAN1 cleaves 5' ssDNA substrates including splayed arm and flap structures. (a) Affinity-purified FLAG-SAN1 WT or D90A from 293T cells incubated with 5' biotinylated X4 and unbiotinylated X4 (b) Affinity-purified FLAG-SAN1 WT or D90A from 293T cells incubated with 40 nucleotide ssDNA versus a 20 nucleotide ssDNA 5' overhang followed by 20 bp of dsDNA (c) Affinity-purified FLAG-SAN1 WT or D90A from 293T cells incubated with dsDNA oligonucleotides with an internal nick or gap; (d) FLAG-tagged SAN1 WT or D90A was incubated with 5' ^{32}P labeled splayed

duplex, 3’ flap, or 5’ flap structures for 2 hrs at 37°C. Products were separated by PAGE and ^{32}P-fragments were detected by autoradiography (see Appendix II for sequences) (e) SAN1 WT or SAN1 DA incubated with splayed duplex, 3’ flap, 5’ flap, and replication fork structures where the bottom strand (B) is 5’ ^{32}P labeled. (f) Affinity-purified FLAG-SAN1 WT or -D90A from 293T cells incubated with variants of ssDNA oligonucleotide X1 with 20 Ts 5’ and 3’ to 20 nts of X1, or a tract of 20 Ts bounded by two 20 nucleotide sections of the X1 sequence all for 2 hrs at 37°C. Substrates were 3’ ^{32}P labeled. (g) Affinity purified FLAG-tagged WT or DA SAN1 from 293T cells was incubated 5’ ^{32}P labeled Holliday Junction structures and products were separated by denaturing gel electrophoresis and ^{32}P fragments were detected by autoradiography.

by 20 base pairs of dsDNA, generating a 20 nucleotide overhang of 5' ssDNA (Figure 12b). As the ssDNA length requirements for SAN1 to cleave are greater than 25 nucleotide, I predicted that this ssDNA overhang of 20 nucleotide would not be cut by SAN1. Other members of the FEN1 family including FEN1, XPG, and GEN1 recognize ssDNA/dsDNA junctions and thus it may be possible that SAN1 recognizes a similar type of structure. The results for this assay are shown in Figure 12b where WT SAN1 is active on the 40 nucleotide ssDNA substrate and displays some activity on the 20 nucleotide overhang structure but significantly less activity than on the ssDNA substrate. Although SAN1 was partially able to cleave the 5' 20 nucleotide ssDNA overhang, this is likely not the ideal substrate for the SAN1 nuclease as the nuclease activity observed on the overhang substrate was significantly less than the activity observed previously on ssDNA substrates.

The next potential SAN1 substrates that were tested were a nick structure and a gap structure. Nicks arise often in DNA replication and repair and FEN1 and EXO1 are both able to cleave nicked and gapped structures[49,83]. Nicks and gaps can arise as a product of DNA damage or as a DNA repair intermediate that must be processed by additional nucleases and ligases to repair the lesion. A nick in the DNA is formed when the phosphate backbone is cleaved but there are no missing bases. A gap, on the other hand, is a structure that can be one missing base to several missing bases. The assay in Figure 12c shows WT or D90A SAN1 tested against a nicked structure or a single-base gap structure, and it is clear that SAN1 does not cleave nick or single-base gap structures. This is a very interesting departure in activity from other members of the FEN1 family and sets the nuclease cleavage capabilities of SAN1 apart from the nuclease activity of

FEN1 and EXO1.

After determining that SAN1 does not cleave nicks and gaps like other members of the FEN1 family, we then tested several other substrates that are known to be cleaved by FEN1, EXO1, XPG, and GEN1. A splayed arm, 5' flap, 3' flap, and replication fork structures with the top strand of DNA 5' ^{32}P labeled or the bottom strand of DNA 5' ^{32}P labeled were tested by nuclease assay (Figures 12d, e). When the top strand was 5' ^{32}P labeled and the entire structure was incubated with WT SAN1 I observed cleavage of the splayed arm and 5' flap structures but not the 3' flap or replication fork structures (Figure 12d). It should be noted that the splayed arm was more efficiently cleaved than the 5' flap structure. When the bottom strand of the DNA was 5' ^{32}P labeled there was no detectable nuclease activity confirming previous results that showed that SAN1 is not active on dsDNA (Figure 11e). FEN1 cleaves splayed arm and flap structures readily and also EXO1 is able to cleave splayed arm and flap structures with less efficiency than FEN1.

In addition to determining whether SAN1 was active on replication and repair structures, it was of interest to further understand if SAN1 has a sequence recognition component of substrate binding. Utilizing the original data showing that SAN1 requires longer than 25 nucleotides of ssDNA to cleave the substrate as a tool, additional sequences were generated. Originally it was shown that SAN1 would not cut a 25 nucleotide sequence of ssDNA but would cut a 50 nucleotide sequence of ssDNA down to about 25 nucleotides. Two different sequences were designed: one with 20 nucleotides of regular sequence (with A, T, C, G bases) followed by 20 nucleotides of Ts only, followed by 20 nucleotides of regular sequence and the other sequence with 20 nucleotides of Ts, 20 nucleotides of regular sequence, followed by 20 nucleotides of Ts.

If SAN1 is not recognizing the entire stretch of DNA, but also requires greater than 25 nucleotide of sequence to bind I predicted that SAN1 would be able to cleave only one of these two oligonucleotide sequences. Strikingly, SAN1 was able to cleave the 3' ^{32}P labeled oligonucleotide with 20 nucleotide of internal Ts but not the flanking Ts oligonucleotide (Figure 12f). This result supports a model in which SAN1 recognizes the initial sequence (up to about 20 nucleotides) for binding and cleavage and the sequence following is unimportant.

After determining that SAN1 cleaves splayed arm structures preferentially over other tested structures, I tested two other known FEN1 family substrates that also happen to be DNA repair intermediates. The first was a Holliday Junction, which is a four-stranded DNA structure that results from homologous recombination and must be repaired by enzymes, such as the GEN1 resolvase, for complete DNA repair. As shown in Figure 12g, SAN1 did not cleave the Holliday Junction structure irrespective of which strand of the four-stranded DNA structure was ^{32}P labeled. It is interesting to note that SAN1 does not cleave this specific HR intermediate (the HJ), making the nuclease activity of SAN1 unique from GEN1. In a previous experiment, SAN1 nuclease activity was also tested against a bubble structure, which is a known substrate of another FEN1 family member, XPG. SAN1 was unable to cut the bubble structure (data not shown). Taken together, these results provide evidence for SAN1 having a very intriguing substrate recognition scheme, cleavage activity, and limit the pathway possibilities of the role of SAN1 as a nuclease *in vivo*.

These results along with the initial enzymatic data on SAN1 provide evidence that it is a unique nuclease. SAN1 acts as a 5' ssDNA exonuclease on substrates longer than 25

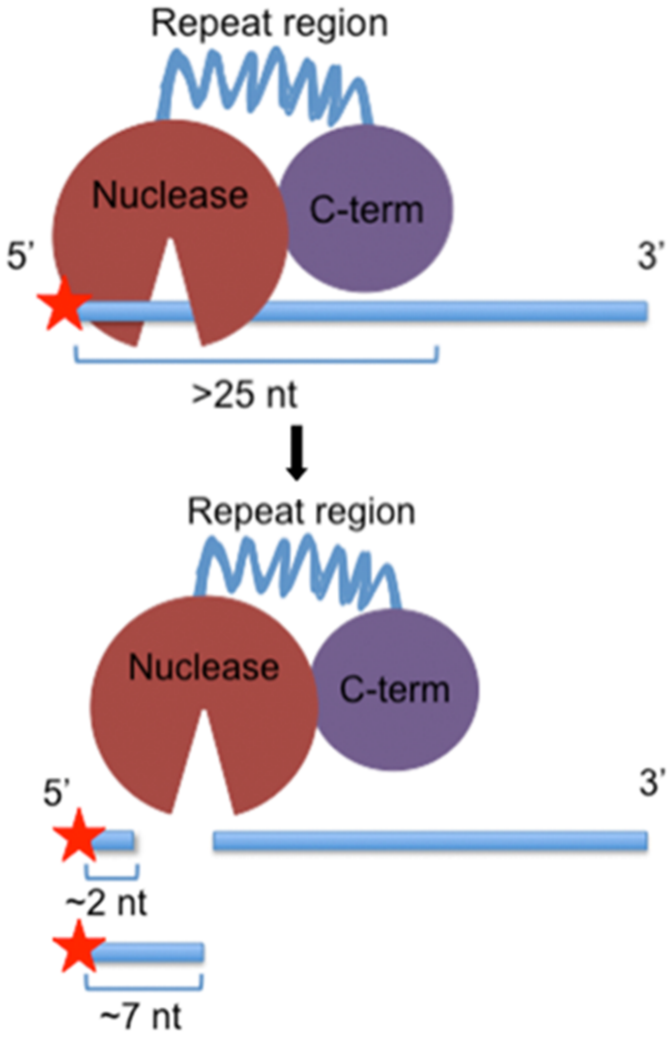


FIGURE 13. SAN1 acts as an exonuclease on 5' ssDNA structures. SAN1 binds ssDNA substrates greater than 25 nucleotide and cleaves from the 5' end. SAN1 requires a free 5' end of ssDNA for activity and after binding SAN1 cleaves 2-7 nucleotide from the 5' end of DNA. SAN1 is not a processive nuclease and falls off after cleaving but may be able to rebind additional substrates.

nucleotides and requires a free 5' end of DNA to bind (Figure 13). SAN1 cleaves 2 or 7 nucleotide fragments from the 5' end of ssDNA including splayed arm structures and 5' flaps. I have also shown that SAN1 utilizes Mg^{2+} as its metal cofactor, like other FEN1 family members, and may be able to fall off of DNA and rebind to initiate another cleavage event creating the laddering pattern of products seen after ssDNA substrates are 3' ^{32}P labeled (Figure 13).

To understand how the various domains of SAN1 contribute to substrate specificity or regulation of the nuclease activity, several deletion mutants were designed. The nuclease domain of SAN1 is homologous to other members of the FEN1 family. However, other family members do not have the same internal repeat region or C-terminus of SAN1 that are of unknown function. As the C-terminus of SAN1 is conserved amongst species that have SAN1, I predicted that this region of the protein may be important for regulating nuclease activity, substrate recognition or a potential protein-protein interaction. An N-terminal FLAG-tagged nuclease domain and a C-terminal Myc-tagged SAN1 C-terminus mutant were developed using Myc-RhoA as a negative control (Figure 14a). The N-terminal FLAG-tagged nuclease domain and the C-terminal Myc-tagged SAN1 C-terminus were co-expressed in 293T cells. The lysates were then immunoprecipitated with anti-FLAG M2 beads, and the results were analyzed by immunoblot (Figure 14b). The Myc Input membrane was re-exposed for a longer time to detect Myc-SAN1 C-terminus as it showed a very low expression level alone. This observation could be the result of the N-terminal and C-terminal interaction providing structural stability to each domain and thus when the C-terminus is expressed alone a significant portion of it is

degraded as the mutant is unstable alone. Interestingly, the SAN1 C-terminal domain but not Myc-RhoA is co-precipitated with the nuclease domain providing evidence that the SAN1 C-terminus and N-terminus interact by co-immunoprecipitation (Figure 14c).

After determining that the nuclease domain of SAN1 and the C-terminus interact by Co-IP, it was important to understand if this interaction was required for nuclease activity. To test this, an N-terminal/C-terminal fusion protein of SAN1 called ΔRepeats and an N-terminal/Repeat region fusion protein called ΔCterminus were produced in 293T cells and affinity purified (Figure 14d). The repeat region was hypothesized to function as a flexible linker to allow the N-terminus and C-terminus of SAN1 to interact for substrate recognition or activation of the nuclease domain. Each construct was overexpressed, immunoprecipitated with FLAG beads and analyzed by western blot to confirm that both deletion mutants were expressed and eluted efficiently (Figure 14e). The purified deletion mutants were then tested for nuclease activity against 5' ^{32}P labeled ssDNA using the filter spin assay described previously. We observed that the ΔRepeats mutant showed similar activity to WT SAN1 while the ΔCterminus mutant showed no nuclease activity similar to D90A SAN1 (Figure 14f).

Taken together, these results characterize a novel nuclease that shares some similarities with other FEN1 family members but also appears to have striking substrate and structure/function regulation differences. SAN1 acts as a 5' ssDNA exonuclease and requires a free 5' end of ssDNA to bind and cleave substrate. SAN1 will only cleave ssDNA substrates greater than 25 nucleotide but if the ssDNA is part of a larger structure,

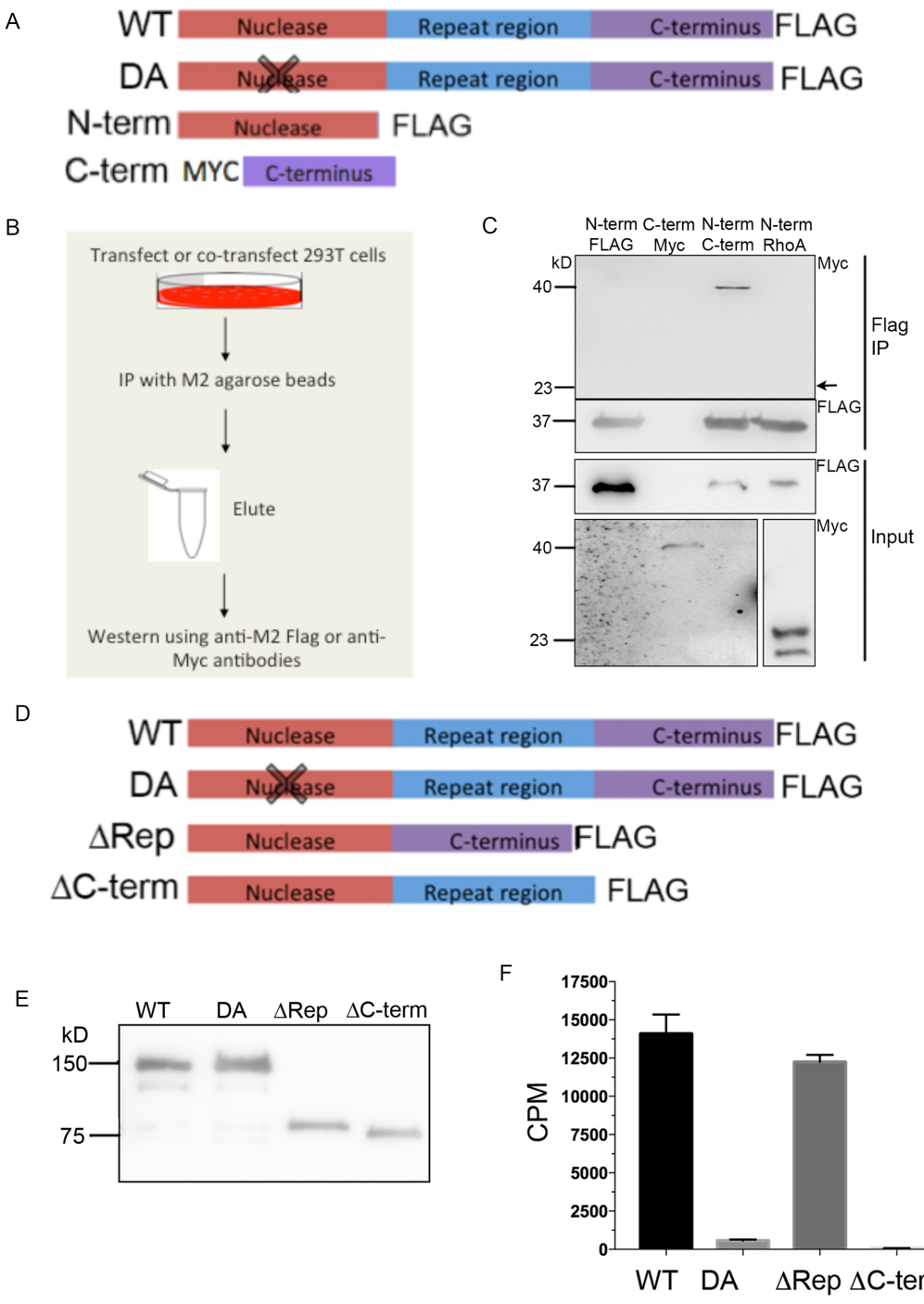


Figure 14. SAN1 C-terminus is required for nuclease activity. (a) Schematic of SAN1 deletion mutants used in b, c (b) Schematic of SAN1 deletion mutant expression and purification from 293 T cells (c) The N-terminal FLAG-tagged nuclease domain was co-expressed in 293T cells with C-terminal Myc-tagged SAN1 C-terminus, or Myc-RhoA as a negative control. Lysates were precipitated with anti-FLAG M2 beads and analyzed by

immunoblot. The Myc Input membrane was re-exposed for a longer time to detect Myc-SAN1 C-terminus. The SAN1 C-terminal domain but not RhoA is co-precipitated with the nuclease domain. (d) Schematic of SAN1 deletion mutants used in e, f (e) immunoblot of WT, DA, ΔRep and ΔC-term proteins purified from 293T cells (f) tested for nuclease activity against 5' ^{32}P labeled ssDNA using the filter spin assay (N=2, error bars show range)

for example a splayed arm or 5' flap structure, it will act on the free 5' end. After showing that the N-terminus and C-terminus interact by Co-IP and this interaction is required for nuclease activity, I propose a new model that combines all known aspects of the nuclease function of SAN1. SAN1 has a ssDNA length requirement of greater than 25 nucleotides because there is a portion of the C-terminus of SAN1 that interacts with the ssDNA and substrates shorter than 25 nucleotides are not readily able to interact with both regions of the protein. This would make ssDNA sequences shorter than 25 nucleotides undesirable substrates. As a feature of a larger splayed arm or 5' flap structure the ssDNA length requirement is less than 25 nucleotides because the C-terminus interacts with the bottom strand of DNA present in the flap or splayed arm structure. This model also provides an explanation for the increased cleavage efficiency of the splayed arm structure over the 5' flap structure. The C-terminus of SAN1 interacts with the bottom ssDNA portion of the splayed arm structure and thus is able to cleave the splayed arm substrate readily whereas in the 5' flap structure the bottom strand of DNA is double stranded and thus provides a less favorable interaction with the C-terminus of SAN1 (Figure 15). This model of SAN1 nuclease activity on ssDNA sequences, as compared to other multi-strand DNA structures, is in line with the defining characteristic of FEN1 family members as nucleases that recognize structural components. From ssDNA/dsDNA junctions to flaps to lesions that produce bubbles and other structures in the DNA, the FEN1 family of nucleases recognizes and binds to structural components of potential substrates. SAN1 therefore represents a previously uncharacterized member of the FEN1 family that increases biological significance of the nucleases in the group by adding to the repertoire of substrates that family members act on.

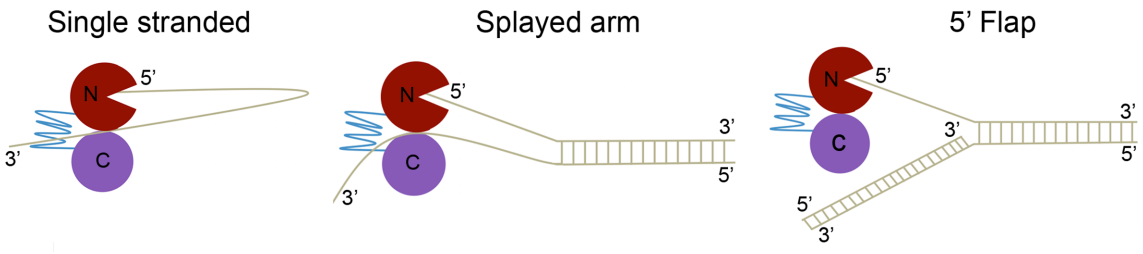


Figure 15. Schematic of a model for SAN1 nuclease activity on DNA structures. SAN1 acts on ssDNA substrates by recognizing the free 5’ end of DNA and cleaving ~3 or ~7 nts from the 5’ end, in a non-processive manner. SAN1 prefers splayed arm substrates over 5’ flap substrates because SAN1 is able to interact with the bottom strand of ssDNA in a splayed arm substrate whereas the bottom strand of DNA in a 5’ flap substrate is double stranded making it a less desirable substrate for the C-terminus of SAN1 to potentially interact with.

CHAPTER 3: SAN1 Involvement in ICL Repair

3.1 Abstract

SAN1 is the first nuclease found to be involved in ICL repair since FAN1 was identified in 2010. Studies show SAN1 is required for colony survival of cells treated with ICL agents MMC or cisplatin but not for survival of cells treated with ionizing radiation which generates single-stranded DNA breaks and double-stranded DNA breaks. SAN1$^{-/-}$ HeLa cells and SAN1$^{-/-}$ mouse embryonic fibroblasts (MEFs) show similar levels of sensitization to ICL agents and in both systems expression of WT SAN1 but not D90A SAN1 is able to rescue resistance to ICLs. As a follow up to the initial colony survival assay (CSA) data we also show that SAN1$^{-/-}$ cells show increased chromosomal aberrations and radials as a result of MMC treatment, a phenotype similar to that seen in FA patients. Although SAN1-/- cells treated with MMC show a FA phenotype by cytogenetic analysis, additional studies have shown that SAN1 is not epistatic to FANCD2, and thus likely does not function in the FA pathway. Identification of Senataxin as a SAN1 interacting protein along with SAN1 being epistatic to SNM1a, a known nuclease involved in transcription-coupled ICL repair provide evidence for the involvement of SAN1 in a replication-independent ICL repair pathway.

3.2 Introduction

DNA interstrand crosslinks (ICLs) are a toxic form of DNA damage that disrupts transcription and replication by covalently and irreversibly joining complementary DNA strands[103]. ICL agents all cause varying levels of distortion in the DNA helix upon binding to the DNA and thus represent a fairly diverse group of damaging agents[60]. Cisplatin causes highly distorting lesions while MMC causes less distortion in the DNA

helix[104]. Various ICL agents also cause different numbers of ICLs and intrastrand crosslinks[105]. Mono-ubiquitylation of the FANCI-FANCD2 (ID2) heterodimer leads to recruitment of multiple nucleases that control nucleolytic incision and ICL unhooking[106], including the endonucleases XPF (FANCQ), which forms an XPF-ERCC1 heterodimer, and SLX1, with additional nucleases such as FAN1, SNM1A and MUS81 contributing independently of the FA pathway[64,66,72]. XPF-ERCC1, which is involved in nucleotide excision repair, is recruited to perform the unhooking incisions[64,72], but under some circumstances it only performs the 3' incision, and another nuclease would be responsible for the 5' incision. The identity of this nuclease remains ambiguous but SLX1 is one candidate[11]. FAN1, a nuclease that interacts with FANCD2, can digest recessed 5' DNA ends and cleave 4 nucleotides 3' to an ICL[56] but is not required for unhooking in *Xenopus* extracts and its function in ICL repair remains unclear[59]. Another nuclease, SNM1A, has no known function in incision, but may participate in repair by digesting past the ICL[72]. It remains unclear whether a single nuclease is responsible for the 5' incision, or if several nucleases act redundantly to complete this process.

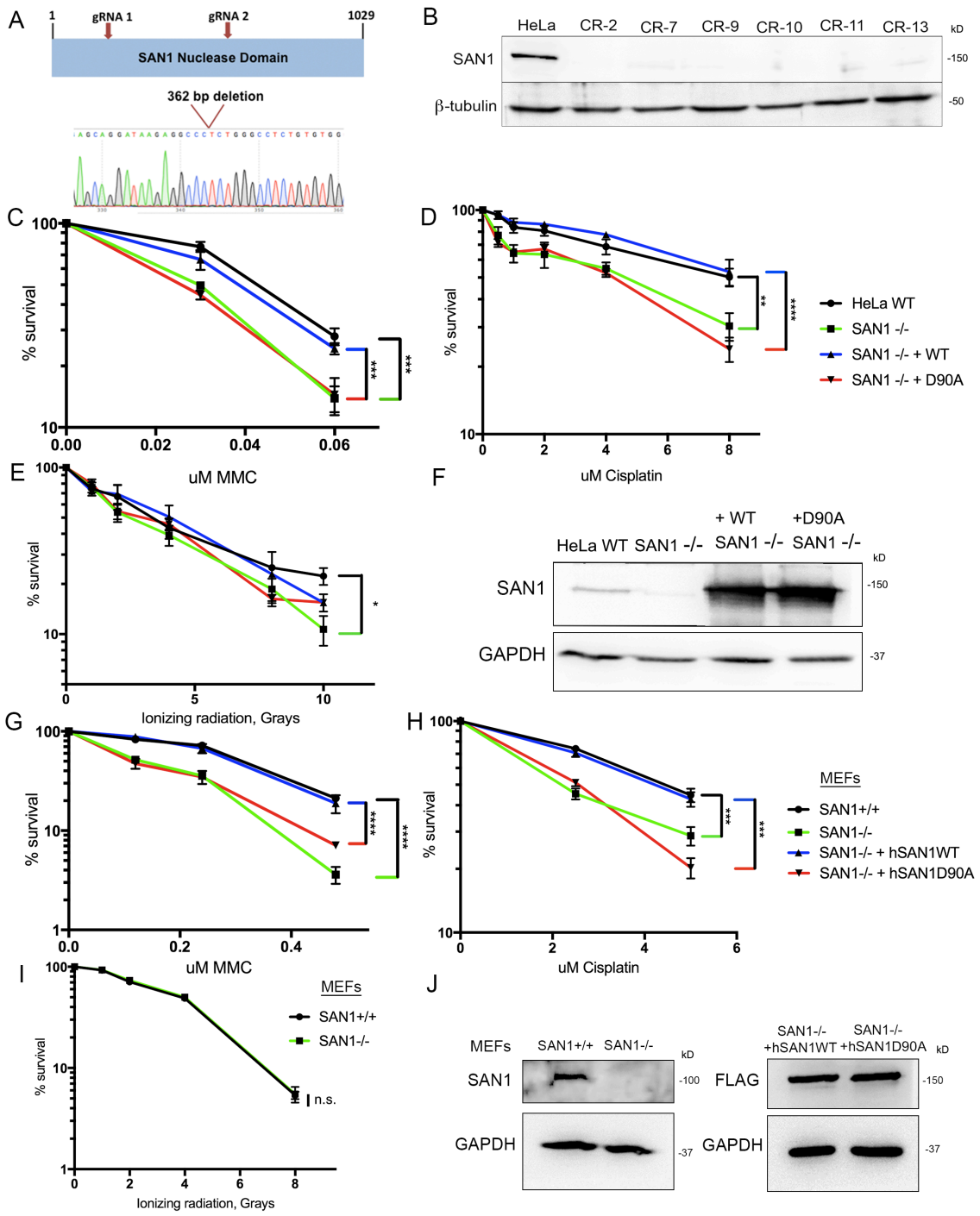
ICL repair can also be triggered by stalling of transcription complexes at lesions during other periods of the cell cycle, including G1[27]. One consequence of transcriptional stalling is the formation of R-loops, which consist of a RNA-DNA hybrid plus the looped single-stranded coding strand of the DNA[107]. R-loops form naturally during transcription at promoters of genes with a high GC content and at termination regions of genes[107]. Persistent R-loops can impede replication and be processed into double stranded breaks (DSBs), leading to genomic instability. ICLs between RNA and DNA strands might also

occur at these structures, although to date there is no direct evidence for their existence. R-loops can be resolved by an endonuclease, RNase H, or by an RNA/DNA helicase, Senataxin (SETX), and if they persist can be aberrantly processed into DSBs by the NER endonucleases XPF and XPG[108]. Interestingly, R-loop resolution has recently been linked to the FA pathway[107,109], and to BRCA1[110], a protein essential for HR and resistance to ICLs.

3.3 Results and Discussion

After identifying SAN1 to be a new and unique 5' ssDNA exonuclease, I also wanted to understand if SAN1 acts as a nuclease *in vivo* and, if so, in what type of DNA repair pathway SAN1 could be functioning in. Data showed that when SAN1 is depleted in cells by shRNA and cells are treated with different DNA damaging agents (MMC, cisplatin, IR, MMS) SAN1 depleted cells are sensitized to only the ICL agents MMC and cisplatin (data not shown). IR causes single- and double-stranded breaks that are commonly repaired by HR or NHEJ pathways, while MMS is an alkylating agent that is repaired by BER.

I generated a SAN1 knockout cell line in HeLa cells using CRISPR/Cas9 genome editing technology to follow up on this result. To abolish nuclease activity of SAN1, two guide RNAs were used that would facilitate deletion of almost 400 base pairs from the nuclease domain of SAN1 (Figure 16a). Clones were screened by PCR, sequencing, and western blot to determine which clones were potential SAN1 knockouts and to validate that they were in fact not producing detectable levels of SAN1 (Figures 16a, b).



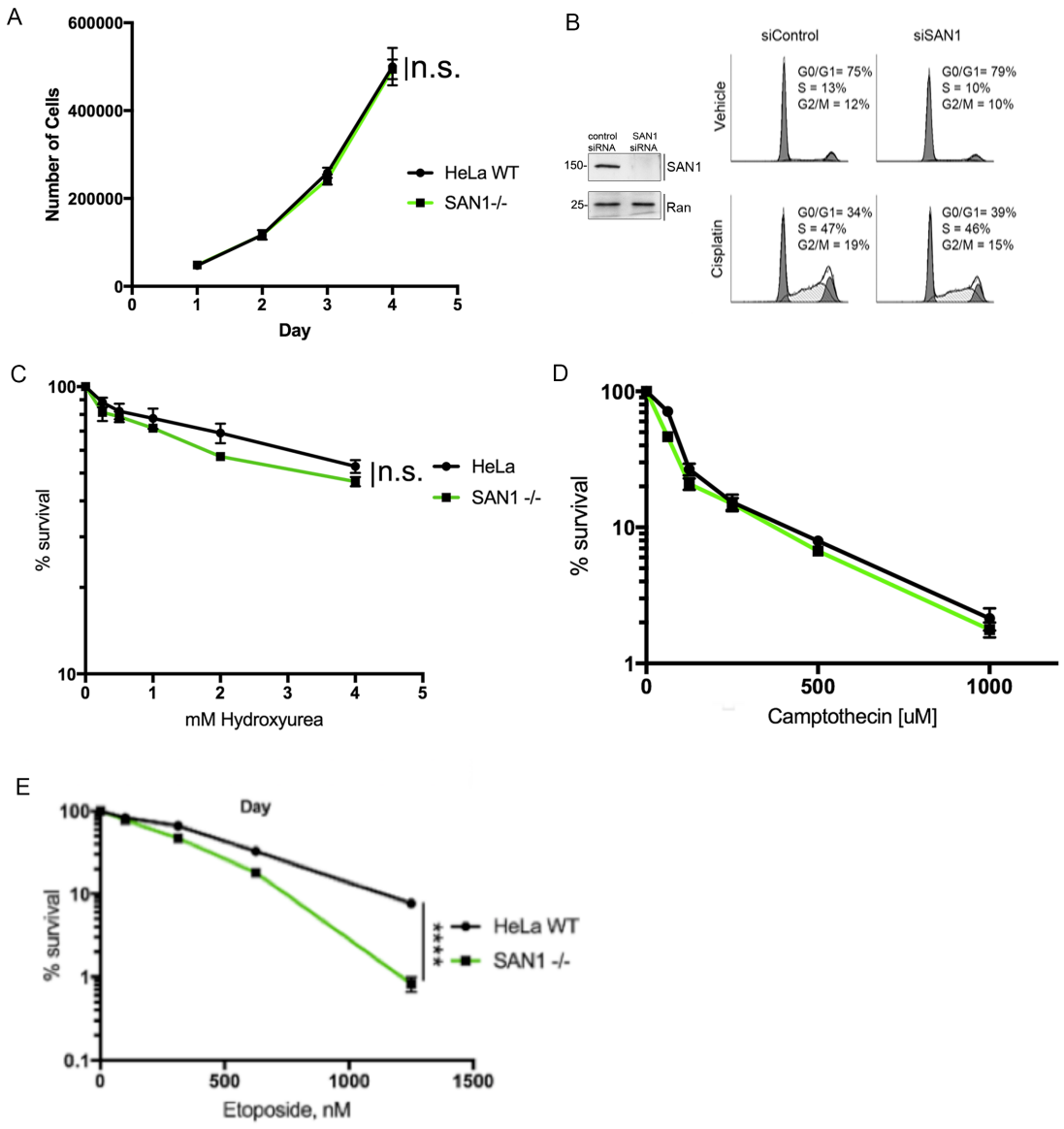
Alex Andrews – CSAs and generation of MEF cell line

FIGURE 16. Loss of SAN1 leads to sensitization of cells to ICL agents (a) Schematic showing CRISPR-Cas9 strategy to create SAN1$^{-/-}$ HeLa cell lines. Two guide RNAs were used to delete a 362 bp region of exon 1 in the *fam120b* gene locus, which contains the conserved FEN1 family nuclease domain. (b) Immunoblot of HeLa WT parental cell line and CRISPR-Cas9 generated SAN1$^{-/-}$ cell lines showing loss of SAN1 expression (lanes

1-7). β-tubulin was used as a loading control. (c-e) SAN1$^{-/-}$ cells were transduced with lentiviral constructs expressing Strep$_2$-FLAG tagged SAN1 WT or the D90A mutant to create stable rescue cell lines**.** Colony survival assays (CSAs) were then performed using HeLa WT, SAN1$^{-/-}$, and WT or D90A rescue lines with MMC, Cisplatin, or ionizing radiation ($N \geq 3$). Statistical significance determined by two-way ANOVA comparing HeLa WT and SAN1$^{-/-}$ or SAN1$^{-/-}$ +WT and SAN1$^{-/-}$ +D90A. Error bars denote s.e.m. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, **** = $p<0.0001$. (f) Immunoblot showing SAN1 expression in HeLa WT, SAN1$^{-/-}$, and SAN1 WT and D90A rescue lines. GAPDH was a loading control**.** (g-i) SAN1$^{+/-}$ mice were crossed to generate SAN1$^{+/+}$ and $^{-/-}$ mouse embryonic fibroblasts (MEFs). The MEFs were immortalized using SV40 large T antigen, and the SAN1$^{-/-}$ MEFs were transduced with lentiviral constructs containing Strep$_2$-FLAG-tagged human SAN1 WT or SAN1 D90A. These cell lines were then used for CSAs with MMC, Cisplatin, or ionizing radiation (N=3). (j) Immunoblot showing mouse SAN1 expression in SAN1 $^{+/+}$ and $^{-/-}$ MEFs (left panel) and hSAN1WT and hSAN1D90A expression in SAN1$^{-/-}$ cells (right panel). GAPDH was used as a loading control. Statistical significance for CSAs was determined by two-way ANOVA test comparing SAN1$^{+/+}$ and SAN1$^{-/-}$ or SAN1$^{-/-}$ + hSAN1WT and SAN1$^{-/-}$ + hSAN1D90A.

Stable cell lines expressing WT or D90A SAN1 were developed from the SAN1$^{-/-}$ HeLa cell line to determine if the nuclease function of SAN1 was required for protection against sensitivity to these DNA damaging agents (Figures 16c, d, e). By colony survival assay (CSA), SAN1$^{-/-}$ cells showed increased sensitivity compared to WT HeLa cells following treatment with MMC or cisplatin. While the SAN1$^{-/-}$ +WT cell line was able to rescue survival similar to HeLa WT levels, the SAN1$^{-/-}$ +D90A line is still sensitized to crosslinking agents similar to the SAN1-/- line alone. This result shows that SAN1 nuclease activity is required for protection against ICL sensitivity (Figures 16c, d). Also in agreement with the initial CSAs, SAN1$^{-/-}$ cells are not sensitized to damage resulting from IR which produces single-stranded DNA breaks and double-stranded DNA breaks (Figure 16e). Together, these results point to SAN1 nuclease activity playing a role in ICL-specific repair before the HR step downstream of ICL unhooking. The repair of ICLs requires the coordination of several repair pathways including NER, TLS, and HR. Further, these data also show that the nuclease function of SAN1 must be important in ICL repair specifically and not in a downstream process of ICL repair like HR. Figure 16f is a western blot showing relative expression levels of WT HeLa cells, SAN1$^{-/-}$ HeLa cells, SAN1$^{-/-}$ +WT cells, and SAN1$^{-/-}$ +D90A cells. The expression levels of WT and D90A SAN1 relative to endogenous SAN1 were increased in the SAN1$^{-/-}$ + WT and SAN1$^{-/-}$ + D90A stable lines, although we do not predict that this increased expression confounds the initial result. Western blots of SAN1-/- HeLa, SAN1-/- + WT, and SAN1-/- +D90A cell lines are shown in Figure 16j and Appendix I (Figure 23a, b).

As an independent method to determine the role of SAN1 in the response to DNA

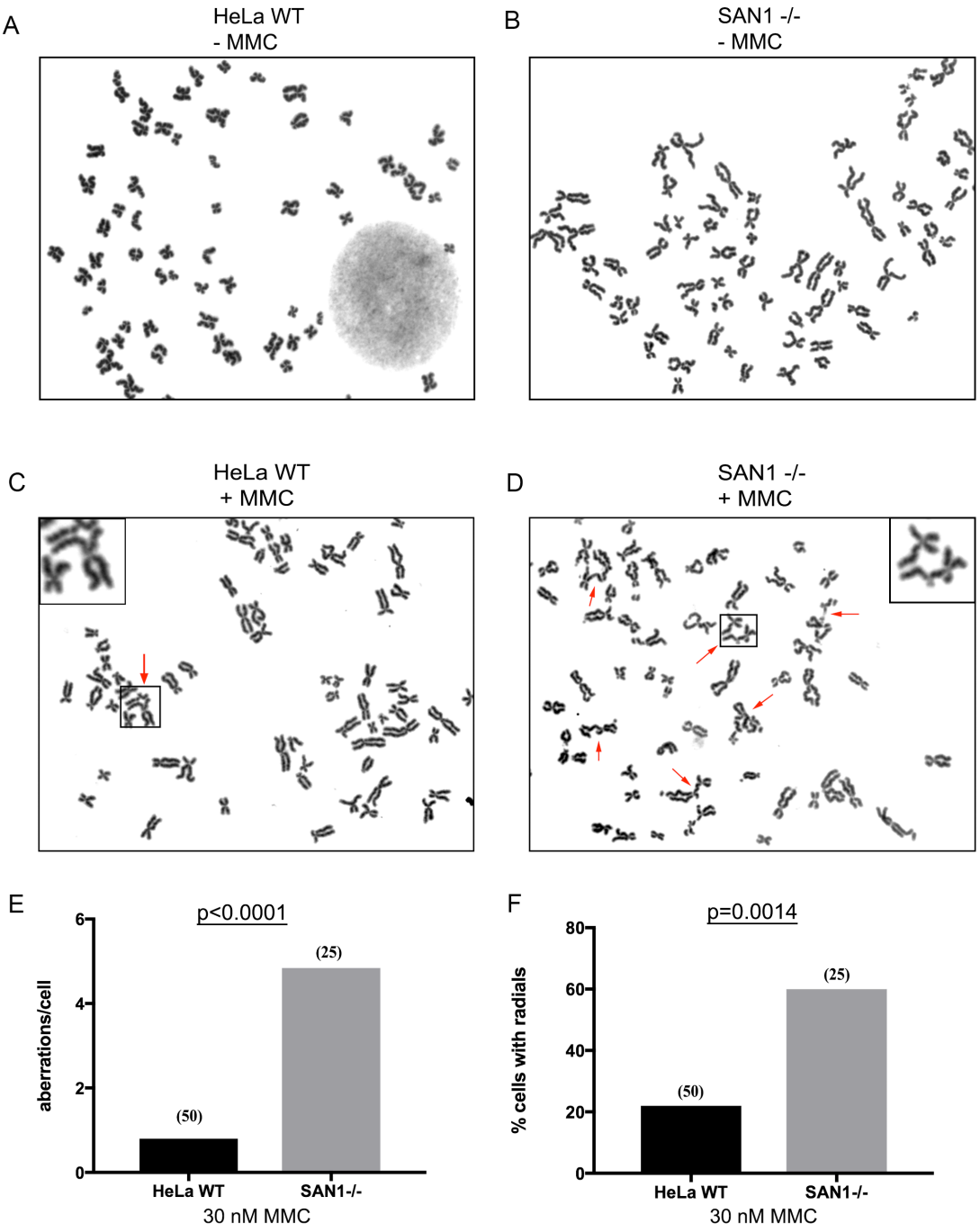


Tim Errington - panel B

FIGURE 17. SAN1$^{-/-}$ cells proliferate normally and are not sensitized to Hydroxyurea or camptothecin. (a) Proliferation assay of WT HeLa cells and SAN1$^{-/-}$ HeLa cells generated by CRISPR/Cas9 technology. Cells were plated in triplicate on day 0 and were trypsinized and counted each day for 4 days. Error bars represent SEM. (b) Cell cycle distribution of cells treated with SAN1 or control siRNA and vehicle or Cisplatin. Immunoblot (left panel) shows knockdown of SAN1 by siRNA. Ran was used as a loading control. Clonogenic survival assays of WT HeLa cells compared to SAN1$^{-/-}$ HeLa cells after treatment with (c) hydroxyurea, (d) camptothecin (e) etoposide. Statistical significance for CSAs was determined by two-way ANOVA test. Error bars denote s.e.m. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, **** = $p<0.0001$.
the cells.

damage, we created a conditional KO mouse, and generated mouse embryonic fibroblasts (MEFs) through the Texas A&M Institute for Genomic Medicine, using EUCOMM ES cells targeting the *fam120b* gene (ES cell clone HEPD0652_5_G10). A *fam120b*/+ male was crossed to a FLPer/+ female to delete the lacz/neo markers, the FLPer transgene was then removed by crossing to a +/+ mouse, and the resulting floxed allele mice were crossed to produce homozygotes. These homozygous mice were then crossed with a Sox2-Cre mouse to obtain a global knockout of the allele. From these $SAN1^{-/-}$ mice, which were viable and fertile, we isolated primary embryonic fibroblasts. Because primary fibroblasts do not proliferate when plated as single cells, they were immortalized using the SV40 the large T antigen, and lines were created from $SAN1^{-/-}$ and $SAN1^{+/-}$ and $WT^{+/+}$ littermates. These lines were then tested for sensitivity to DNA damaging agents. The homozygous knockout cells, $SAN1^{-/-}$ MEFs, showed significantly increased sensitivity to ICL agents as compared to WT cells (Figures 16g, h), and were not sensitive to IR (Figure 16i). Moreover, re-introduction of human SAN1-ssf (strep-strep-FLAG tag) WT rescued survival but the D90A mutant did not (Figures 16g, h, i). Taken together, these data demonstrate that SAN1 nuclease activity is required for survival in response to ICL agents, but not to drugs that induce single- or double-stranded DNA breaks, suggests the finding that SAN1 plays a role in ICL repair but does not participate in HR or non-homologous end joining (NHEJ). Raw western blots of SAN1-/- MEFs, SAN1-/- + hWT MEFs, and SAN1-/- + hD90A MEFs cell lines are shown in Appendix I (Figure 23c).

In addition, $SAN1^{-/-}$ cells proliferate similarly to WT HeLa cells and SAN1



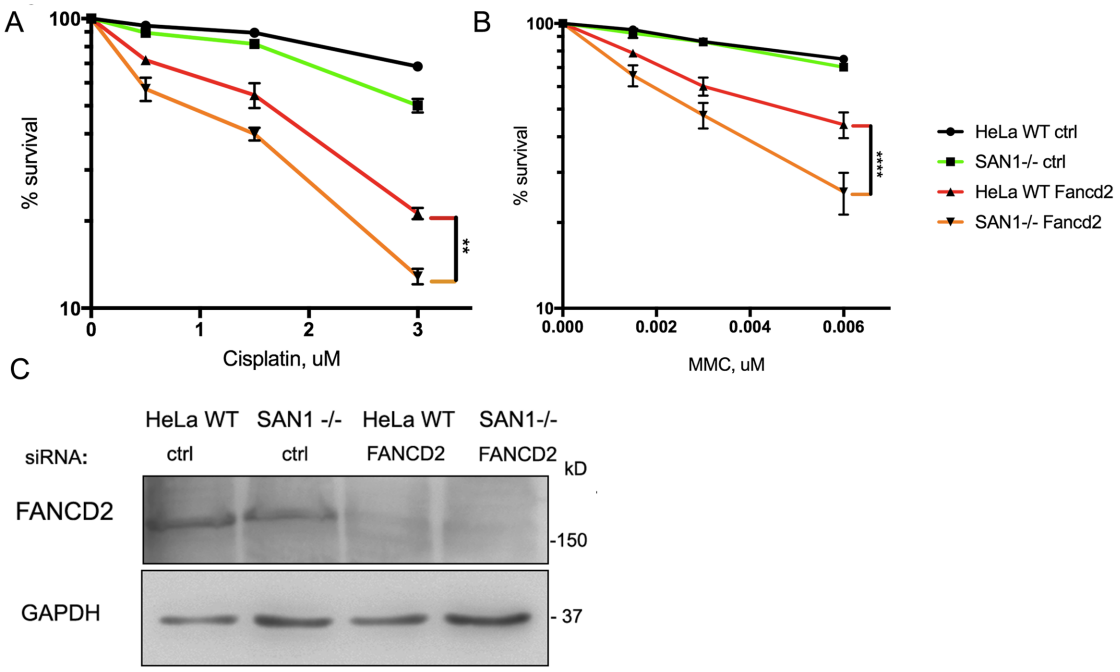
Alan D'Andrea Lab- Cytogenetic Analysis

FIGURE 18. SAN$^{-/-}$ cells display increased levels of radial chromosomes in response to MMC. (a, b) Micrographs of metaphase spreads from untreated HeLa WT and SAN1$^{-/-}$ cells. (c, d) Micrographs of metaphase spreads of HeLa WT and SAN1 $^{-/-}$ cells following treatment with 0.03 mM MMC, showing large increase in radials or other chromosomal aberrations in SAN1$^{-/-}$ cells. Red arrows indicate radial chromosomes or aberrations. (e,f)

Quantification of aberrations/cell and percentage of radials/cell for HeLa WT and SAN1 $^{-/-}$ cells treated with MMC. Metaphase spreads from 50 HeLa WT cells were analyzed (11 radial forms, 11 cells with radials, 40 aberrations). Metaphase spreads from 25 SAN1$^{-/-}$ cells were analyzed (42 radial forms, 15 cells with radials, 121 aberrations). Data were analyzed in Prism GraphPad from contingency tables using Fisher's exact-test (two-sided P value).

knockdown cells shown a normal cell cycle profile as compared to cells treated with control siRNA with and without cisplatin (Figures 17a, b). These results, along with the SAN1$^{-/-}$ HeLa CSA data and the SAN1$^{-/-}$ MEF CSA data, confirm that removal of SAN1 from cells does not alter normal proliferation or cell cycle progression but does strongly sensitize cells to ICL agents MMC and cisplatin.

In addition to testing the DNA damaging agents MMC, cisplatin, and IR, we wanted to determine if SAN1 knockout cells showed sensitivity to any other replication-dependent DNA damaging agents. Using WT HeLa cells and SAN1$^{-/-}$ HeLa cells CSAs were performed with hydroxyurea, camptothecin, and etoposide. Hydroxyurea is a potent inhibitor of the enzyme ribonucleotide reductase that prevents replication by depleting the pool of dNTPs available at the replication fork for DNA synthesis. Camptothecin is a topoisomerase I poison that blocks disassociation of the topoisomerase I from DNA causing single-stranded DNA breaks. Etoposide, on the other hand is a topoisomerase II poison that blocks disassociation of the topoisomerase II from DNA, resulting in persistent DNA double-stranded breaks that can cause chromosomal aberrations or rearrangements. Each of these drugs causes significant replication stress. SAN1 deletion had little effect on sensitivity to hydroxyurea and camptothecin but did show a slight sensitivity to etoposide (Figures 17c, d, e). FA proteins are known to have roles in other pathways including replication fork protection and even cytokinesis[11]. Although we predict that SAN1 may have an additional function in an etoposide-related repair pathway, these results broadly confirm that SAN1 knockout cells show the most significant sensitivity to ICL-inducing agents including MMC and cisplatin. Additional

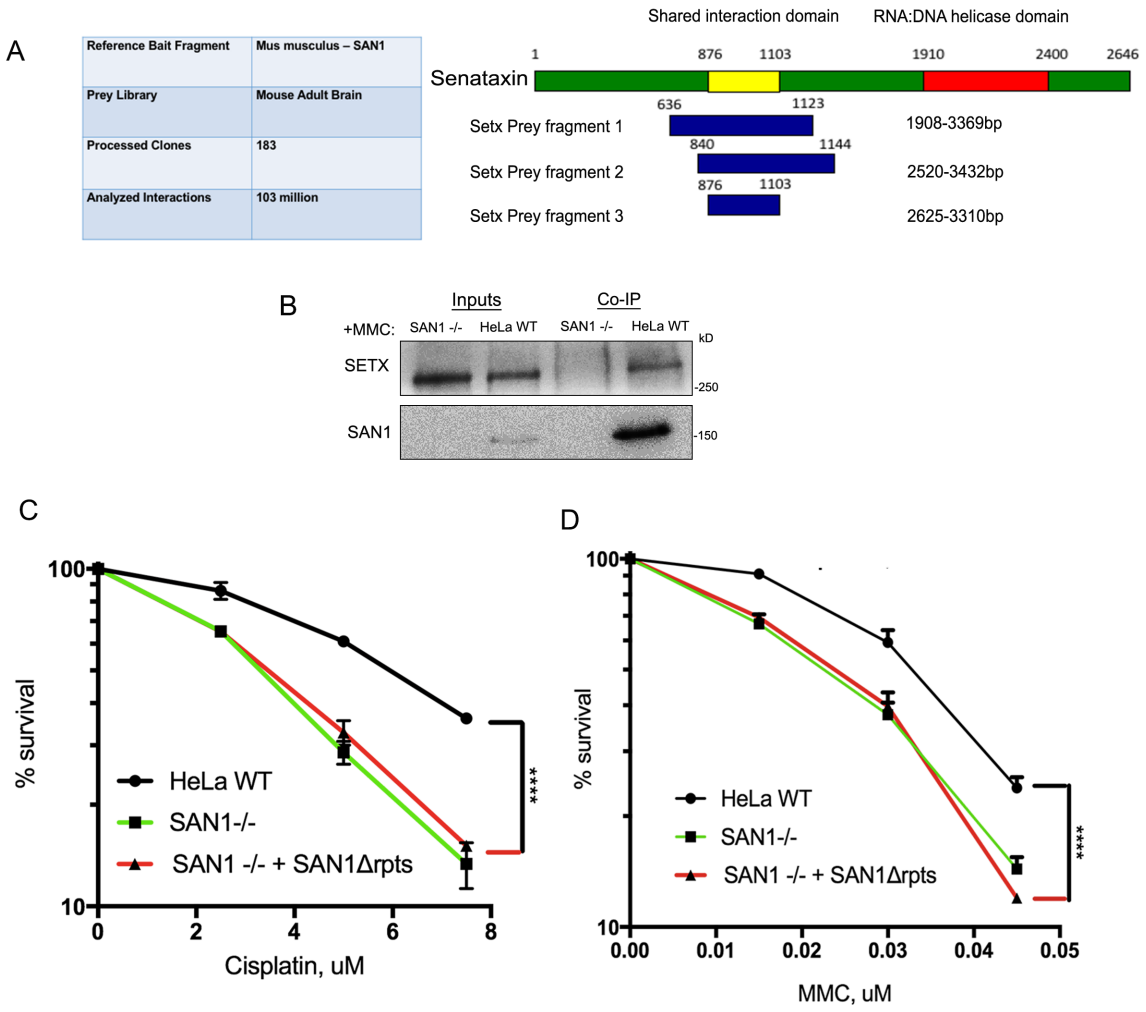


Alex Andrews – Figure 19

FIGURE 19. SAN1 functions independently of the FA pathway and does not affect FA pathway activation. (a. b) CSAs of HeLa WT and SAN1 $^{-/-}$ cells treated with scrambled ctrl siRNA or FANCD2 siRNA, in response to Cisplatin and MMC (N=3). Statistical significance determined by two-way ANOVA. Error bars denote s.e.m. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, **** = $p<0.0001$. (c) Immunoblot showing siRNA knockdown of FANCD2 in HeLa WT and SAN1$^{-/-}$ cells.

investigation is required to determine if SAN1 plays a minor role in another pathway.

Because SAN1$^{-/-}$ cells show an increased sensitivity to crosslinking agents such as MMC and cisplatin in colony survival assays, it was critical to determine if the SAN1$^{-/-}$ HeLa line showed a similar phenotype to FA patients by cytogenetic analysis. The results from the cytogenetic analysis are shown in Figure 18. Untreated SAN1$^{-/-}$ cells show few chromosomal breaks or radials but treatment with 0.03 mM MMC significantly increased the number of breaks and radials relative to WT HeLa treated cells. As our SAN1$^{-/-}$ cell line showed a similar phenotype to FA patient cells (who display severe sensitivity to ICL-inducing agents) the results pointed to a role for SAN1 in FA-mediated ICL repair. This assay is the gold standard for diagnosing patients with FA. In the clinic, patient blood samples are taken and treated with ICL-inducing agents MMC, cisplatin, or diepoxybutane (DEB) before being scored for chromosomal breaks and radials. These breaks and radials result from the inability to repair ICLs, resulting in double-stranded DNA breaks and the aberrant activation of NHEJ pathway, which is highly error prone. Chromosomes can be mistakenly repaired to other chromosomes resulting in radial structures, or fragments can be lost entirely resulting in cell death. It is also important to note that in the assay with the parental WT HeLa cell line, WT HeLa cells showed minimal chromosomal breaks or aberrations without DNA damaging agents and significantly fewer breaks or radials when treated compared to the SAN1$^{-/-}$ HeLa line treated (Figures 18a-d). The SAN1$^{-/-}$ HeLa line treated with MMC showed five-fold greater aberrations per cell and three-fold greater cells with radials than treated WT HeLa



Alex Andrews – Panels B, C, D

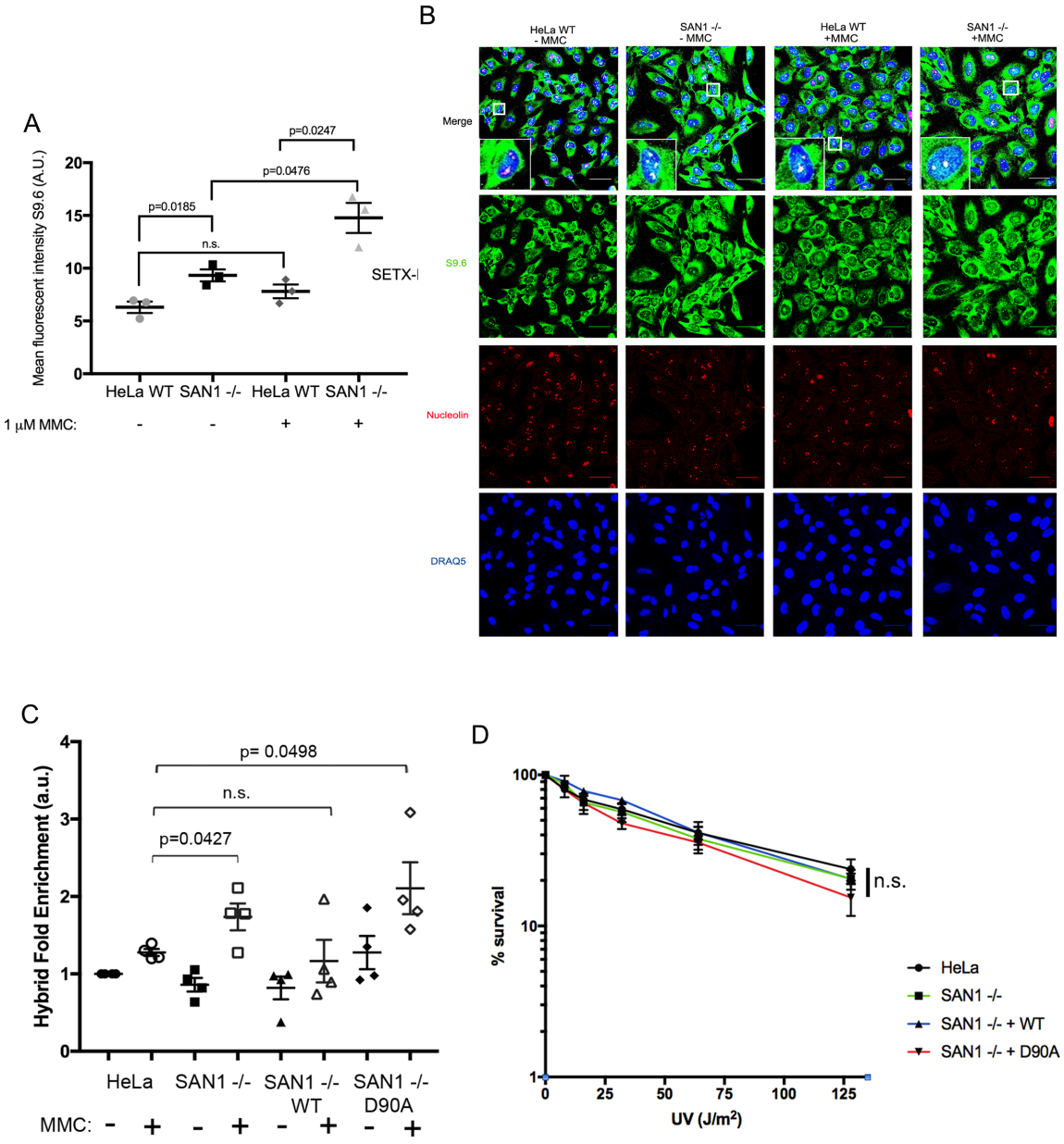
FIGURE 20. SAN1 interacts with Senataxin and interaction is required for ICL resistance. (a) Yeast two-hybrid (Y2H) screen and library information in addition to a schematic detailing mSETX prey fragments that interacted with mSAN1 in Y2H screen. The shared interaction domain of mSETX prey fragments shows that SAN1 interacts with SETX in a region of the N-terminus and not near the RNA/DNA helicase domain at the C-terminus. (b) Endogenous SAN1 was co-immunoprecipitated from HeLa WT and SAN1 $^{-/-}$ after treatment with 1 mM MMC. Top panel: immunoblot (IB) of Senataxin inputs (lanes 1-2) and Co-IP (lanes 3-4), bottom panel: IB of SAN1 inputs (2%) (lanes 1-2), and Co-IP (lanes 3-4). (c-d) CSAs for HeLa WT, SAN1$^{-/-}$, and SAN1$^{-/-}$ +SAN1ΔRep-

ssf cells exposed to Cisplatin and MMC. Statistical significance determined by two-way ANOVA.

cells which represent very significant increases in response to these ICL agents (Figures 18e, f).

Next, epistasis experiments were used to determine whether SAN1 functions as a nuclease within the FA pathway, the canonical mechanism for ICL repair. FANCD2, a central component of the FA pathway was depleted in WT HeLa cells or in the SAN1-/- HeLa background. FANCD2 plays a critical role in the activation of the FA pathway, and acts upstream of proteins that function in the nucleolytic steps of ICL repair such as SLX1-SLX4, XPF-ERCC1, and FAN1[9,26,27]. Treatment of WT HeLa cells with FANCD2 siRNAs led to a sharp decrease in cell survival when cells were exposed to low doses of Cisplatin or MMC, as previously reported (Figures 19a-c)[28]. Strikingly, when FANCD2 was depleted in $SAN1^{-/-}$ cells, a synergistic increase in the sensitivity of $SAN1^{-/-}$ cells to ICLs occurred. In these experiments, we observed only a small to negligible decrease in survival between HeLa WT and $SAN1^{-/-}$ cells at very low concentrations of Cisplatin and MMC. The synergistic increase in sensitivity after depletion of FANCD2 in $SAN1^{-/-}$ cells suggests that ICLs are predominantly processed by the FA pathway and that SAN1 might function in a secondary pathway if the FA pathway is overwhelmed by abundant ICLs. Collectively, our data argue that SAN1 is not epistatic to FANCD2, and functions independently of the FA pathway (Figures 19a-c).

To probe the function of SAN1, we sought potential interacting proteins using a Hybrigenics genome-wide yeast two-hybrid (Y2H) screen. Full length murine SAN1 was used as bait, and was screened against a mouse adult brain library with a complexity of



Alex Andrews- Panels A, B

FIGURE 21. SAN1$^{-/-}$ cells show increased R-loop formation with MMC treatment. (a) Quantification of nuclear R-loop intensity (N=3). HeLa WT and SAN1$^{-/-}$ cells were treated with vehicle or 1 mM MMC and labeled with a monoclonal antibody to detect RNA/DNA hybrids (S9.6), nucleolin, and Draq5. Statistical significance calculated using unpaired t-test (N=3 biological replicates, at least 60 cells per sample were analyzed). (b)

HeLa WT and SAN1$^{-/-}$ cells were treated with vehicle or 1 mM MMC and labeled with a monoclonal antibody to detect R-loops (S9.6), a polyclonal antibody against nucleolin, and Draq5 to label DNA. Intensity of nuclear R-loop staining was quantified from the nucleus following masking with the DRAQ5 channel, and subtraction from nucleolin regions. (c) Dot blot assay for quantification of RNA/DNA hybrids. (N=4) Statistical significance determined by unpaired t-test comparing each condition to HeLa treated. (d) Colony survival assay of HeLa WT, SAN1$^{-/-}$, SAN1$^{-/-}$ +WT and SAN1$^{-/-}$ +D90A cell lines following exposure to UV.

~1x10^7, at 10-fold coverage. From 103 million analyzed reactions, 183 clones were processed, and 9 positive clones were in-frame and were present more than once (Figure 20a, Appendix I - Figure 24). Amongst these high-confidence hits was Senataxin (SETX), a gene mutated in ataxia oculomotor apraxia type 2 (AOA2) and juvenile amyotrophic lateral sclerosis (ALS4)[33,34]. SETX has been shown to participate in several different DNA damage responses, and functions centrally as a RNA/DNA helicase in the resolution of R-loops[35]. A small region of this protein, mSETX (876-1103), N-terminal to the helicase domain, bound to SAN1 (Figure 20a). Based on the role of Senataxin (SETX) in various areas of the DNA damage response, we explored whether an interaction between SAN1 and SETX is important for SAN1 function. Immunoprecipitation of endogenous SAN1 co-precipitated a protein detected by an anti-SETX antibody from HeLa WT cell nuclear extracts, but not from SAN1$^{-/-}$cells (Figure 20b). Moreover, unlike SAN1WT-ssf, SAN1ΔRep-ssf was unable to interact with full length SETX-Flag-GFP (data not shown) and was also still active as a nuclease (Figure 14f). We next used this mutant to test if SAN1 binding to SETX is essential for ICL resistance. Strikingly, the SAN1ΔRep-ssf mutant was unable to rescue survival in SAN1$^{-/-}$ cells treated with Cisplatin or MMC (Figures 20c, d).

Finally, given the known role of SETX in resolving R-loops, we asked whether the loss of SAN1 might lead indirectly to R-loop accumulation, from a failure to repair and remove ICLs. To test this possibility, we performed immunofluorescence staining of HeLa WT and SAN1$^{-/-}$ cells using the S9.6 monoclonal antibody, which specifically detects RNA-DNA hybrids. Co-staining with nucleolin antibody was used to subtract the

constitutive presence of RNA-DNA hybrids in the nucleoli. No significant difference in R-loops was found between untreated and MMC treated HeLa WT cells. However, the abundance of R-loops in SAN1$^{-/-}$ cells after MMC treatment was significantly increased (Figures 21a, b). To provide an independent assessment of R-loop formation, a dot blot assay was used in which isolated genomic DNA is probed for RNA/DNA hybrids with the S9.6 antibody, and normalized to total ssDNA (Figure 21c). Treatment of cells with MMC caused a significant increase in R-loops in SAN1$^{-/-}$ cells compared to WT cells. Moreover, the re-expression of WT SAN1, but not of the nuclease-defective D90A mutant, blocked any increase in R-loop formation in response to MMC (Figure 21c). I did not observe a similar increase between SAN1$^{-/-}$ cells compared to HeLa WT cells in untreated conditions as in the S9.6 staining in Figures 21a, b; however this might reflect differences in assay sensitivity or dynamic range. Representative slot blots shown in Appendix I (Figure 23d).

Chapter 4 – Discussion and Future Directions

4.1 Discussion

Every day, cells in the body are exposed to many DNA damaging agents. DNA damaging agents create tens of thousands of DNA lesions that when left unrepaired result in genomic instability and ultimately diseases such as premature aging and cancer[111]. To repair these lesions in a timely manner and to prevent genomic stress, an extensive network of repair pathways have evolved, collectively called the DNA damage response[4]. Although a vast array of protein components of many of the DNA damage response pathways have been identified, there remains a lack of understanding of how the network of DDR pathways interact to repair certain types of lesions throughout the cell cycle. One particularly complicated lesion to repair is the interstrand crosslink (ICL). ICLs are extremely cytotoxic, as they covalently link the strands of DNA resulting in a block of normal cellular processes such as replication and transcription[56]. Repair of ICLs also requires the interaction of several different repair pathways including nucleotide excision repair (NER), homologous recombination (HR), and translesion synthesis (TLS)[71]. Mammals have also developed an additional pathway to repair ICLs called the FA pathway which is critical in replication-dependent ICL repair[63]. The FA pathway is an extensive network of over 20 proteins, deficiency in any one of which leads to FA development[77]. The FA pathway and replication-dependent ICL repair are only active in S and G2 phases of the cell cycle, leaving cells susceptible to ICL damage outside of replication that must also be repaired. There is significantly less known about replication-independent ICL repair, thus representing an area of the field requiring additional research focus to further understand. This study characterizes a new nuclease component

of replication-independent ICL repair that acts on various ssDNA structures and is required for protection against ICL sensitivity.

Senataxin-associated nuclease 1 has been shown to display a unique nuclease activity on 5' ssDNA. SAN1 shares sequence homology with another known nuclease family, the FEN1 family of structure specific nucleases that function in various aspects of DNA replication and repair. Importantly, SAN1 has the same N and I regions of the nuclease domain that represent the regions of DNA binding and Mg^{2+} binding within the active site. SAN1 requires a free 5' end of DNA for binding proving it to be a true exonuclease, and prefers Mg^{2+} as a metal cofactor for nuclease reactions, similar to other members of the family. The FEN1 family represents a group of nucleases with a wide range of substrate preferences. FEN1, EXO1, GEN1, and XPG all cleave very different substrates but have homologous nuclease domains providing an interesting look into how these proteins employ different structural motifs to probe the nucleic acid environment surrounding potential substrates[78]. These structural regions, or pockets, on the enzymes increase substrate recognition and binding affinity. Studies examining the structures of FEN1 and EXO1 bound to DNA substrates have also identified other important features of the FEN1 family, including a K+-binding site, a hydrophobic wedge, and a helical gateway formed by alpha helices[83,87,88].

In addition to SAN1 possessing the same catalytic residues and structural components as other members of the FEN1 family, SAN1 also has a unique C-terminus and internal repeat region. The C-terminus interacts with the N-terminus and the C-terminus is

required for nuclease activity. As the regulation of nucleases is critical in preserving genomic integrity, the C-terminus of SAN1 could function as a regulator of SAN1 nuclease activity, thus preventing aberrant cleavage. SAN1 also possesses a unique repeat region, which we hypothesize allows for flexibility for the interaction between the N-termini and C-termini to stimulate substrate binding or nuclease function. Interestingly, although this repeat motif has not been found in other proteins, the SAN1 repeat motif has been shown to possess different numbers of repeats across species that have SAN1. For example, humans typically have twelve repeat motifs and mice typically have six repeat motifs. The internal repeat region is also not required for nuclease activity. Future studies examining if the internal repeat region functions as a flexible linker between the N- and C-terminal regions are necessary to more fully understand the contributions of various domains of SAN1. One possible avenue would be to design an alternative flexible linker and create a fusion protein of the N-terminus of SAN1 and the C-terminus of SAN1 connected by this alternative linker. After expressing and purifying this mutant protein, nuclease activity could be examined. If the nuclease domain and C-terminus are still able to interact through the alternative flexible linker then the mutant protein should show nuclease activity on a single-stranded DNA substrate. In addition to possibly functioning as a flexible linker, we know from the mapping experiments that the internal repeat region is the site of SAN1 interaction with Senataxin, an RNA/DNA helicase. We have also shown that SAN1ΔRepeats expressed in SAN1-/- cells does not restore protection against ICL sensitivity, providing evidence that the interaction with Senataxin is potentially important in the relationship between SAN1 nuclease activity and ICL repair.

FEN1 family members are thought to recognize substrates using either a "threading" or "tracking" mechanism. In the threading model, the nuclease recognizes the free 5' end of DNA and threads down to the ssDNA/dsDNA junction. For example, EXO1 cleaves flap structures via this threading mechanism, albeit inefficiently. EXO1 is able to recognize the free 5' flap and thread down to the ssDNA/dsDNA junction allowing for cleavage 1 nucleotide in from the ssDNA/dsDNA junction[78,86]. Conversely, in the tracking model the nuclease recognizes the ssDNA/dsDNA junction alone and is able to bind and cleave without recognizing the ssDNA end alone. FEN1, GEN1 and XPG can all recognize substrates using the tracking mechanism, recognizing the ssDNA/dsDNA junction and acting to resolve those structures (in a flap, Holliday Junction, or bubble) then proceeding with DNA replication or repair[80]. Here, I provide evidence based on the requirement of a free 5' end of DNA that SAN1 utilizes a threading mechanism to bind substrate and initiate cleavage. In addition to threading the ssDNA flap, the requirement of the C-terminus for nuclease activity and the increased cleavage capability for a splayed arm structure provide evidence for a model in which the C-terminus interacts with the bottom strand of ssDNA to increase nuclease activity.

SAN1 shows increased cleavage capability on splayed arm structures over ssDNA or 5' flap structures. These data support a model in which increased cleavage of splayed arm structures together with the binding requirement of the C-terminus for nuclease activity could account for increased substrate affinity or selection similar to other FEN1 family members. The interaction of the C-terminal region with the N-terminal region may

stimulate substrate binding or cleavage activity on splayed substrates more efficiently than on ssDNA substrates (that have a size requirement) or 5' flap structures (that have dsDNA on the bottom strand). Additional experimentation is required to understand the detailed mechanism of how, beyond the requirement of a free 5' end of DNA, SAN1 recognizes and binds to various substrates. Studies using EMSA assays to probe and analyze the binding capabilities of SAN1 D90A to various substrates would shed light on some of these possible mechanisms. Increasing concentrations of SAN1 D90A should lead to increased binding of substrates and a quantitative analysis of binding would identify substrates that are more efficiently bound by SAN1 D90A at various concentrations of enzyme.

In nuclease assays, SAN1 cleaves splayed arm structures the most efficiently, supporting a model in which SAN1 acts similarly to other FEN1 family members and recognizes or interacts with additional regions of the DNA structure for increased substrate specificity. The preference for cleaving the 5' flap of a splayed arm structure indicates that SAN1 acts on an intermediate 5' ssDNA flap that is part of a splayed arm structure after resolution of an R-loop structure and the action of another endonuclease such as XPF. SAN1 may then act on an intermediate structure after the initial cleavage event, processing the substrate for further repair. This would open the possibility for a nuclease like SNM1a to act downstream of SAN1 to digest past the ICL allowing for unhooking and subsequent repair steps. Persistent R-loops can result in double-stranded break formation and other genomic stress[109,112]. SAN1$^{-/-}$ cells have increased R-loop

formation following MMC treatment and the inability to repair ICLs in the absence of SAN1 could account for the observed persistent R-loops.

In addition to understanding the structure and function of SAN1, I also investigated if, and how, SAN1 acted as a nuclease *in vivo*. Disruption of the gene for the SAN1 nuclease, *Fam120b*, had no effect on cell cycle, but specifically increased sensitivity to cross-linking agents, and resulted in DNA damage and radial chromosome formation in the presence of Mitomycin C (MMC). Global analysis of gene expression by RNA-Sequencing failed to detect any significant differences either in expression levels or splice variants between MEFs from WT and Fam120b deleted mice, indicating that this protein does not significantly impact gene expression. Moreover, Fam120b knockout mice are viable and fertile.

SAN1 is not epistatic with components of the FA pathway, including FancD2 and XPF. However, the interaction with Senataxin is required for SAN1 function. SAN1 cooperates with SETX in an FA-independent repair process to protect cells from ICLs. Interestingly, the nuclease activity of SAN1 is the most similar to FAN1, a FANCD2 interacting nuclease that was identified in 2010[65,67]. Upon its discovery and connection to FANCD2 and the FA pathway, FAN1 was predicted to be the definitive 5' nuclease to act in the unhooking step of ICL repair. Although FAN1 is able to unhook the ICL, it is not required for ICL resolution and thus XPF is the only nuclease that has been shown to be both required and sufficient for ICL unhooking. Patients with FAN1 mutations do not have FA but instead have a unique kidney disease called karyomegalic interstitial

nephritis, a disease that is entirely separate from FA. These observations led to a search for the true *in vivo* substrate of FAN1 to provide an understanding of the unique nuclease activity of SAN1 in the context of the landscape of ICL repair pathways. SAN1 is able to digest small fragments of DNA and does not appear to be processive, supporting a model in which SAN1 acts after an initial cleavage event by an endonuclease. As it is still unclear if and how each nuclease specifically functions in ICL repair, there remains the possibility that SAN1 has redundant function to other known nucleases including SNM1A or FAN1. Colony survival assay evidence shows that SAN1 does not appear to be acting epistatically to FAN1 but this does not rule out the possibility that these nucleases could perform similar functions in different pathways. This lack of epistasis also provides an explanation for why there would be the need for an additional nuclease in ICL repair. We do not yet know the exact step in ICL repair during which SAN1 functions. A remaining question is whether SAN1 acts in coordination with protein cofactors other than Senataxin that have not yet been identified.

Although I have shown that SAN1 does not appear to function in the FA pathway and we believe SAN1 functions in a transcription-coupled ICL repair pathway, there are still many areas of TC-ICL repair that are not clear. The fact that SAN1 was not previously identified as a nuclease involved in ICL repair is likely due to its lack of involvement in the FA pathway, and accordingly, it would not have been identified in a FA pathway-related screen. It is now clear that Senataxin, a SAN1 interacting protein, acts with BRCA1 to prevent the accumulation of R-loops in ICL repair. The connection between R-loop formation, and ICL repair, though, remains poorly understood[110]. SAN1 $^{-/-}$ cells

are sensitive to crosslinking agents and thus we predict that R-loops form as a result of stalling of the transcription complex at an ICL. An interesting way to test the relationship between ICL formation and R-loop generation would be to stimulate R-loop formation on a plasmid with a single site specific ICL and observe which factors are required for resolution of the lesion by immunodepletion. Several studies have been published using plasmids with single site-specific ICLs in a replication-dependent repair context; development of a similar system in a replication-independent repair context would shed light on the molecular mechanisms of repair[59,60]. The R-loop must be resolved by a resolvase enzyme, such as Senataxin, that recruits SAN1 to the site of the ICL, allowing for SAN1 to act on an intermediate splayed arm structure after resolution of the R-loop and before the action of a nuclease such as SNM1a that can unhook crosslinks. Preliminary data shows that SAN1 appears to act epistatically to SNM1a, as there was no significant decrease in cell survival in a $SAN1^{-/-}$ SNM1a knockdown cell line treated with MMC (data not shown).

Regulation of nucleases in ICL repair is critical for maintaining genomic integrity, and identifying different modes of SAN1 regulation would provide a stronger understanding of the TCR pathway. FA pathway nucleases are regulated by ubiquitylation of the FANCD2-FANCI complex and recruitment of the critical scaffolding protein SLX4. Although SLX4 does not have nuclease activity it is absolutely essential for FA pathway-dependent ICL repair and acts as a recruitment factor and docking location for the ICL repair nucleases, XPF/ERCC1, SLX1, and MUS81/EME1.

Taken together, this work shows that SAN1 binding to the RNA/DNA helicase SETX is required for SAN1 to function in the response to ICLs. SAN1 and SETX likely cooperate in the same pathway, as depletion of SETX is epistatic with loss of SAN1. Consistent with this model, loss of SAN1 leads to an increase in R-loops following treatment with MMC, likely as a result of the inability of SAN1 to participate in a step of ICL repair. Collectively these studies support the existence of repair pathway in which SAN1 acts independently of the FA pathway to protect cells from ICLs. SAN1 might act in a backup pathway that is required when the FA pathway is overwhelmed. It is tempting to speculate that SAN1 acts with SETX at ICL sites where transcription complex stalling has occurred, and that the subsequent formation of R-loops recruits SETX, which in turn recruits SAN1. The formation of a ssDNA break near an RNA:DNA hybrid could lead to a 5' splayed arm-like substrate after resolution of the R-loop by SETX. This would allow for digestion of the ssDNA flap adjacent to an ICL by SAN1, possibly providing a better substrate for the 5' exonuclease SNM1A, which might unhook the lesion by digesting past the crosslink (Figure 22). Although further experimentation is required to understand the exact mechanism of SAN1 function in ICL repair, this study represents a significant advance in the understanding of a previously uncharacterized FEN1 family member and provides evidence for a new member of a replication-independent ICL repair pathway.

4.2 Future Directions

One of the main remaining questions regarding the function of SAN1 is in which step of ICL repair and on which ICL repair intermediates it is acting. The evidence that SAN1

acts epistatically of SNM1a provides support that SAN1 may be cleaving a precursor substrate that then is acted on by SNM1a. This could be tested by examining the nuclease activity of each protein separately followed by examining the nuclease activity of the combination of SAN1 and SNM1a in a nuclease assay on an ICL substrate. If SAN1 is able to act on a splayed arm structure with an ICL generating a substrate that SNM1a could then digest the ICL would be unhooked. However, we have not yet tested that experimentally.

In addition to determining in which step of ICL repair SAN1 is acting, it would also be interesting to understand the preference of a splayed arm substrate. I developed several new oligonucleotides with several different lengths of 3' ssDNA regions hypothesizing that if SAN1 is able to interact with a longer 3' ssDNA region the 5' strand will be more efficiently cut by the nuclease domain. Because I observed that SAN1 is able to cleave ssDNA, splayed arm, and 5' flap structures but the splayed arm structures are cleaved the most efficiently, I posit that SAN1 would show increased activity against the substrate if the 3' ssDNA region of the splayed arm is longer and thus provided a longer region for SAN1 to interact with. Understanding how SAN1 recognizes substrates beyond just requiring a free 5' end of DNA would significantly advance our understanding of how the protein functions *in vivo*.

Another important next step in the study of SAN1 will be to understand the regulation of the nuclease. FA pathway nucleases are regulated by activation of the FA pathway through phosphorylation and ubiquitylation along with recruitment to ICL sites through

SLX4[113]. SLX4 is a critical scaffolding protein that does not have nuclease activity but is required in the FA pathway to recruit the toolkit of nucleases required for unhooking the lesion[64]. SLX4 is recruited to the site of the ICL by monoubiquitylated FANCD2-FANCI[63]. Identifying additional proteins with which SAN1 interacts or that regulate SAN1 will provide a deeper understanding of replication-independent ICL repair. As a critical SAN1 interacting protein, Senataxin binding to SAN1 is important for the ability of SAN1 to function in ICL repair. Many nucleases in DDR pathways require or are stimulated by the action of helicases[45]. One particularly interesting interaction is between the helicase BLM and the FEN1 family nuclease EXO1. BLM stimulates EXO1 nuclease activity on DNA substrates during HR[45,95]. It will be important to further characterize the interaction between the Senataxin helicase and SAN1 nuclease activity, and in particular, to understand if Senataxin is able to stimulate nuclease activity on substrates.

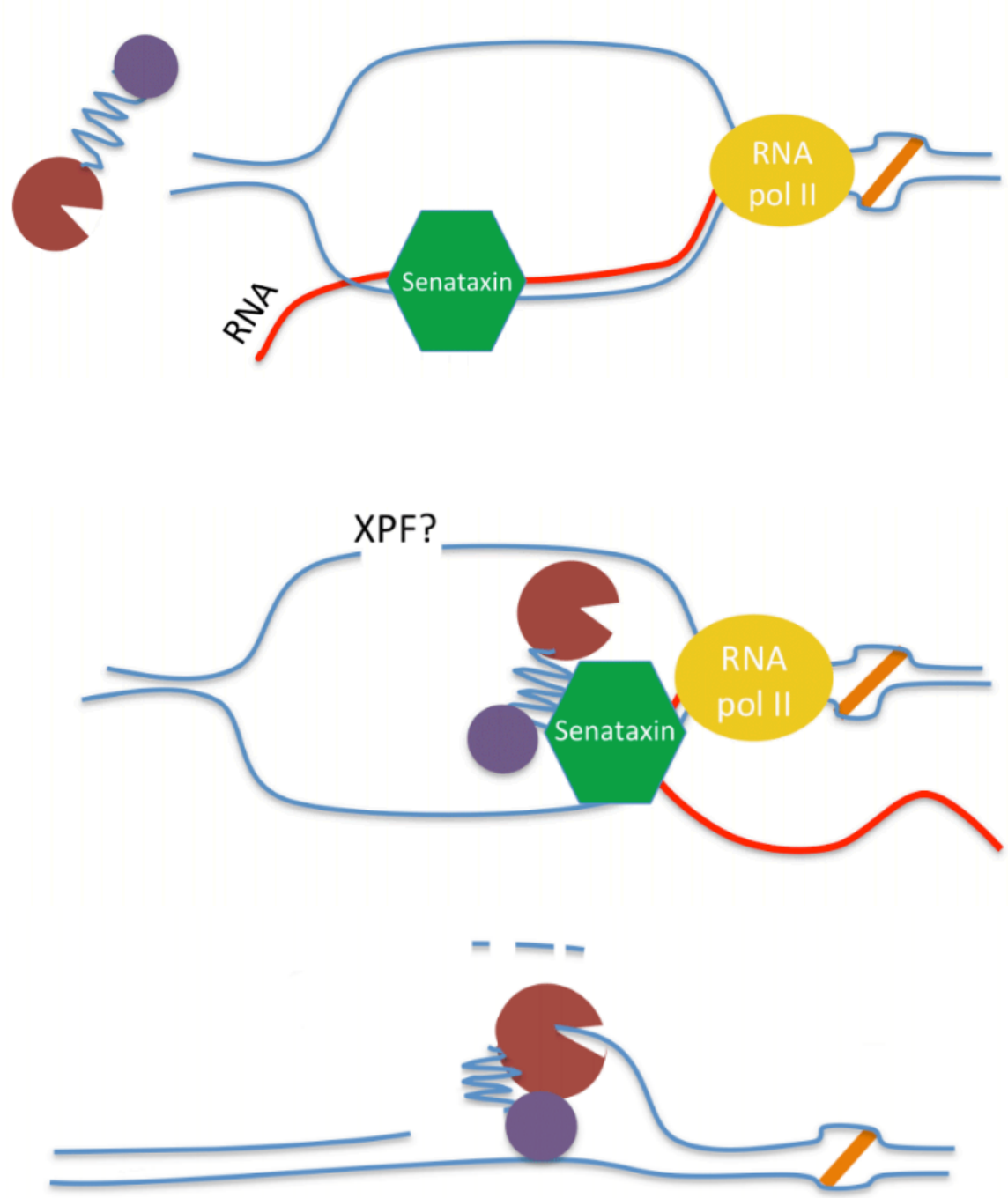
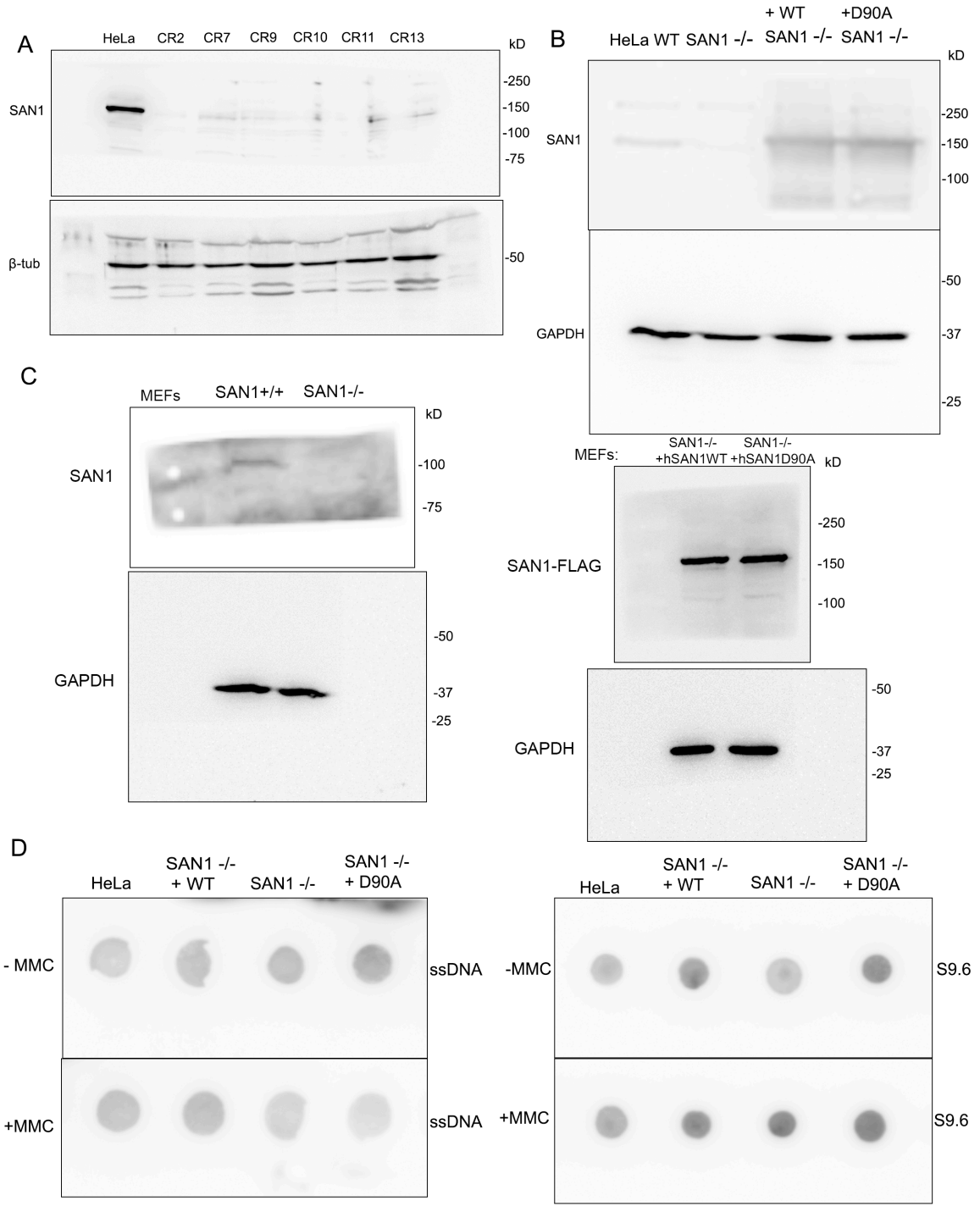


FIGURE 22. Speculative model for SAN1 function with SETX in processing of ICLs. Collision of transcription complexes with an ICL results in the formation of R-loops. The helicase SETX is recruited to the RNA/DNA hybrid and unwinds it. SETX also recruits the SAN1 nuclease. If an incision is generated in the ssDNA loop, perhaps by XPF, SAN1 can digest the free 5' end back to the ICL, where other nucleases participate in unhooking of the lesion.



Alex Andrews – Figures B, C

Figure 23. Raw blot data from Figures 16, 17, 21. (a) SAN1-/- HeLa cell clones generated using CRISPR/CAS9 and loading controls as shown in Figures 16, 17. (b) SAN1WT and SAN1D90A rescue lines and loading control as shown in Figure 16. 17.

(c) SAN1+/+ and SAN1-/- MEFs raw western blots from Figure 16. Middle right panels: MEF rescue cell line raw western blots and loading control from Figure 16. (d) Representative slot blot images (ssDNA blot and S9.6 blot) for HeLa, SAN1-/-, SAN1-/- +WT, SAN1-/- +D90A cells treated with or without MMC from Figure 21.

mSAN1 Yeast Two-Hybrid - Hybrigenics							
Gene Name	Gene ID (NCBI)	Global PBS	Start	Stop	Frame	Sense	% Id 5p/3p
Mus musculus - Clpb	20480	B	231	1005	IF	Sense	99.1 / 97.4
Mus musculus - Clpb	20480	B	258	924	IF	Sense	100.0 / 99.7
Mus musculus - Clpb	20480	B	282	1014	IF	Sense	99.5 / 99.7
Mus musculus - Hipk3	15259	B	1995	3068	IF	Sense	99.3 / 99.6
Mus musculus - Hipk3	15259	B	2151	3147	IF	Sense	99.4 / 99.0
Mus musculus - Hipk3	15259	B	2331	3159	IF	Sense	98.3 / 99.6
Mus musculus - Hipk3	15259	B	2373	3385	IF	Sense	98.6 / 99.7
Mus musculus - Lonrf1	244421	B	1125	2202	IF	Sense	98.5 / 99.5
Mus musculus - Lonrf1	244421	B	1443	2214	IF	Sense	98.7 / 98.5
Mus musculus - Lonrf1	244421	B	1443	2214	IF	Sense	99.5 / 99.3
Mus musculus - Lonrf1	244421	B	1443	2214	IF	Sense	99.3 / 99.4
Mus musculus - Pias2	17344	A	627	1792	IF	Sense	99.2 / 99.3
Mus musculus - Pias2	17344	A	1002	1747	IF	Sense	99.9 / 99.8
Mus musculus - Pias2	17344	A	1245	1746	IF	Sense	99.4 / 99.4
Mus musculus - Pias2	17344	A	1248	1755	IF	Sense	99.4 / 99.4
Mus musculus - Pias2	17344	A	1248	1755	IF	Sense	99.4 / 99.4
Mus musculus - Pias2	17344	A	1248	1755	IF	Sense	99.4 / 99.4
Mus musculus - Pias2	17344	A	1248	1755	IF	Sense	99.4 / 99.4
Mus musculus - Pias2	17344	A	1248	1755	IF	Sense	99.4 / 99.4
Mus musculus - Pias2	17344	A	1248	1755	IF	Sense	99.4 / 99.4
Mus musculus - Pias2	17344	A	1248	1755	IF	Sense	99.4 / 99.4
Mus musculus - Pias2	17344	A	1251	1846	IF	Sense	92.5 / 90.3
Mus musculus - Pias2	17344	A	1251	2235	IF	Sense	98.2 / 92.2
Mus musculus - Pias2	17344	A	1257	1680	IF	Sense	100.0 / 100.0
Mus musculus - Pias2	17344	A	1290	1871	IF	Sense	85.7 / 85.7
Mus musculus - Pias2	17344	A	1290	1871	IF	Sense	85.7 / 85.7
Mus musculus - Pias2	17344	A	1332	1755	IF	Sense	100.0 / 100.0
Mus musculus - Pias2	17344	A	1332	1755	IF	Sense	100.0 / 100.0
Mus musculus - Pias2	17344	A	1332	1755	IF	Sense	100.0 / 100.0
Mus musculus - Rnf111	93836	B	861	1560	IF	Sense	99.9 / 99.6
Mus musculus - Rnf111	93836	B	861	1560	IF	Sense	99.9 / 100.0
Mus musculus - Rnf111	93836	B	954	1548	IF	Sense	100.0 / 99.8
Mus musculus - Setx	269254	B	1908	3369	IF	Sense	99.6 / 99.9
Mus musculus - Setx	269254	B	1908	3369	IF	Sense	98.1 / 98.5
Mus musculus - Setx	269254	B	2520	3432	IF	Sense	99.9 / 99.9
Mus musculus - Setx	269254	B	2625	3310	IF	Sense	99.6 / 99.6
Mus musculus - Smek1	68734	A	-52	947	IF	Sense	99.2 / 99.1
Mus musculus - Smek1	68734	A	-40	1133	IF	Sense	99.6 / 98.7
Mus musculus - Smek1	68734	A	-40	1133	IF	Sense	98.6 / 99.9
Mus musculus - Smek1	68734	A	330	948	IF	Sense	100.0 / 99.7

Mus musculus - Smek1	68734	A	375	1062	IF	Sense	99.9 / 99.0
Mus musculus - Smek1	68734	A	375	1062	IF	Sense	100.0 / 99.0
Mus musculus - Zfp142	77264	B	855	1518	IF	Sense	99.4 / 99.5
Mus musculus - Zfp142	77264	B	1026	1889	IF	Sense	99.0 / 97.7
Mus musculus - Zfp142	77264	B	1128	1954	IF	Sense	99.0 / 99.9
Mus musculus - Zfp142	77264	B	1155	1512	IF	Sense	100.0 / 100.0
Mus musculus - Zfp142	77264	B	1155	1512	IF	Sense	100.0 / 100.0
Mus musculus - Zfp142	77264	B	1155	1512	IF	Sense	100.0 / 100.0
Mus musculus - Zfp142	77264	B	1155	1512	IF	Sense	100.0 / 100.0
Mus musculus - Zfp142	77264	B	1155	1512	IF	Sense	100.0 / 100.0
Mus musculus - Zfp142	77264	B	1155	1512	IF	Sense	100.0 / 100.0
Mus musculus - Zfp142	77264	B	1155	1512	IF	Sense	100.0 / 100.0
Mus musculus - Zfp142	77264	B	1155	1512	IF	Sense	100.0 / 100.0
Mus musculus - Zfp142	77264	B	1155	2013	IF	Sense	99.7 / 100.0
Mus musculus - Zfp142	77264	B	1155	1512	IF	Sense	100.0 / 100.0
Mus musculus - Zmym2	76007	A	951	1956	IF	Sense	99.7 / 99.8
Mus musculus - Zmym2	76007	A	951	1956	IF	Sense	99.9 / 99.9
Mus musculus - Zmym2	76007	A	1056	2202	IF	Sense	99.1 / 99.9
Mus musculus - Zmym2	76007	A	1056	2202	IF	Sense	97.9 / 99.3
Mus musculus - Zmym2	76007	A	1086	2307	IF	Sense	95.9 / 98.2
Summary of PBS categories							
A : Very high confidence in the interaction							
B : High confidence in the interaction							

Figure 24. Genome-wide yeast two-hybrid results from Hybrigenics. Full-length murine SAN1 was used as bait and was screened against a mouse adult brain library with a complexity of ~1x10^7, at 10-fold coverage. From 103 million analyzed reactions, 183 clones were processed, and 9 positive clones were in-frame and were present more than once.

Appendix II: Oligonucleotides used in nuclease assays

Table 1. Oligos used in nuclease assays

X1	5'-TGGGTCAACGTGGGCAAAGATGTCCTAGCAATGTAATCGTCTATGACGTT-3'	ssDNA	50 nt
X4	5'-AACGTCATAGACGATTACATTGCTAGGACATCTTTGCCCACGTTGACCCA-3'	ssDNA	50 nt
a3	5'-CCTCGATCCTACCAACCAGATGACGCGCTGCTACGTGCTACCGGAAGTCG-3'	Splayed arm, 3' flap, 5' flap, replication fork	50 nt
b	5'-CGACTTCCGGTAGCACGTAGCAGCGGCTCGCCACGAACTGCACTCTAGGC-3'	Splayed arm, 3' flap, 5' flap, replication fork	50 nt
c	5'-GCCTAGAGTGCAGTTCGTGGCGAGC-3'	Splayed arm, 3' flap, 5' flap, replication fork	25 nt
d3	5'-CGTCATCTGGTTGGTAGGATCGAGG-3'	Splayed arm, 3' flap, 5' flap, replication fork	25 nt
X3	5'-TGGGTCAACGTGGGCAAAGATGTCC-3'	ssDNA, first 25 nt of X1	25 nt
N1	5'-TGGGTCAACGTGGGCAAAGA-3'	Nicked	20 nt
N2	5'-ATGTAATCGTCTATGACGTT-3'	Nicked	20 nt
N3	5'-AACGTCATAGACGATTACATTCTTTGCCCACGTTGACCCA-3'	Nicked	40 nt
G1	5'-TGGGTCAACGTGGGCAAAG-3'	Gapped	19 nt
G2	5'-ATGTAATCGTCTATGACGTT-3'	Gapped	20 nt
G3	5'-AACGTCATAGACGATTACATTCTTTGCCCACGTTGACCCA-3'	Gapped	40 nt
X4 Biotin 5'	5'-Biotin-AACGTCATAGACGATTACATTGCTAGGACATCTTTGCCCACGTTGACCCA-3'	X4 - 5' Biotin	50 nt
X1A	5'-TGGGTCAACGTGGGCAAAGATTTTTTTTTTTTTTTTTTTTTTTTTGTCCTAGCAATGTAATCTG-3'	Variation of X1 - 1st 20 nt of X1, 20 internal Ts, middle 20 nt of X1	60 nt
X1B	5'-TTTTTTTTTTTTTTTTTTTTTTGGGTCAACGTGGGCAAAGATTTTTTTTTTTTTTTTTTTT-3'	Variation of X1 - 20 Ts, 1st 20 nt of X1, 20 Ts	60 nt
S3	5'-TGGGTCAACG-3'	Ladder - 1st 10 nt from X1 oligo	10 nt
S4	5'-TGGGTCAACGTGGGC-3'	Ladder - 1st 15 nt from X1 oligo	15 nt
S5	5'-TGGGTCAACGTGGGCAAAGA-3'	Ladder - 1st 20 nt from X1 oligo	20 nt
S6	5'-TGGGTCAACGTGGGCAAAGATGTCC-3'	Ladder - 1st 25 nt from X1 oligo	25 nt
S7	5'-TGGGTCAACGTGGGCAAAGATGTCCTAGCA-3'	Ladder - 1st 30 nt from X1 oligo	30 nt
X0-1	5'-ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGCAGGTTCACCC-3'	Holliday Junction	60 nt
X0-2	5'-GGGTGAACCTGCAGGTGGGCAAAGATGTCCATCTGTTGTAATCGTCAAGCTTTATGCCGT-3'	Holliday Junction	60 nt
X0-3	5'-ACGGCATAAAGCTTGACGATTACAACAGATCATGGAGCTGTCTAGAGGATCCGACTATCG-3'	Holliday Junction	60 nt
X0-4	5'-CGATAGTCGGATCCTCTAGACAGCTCCATGTAGCAAGGCACTGGTAGAATTCGGCAGCGT-3'	Holliday Junction	60 nt

1. Plasmids and siRNA

-Human SAN1 cDNA (FLJ56631) was purchased from the NBRC, Japan, and cloned into the pRK7 expression vector with a C-terminal FLAG tag. SAN1 (D90A) and SAN1 rescue constructs were made by QuikChange site-directed mutagenesis (Agilent Technologies). Partial mouse SAN1 open reading frames were obtained from GE Healthcare Dharmacon MGC cDNAs library and cloned into pASK-IBA3Plus to express full length SAN1-Strep-tag II. The D90A mutation was introduced using Stratagene QuikChange. pICE-RNaseHI-WT-NLS-mCherry was a gift from Patrick Calsou (Addgene plasmid # 60365). For RNAi knockdown experiments SMARTpool siGENOME (Thermo Fisher and Dharmacon) siRNAs were used for SETX (M-021420-01-0005) and FAM120B (M-014898-00-0005). Cells were transfected, split 24 h later, and transfected again, with specific siRNA pools or a scrambled control (D-001810-01-05) using RNAiMAX (Invitrogen) and Opti-MEM media (Thermo Fisher) according to manufacturer's instructions. Colony survival assays were performed and lysates made 36 h after the second round of siRNA transfection. Knockdown was determined by western blotting using FANCD2 (1:500, Novus Biologicals NB100-182), SETX (1:1000, Novus Biologicals NB100-57542), SNM1a (1:1000, Bethyl Labs A303-747A-M), XPF (1:1000, Bethyl Laboratories A301-315A), FAN1 (1:1000, non-commercial), FAM120b (1:1000, Abcam ab106455).

2. Production of hSAN1 from mammalian cells

-pKSAN1WT-FLAG or pKSAN1D90A-FLAG constructs were transfected into 293T cells. At 24 hrs post-transfection cells were washed once with 1× PBS followed by addition of Lysis Buffer (25 mM HEPES pH 7.4, 10% glycerol, 1 mM EDTA, 1 mM DTT, 20 nM calyculin A (Sigma) and protease inhibitors (Roche)) and incubation on ice for 5 min. Samples were snap frozen in liquid nitrogen and subsequently thawed at 37 °C three times. NaCl was added to 300 mM. Lysates were then centrifuged at 16,100×*g*for 20 min at 4 °C and the supernatant was added to an equal amount of Lysis Buffer plus 0.2% NP-40 and then centrifuged at 16,000×*g* for 5 min at 4 °C. Lysates were added to washed mouse anti-FLAG M2 agarose (Sigma) for 1 h at 4 °C. Samples were washed 2× with Wash Buffer (Lysis Buffer plus 0.1% NP-40 and 150 mM NaCl) for 15 min per wash at 4 °C. Samples were then washed 2× with Elution Buffer (62.5 mM HEPES pH 7.4, 62.5 mM KCl, 5% glycerol, 1 mM DTT, 50 μg/mL BSA) for 15 min per wash at 4 °C, and SAN1 protein was eluted from beads with Elution Buffer + 150 ng/μL FLAG peptide (20 μL/10 cm, 60 μL/15 cm) by shaking (600 rpm) at 4 °C for 30 min. Protein concentration was estimated by SDS-PAGE followed by Coomassie brilliant blue staining in parallel with known amounts of BSA.

3. ^{32}P Labeling and oligonucleotide annealing

-Defined oligonucleotide structures (ssDNA, dsDNA, nicked and gapped structures, replication fork, splayed arm structure, 5' and 3' flap structures) were 5' -labeled with 5 pmol γ -[^{32}P] ATP using T4 polynucleotide kinase (NEB) or 3' -labeled with 5 pmol α -[^{32}P] CordycepinTP using DNA terminal transferase (20 U) in 1× TdT buffer, supplemented with 2.5 mM $CoCl_2$ in a 10 μL reaction at 37 °C, according to standard methods (NEB). Oligonucleotides were annealed in 6× annealing buffer (0.9 M NaCl, 90 mM sodium citrate) by slow cooling from 95 °C and purified from 12% native PAGE gel by the crush and soak method. Oligonucleotide sequences are given in Appendix II.

4. Nuclease assay standard reaction conditions

-All reactions were carried out in nuclease buffer (62.5 mM HEPES pH 7.4, 62.5 mM KCl, 5 % Glycerol, 1 μM DTT, 50 μ g/mL BSA). Five micromolar of substrate was added to 4 μL of protein (27.5 nmol) or water and after the addition of Start Buffer (6 mM $MgCl_2$, 2 mM β -ME, 0.05 μg/μL BSA) reactions were incubated at 37 °C for 2 h (unless otherwise noted). Reactions were stopped using 2.5 μL of 5 × Stop Buffer (15 mM EDTA, formamide, 0.005% Bromophenol blue, 0.005% Xylene cyanol, 5% glycerol) followed by boiling the samples for 3 min. Unless otherwise stated DNA substrate sequences for individual experiments are shown schematically in each figure and listed in Appendix II. Reactions were analyzed on 12% denaturing PAGE gel (1× TBE, 7 M urea, 12% 19:1 acrylamide/bis-acrylamide 19:1).

5. RNA/DNA hybrid slot-blot assay

-Hybrids were detected as described by Sollier et al (2014). Total nucleic acids were extracted from cells with the Qiagen DNeasy kit. DNA (1 μg) was spotted in duplicate wells onto positively-charged nylon membrane using a slot-blot apparatus, cross linked by UV treatment, and one well was probed with the S9.6 antibody (1:1000, Kerafast, 1:200 ENH001). The DNA in the duplicate well was denatured for 10 min in 0.5N NaOH, 1.5M NaCl, then neutralized for 10 min in 1M NaCl, 0.5 M Tris-HCl (pH 7.0) and probed with an antibody against ssDNA (1:10000, Millipore MAB3034). Spots were detected with an anti-HRP secondary anti-mouse antibody, and imaged and quantified with an Amersham Imager 600.

6. Yeast-two hybrid screen

-The yeast two-hybrid screen using full-length murine SAN1 as bait was carried

out by Hybrigenics Corporation, Cambridge, MA using a mouse brain library. Additional putative interacting proteins shown in Appendix I (Figure 24).

7. Cell cycle analysis

-HeLa cells were transfected with indicated siRNA and then treated with vehicle or 5 μM Cisplatin for 2 hrs, replaced with fresh media and 24 hrs later harvested for flow cytometric analysis. Cells were washed in PBS and then fixed in ice-cold 70% ethanol for at least 2 hrs. The cells were washed in PBS and resuspended in propidium iodide staining solution (PBS, 0.1% TritonX-100, 0.2 mg/mL DNase-free RNase A (Sigma), 20 μg/m LPI (Sigma)) and analyzed using a FACSCalibur machine (BD).

8. Radial Chromosome Assays

-HeLa cells were incubated for 48 hrs in the presence or absence of MMC. Colcemid (0.1 g/ml) was added to the medium 2 hrs before the cells were collected. For each sample, 50 metaphases were analyzed for chromosomal abnormalities, as previously described (Shimamura et al., 2002). Data were analyzed in Prism GraphPad from contingency tables using Fisher's exact-test (two-sided P value).

9. Proliferation Assay

-HeLa cells and SAN1$^{-/-}$ HeLa cells were plated in 6 well plates at 100 cells/well and allowed to adhere. Each day a number of the wells were counted. Relative cell growth was determined by dividing each time point by day 1.

10. Electrophoretic Mobility Shift Assay with SAN1 D90A

-^{32}P-labeled X4 (1 nM) was incubated with different concentrations of SAN1 D90A (0, 1, 5, 10, 20 nM) or RPA (10 nM) in binding buffer (20 mM HEPES pH 7.4, 0.1% NP-40, 0.1 M KCl, 5 mM $MgCl_2$, 1% glycerol, 0.25 μg/μl BSA, 1 mM DTT) for 30 minutes at room temperature. 4 μl of 6X Ficoll loading dye was added to the reactions and 10 μl of each reaction was separated by electrophoresis on a 5% gel (37.5:1 polyacrylamide, 0.5X TBE) at 30 V for 3 hrs in a cold room with pre-chilled 0.5X TBE. Gels were dried and visualized by autoradiography.

11. Double-affinity Strep-FLAG WT and D90A SAN1 purification

-HeLa SAN1$^{-/-}$ cells + SAN1 WT-SSF or SAN1 D90A-SSF cells were washed once with 1X PBS followed by addition of Lysis Buffer 1 (50 mM Tris HCl pH 7.4, 1 mM $MgCl_2$, protease inhibitors (Roche). Samples were snap frozen in liquid nitrogen and subsequently thawed at 37°C three times. NaCl was added to 150 mM. Lysates were then centrifuged at 16,100 x g for 20 min at 4°C and the supernatant was added to washed Strep-Tactin beads (IBA) and rotated for 1 hour at 4°C. Samples were washed 2X with Wash Buffer (50 μM Tris HCl pH 7.4, 150 mM

NaCl, protease inhibitors) for 15 min rotating per wash at 4°C. SAN1 protein was eluted from beads with Elution Buffer (50 mM Tris HCL pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 2.5 mM desthiobiotin) by shaking (600 rpm) at 4°C for 30 min. Samples were then added to equal volume of 2X Lysis Buffer 2 (50mM HEPES pH 7.4, 20% glycerol, 2 mM EDTA, 2mM DTT, 40 nM calyculin A (Sigma) and protease inhibitors (Roche)) and added to washed mouse anti-FLAG M2 agarose (Sigma) rotating for 1 hr at 4°C. Samples were washed 1X with Wash Buffer (Lysis Buffer 2 plus 0.1% NP-40 and 150 mM NaCl) for 15 min rotating at 4°C. Samples were then washed 1X with Elution Buffer (62.5 mM HEPES pH 7.4, 62.5 mM KCl, 5% glycerol, 1 mM DTT) for 15 min rotating at 4*C. SAN1 protein was eluted from beads with Elution Buffer + 150 ng/μL FLAG peptide by shaking (600 rpm) at 4°C for 30 min. Samples from each step of the double-affinity Strep-FLAG purification were visualized by SDS-PAGE followed by silver stain or immunoblot (using mouse M2 anti-FLAG antibody 1:1000, Sigma) or were tested for nuclease activity using the nuclease filter-spin assay described previously.

12. Kinetic Characterization of hSAN1

-Reactions were carried out with varied concentrations of substrate (10 -1000 nM) and 240 pmol enzyme in Elution Buffer. Reactions containing SAN1 WT and substrate were pre-warmed to 37 °C and initiated by the addition of Start Buffer. Reactions were sampled at 3 time points by addition of Stop buffer, and processed using the Nucleotide Removal kit (Qiagen). Products were quantified using Ecoscint Original scintillation liquid (National Diagnostics) and a scintillation counter and were analyzed using GraphPad Prism.

13. Co-IP of N-terminus and C-terminus of SAN1

-15 μg of pKhSAN1-Nuclease-FLAG and 15 μg of pKMyc-mSAN1-Cterminus were co-transfected into 293T cells by calcium phosphate transfection. In samples where only pKhSAN1-Nuclease-FLAG or pKMyc-mSAN1-Cterminus was transfected, 30 μg of DNA was used. Cells were lysed in buffer containing 25 mM HEPES pH 7.4, 150 mM NaCl, 0.5% Triton-X 100, 0.5 mM EDTA, 1 mM $MgCl_2$, 2 mM DTT, 1 mM PMSF, 10 μg/ml leupeptin, 20 ug/ml aprotinin 24 hrs later. Cell lysates were then incubated with anti-FLAG M2 agarose for 1 hr at 4°C rotating before 3 washes. Samples were boiled for 8 min in 4X LSB and analyzed by immunoblot (Myc 1:1000 9E10 clone, FLAG 1:1000 Sigma F1804). pKMycRhoA was co-expressed with pKhSAN1-Nuclease-FLAG as a negative control.

14. Recombinant murine SAN1 expression and purification

-The pASK-SAN1 plasmids were transformed into BL21-CodonPlus (DE3)-RIPL bacterial cells. SAN1 expression was induced by the addition of tetracycline (0.2 μg/mL) to 0.5 L of bacteria in early exponential phase in liquid culture followed by overnight incubation at 18°C with shaking at 250 rpm. Cells were centrifuged (5000 rpm, 20 min, 4°C) and the pellet was resuspended in lysis buffer (50 mM

Tris HCl pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM PMSF, protease inhibitors (Roche)). Lysates were treated with lysozyme (1mM, 4°C, 15 min) before being ruptured in a French Press (1500 psi) twice. Lysates were then centrifuged (13.2 rpm, 20 min, 4°C). The supernatant was loaded onto a pre-equilibrated 1 mL CV Gravity flow Strep-Tactin Sepharose column (IBA). The column was washed with 20 CV buffer w (50 mM Tris HCl pH 7.4, 150 mM NaCl) and eluted with buffer e (50 mM Tris HCl pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 2.5 mM desthiobiotin) in 2.5 mL fractions. Samples from each elution fraction were subjected to SDS-PAGE. Fractions containing SAN1 were combined and protein concentration was estimated by SDS-PAGE followed by brilliant blue staining in parallel with known amounts of BSA.

15. Colony Survival Assays

-HeLa or mouse embryonic fibroblast cells were seeded at 300-400 cells per 6 well dish for 16 hrs overnight, treated with DNA damaging agents, and allowed to form colonies for 7-10 days. Colonies were fixed in ice cold 70% EtOH, stained with crystal violet, and counted.

16. Cell Culture and Transfections

-HeLa (ATCC) and 293T (ATCC) cells were grown in DMEM supplemented with 10% FBS, 100 U/mL penicillin and 100 U/mL streptomycin (GIBCO). Transfection of plasmids or siRNA was performed with calcium phosphate, Lipofectamine 2000 (Invitrogen), or Lipofectamine RNAiMAX (Invitrogen). Virus was made, collected and titered as described previously (McCaffrey et al. 2009). Cell lines verified mycoplasm free by DAPI staining. Cell lines validated by DNA sequencing.

17. Production of $SAN1^{-/-}$ HeLa cells

-Two 60-bp guide sequences (sgRNA1 F: 5′-TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGCAGGATAAGAGAGATGAAT -3′ and sgRNA2 F: 5’ – TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGAGAAGCTCTGTGAGAGTCT – 3’) were designed to target sites in the first exon of SAN1, to remove the majority of the nuclease domain. The sgRNA1 and sgRNA2 were cloned into gRNA cloning vectors and verified by sequencing. gRNA1, gRNA2, and NLS-hCas9-NLS constructs were transfected into HeLa cells using Lipofectamine 2000 (Invitrogen 11668-030) per Invitrogen protocol publication number MAN0007824 Rev 1.0 and seeded at single cell density on 15 cm dishes. Individual colonies were isolated by 0.25% trypsin and were plated separately in 6 well plates to grow. Genomic DNA was isolated from WT HeLa cells and from 12 clones. Exon 1 of SAN1 was PCR amplified using SAN1 genomic PCR primers (SAN1 genomic F: 5’ – ACTGATTAATTTATCTTTCTTTCCAGATCC – 3’ and SAN1 genomic R: 5’TCTGGGATTATGTCGTTGCCAAGGAGG – 3’) and clones

that showed a deletion were sequenced and analyzed by immunoblot (SAN1 1:1000). 6 SAN1$^{-/-}$ HeLa clones were identified and experiments in this study were completed with SAN1$^{-/-}$ HeLa Clone 2 unless otherwise specified.

18. Generation of Mouse Embryonic Fibroblasts (MEFs)

-A conditional KO mouse was created through the Texas A&M Institute for Genomic Medicine, using EUCOMM ES cells targeting the *fam120b* gene (ES cell clone HEPD0652_5_G10). A *fam120b*/+ male was crossed to a FLPer/+ female to delete the lacz/neo markers, the FLPer transgene was then removed by crossing to a +/+ mouse, and the resulting floxed allele mice were crossed to produce homozygotes. These were then crossed with a Sox2-Cre mouse to obtain a global knockout of the allele. *Fam120b* +/+ and $^{-/-}$ MEFs were isolated from day 13.5 embryos, from matings of heterozygous parents. Embryos were incubated in 0.25% Trypsin overnight at 4° C for digestion to single cells. Cells were plated and cultured in 10% FBS, 100 U/mL penicillin and 100 U/mL streptomycin (GIBCO). MEFs were immortalized by transfection with a plasmid encoding SV40 Large T antigen (Addgene 21826), followed by a 1/10 split of the cells 5-6 times. Genotypes were validated by genomic PCR and western blotting using a non-commercial rabbit polyclonal antibody against murine SAN1 (1:1000).

References

1. Lukas J, Lukas C, Bartek J. Mammalian cell cycle checkpoints: Signalling pathways and their organization in space and time. *DNA Repair (Amst)*. 2004;3(8-9):997-1007. doi:10.1016/j.dnarep.2004.03.006.
2. Shaltiel IA, Krenning L, Bruinsma W, Medema RH. The same, only different - DNA damage checkpoints and their reversal throughout the cell cycle. *J Cell Sci*. 2015;128(4):607-620. doi:10.1242/jcs.163766.
3. AlexandreMaréchal, Zou L. DNA Damage Sensing by the ATMand ATR Kinases. *Cold Spring Harb Perspect Biol*. 2013:1-18. doi:10.1101/cshperspect.a012716.
4. Harper JW, Elledge SJ. The DNA damage response: ten years after. *Mol Cell*. 2007;28(5):739-745. doi:10.1016/j.molcel.2007.11.015.
5. Bonilla CY, Melo JA, Toczyski DP. Colocalization of Sensors Is Sufficient to Activate the DNA Damage Checkpoint in the Absence of Damage. *Mol Cell*. 2008;30(3):267-276. doi:10.1016/j.molcel.2008.03.023.
6. Zhou BS, Elledge SJ. checkpoints in perspective. 2000;408(November):433-439.
7. Bartek J, Lukas J. DNA damage checkpoints: from initiation to recovery or adaptation. *Curr Opin Cell Biol*. 2007;19(2):238-245. doi:10.1016/j.ceb.2007.02.009.
8. Voulgaridou GP, Anestopoulos I, Franco R, Panayiotidis MI, Pappa A. DNA damage induced by endogenous aldehydes: Current state of knowledge. *Mutat Res - Fundam Mol Mech Mutagen*. 2011;711(1-2):13-27. doi:10.1016/j.mrfmmm.2011.03.006.
9. Stone MP, Cho Y, Huang H a I, et al. Unsaturated Aldehydes Derived from Lipid Peroxidation and Environmental Sources. *Acc Chem Res*. 2008.

10. Langevin F, Crossan GP, Rosado I V., Arends MJ, Patel KJ. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature*. 2011;475(7354):53-59. doi:10.1038/nature10192.

11. Ceccaldi R, Sarangi P, D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol*. 2016;17:337. http://dx.doi.org/10.1038/nrm.2016.48.

12. Tham KC, Kanaar R, Lebbink JHG. Mismatch repair and homeologous recombination. *DNA Repair (Amst)*. 2016;38(Muller 1916):75-83. doi:10.1016/j.dnarep.2015.11.010.

13. Deans AJ, West SC. DNA interstrand crosslink repair and cancer. *Nat Rev Cancer*. 2011;11(7):467-480. doi:10.1038/nrc3088.

14. Li GM. Mechanisms and functions of DNA mismatch repair. *Cell Res*. 2008;18(1):85-98. doi:10.1038/cr.2007.115.

15. Bagshaw ATM. Functional mechanisms of microsatellite DNA in eukaryotic genomes. *Genome Biol Evol*. 2017;9(9):2428-2443. doi:10.1093/gbe/evx164.

16. Jiricny J. The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol*. 2006;7(5):335-346. doi:10.1038/nrm1907.

17. Genschel J, Bazemore LR, Modrich P. Human exonuclease I is required for 5′ and 3′ mismatch repair. *J Biol Chem*. 2002;277(15):13302-13311. doi:10.1074/jbc.M111854200.

18. Kadyrova LY, Dahal BK, Kadyrov FA. Evidence that the DNA mismatch repair system removes 1-nucleotide Okazaki fragment flaps. *J Biol Chem*. 2015;290(40):24051-24065. doi:10.1074/jbc.M115.660357.

19. Robertson AB, Klungland A, Rognes T, Leiros I. Base excision repair: The long and short of it. *Cell Mol Life Sci*. 2009;66(6):981-993. doi:10.1007/s00018-009-8736-z.

20. Hegde ML, Hazra TK, Mitra S. Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res*. 2008;18(1):27-47. doi:10.1038/cr.2008.8.

21. Nakamura J, Swenberg J a. Endogenous Apurinic / Apyrimidinic Sites in Genomic DNA of Mammalian Tissues Endogenous Apurinic / Apyrimidinic Sites in Genomic DNA of Mammalian Tissues 1. 1999:2522-2526.

22. Fromme JC, Verdine GL. Base Excision Repair. *Adv Protein Chem*. 2004;69:1-41. doi:10.1016/S0065-3233(04)69001-2.

23. David SS, O'Shea VL, Kundu S. Base-excision repair of oxidative DNA damage. *Nature*. 2007;447(7147):941-950. doi:10.1038/nature05978.

24. Prasad R, Dianov GL, Bohr VA, Wilson SH. FEN1 stimulation of DNA polymerase β mediates an excision step in mammalian long patch base excision repair. *J Biol Chem*. 2000;275(6):4460-4466. doi:10.1074/jbc.275.6.4460.

25. Schärer OD. Nucleotide Excision Repair in Eukaryotes. *Cold Spring Harb Perspect Biol*. 2013;5(10):doi: 10.1101/cshperspect.a012609. doi:10.1016/j.dnarep.2015.09.003.

26. Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol*. 2014;15(7):465-481. doi:10.1038/nrm3822.

27. Fousteri M, Mullenders LHF. Transcription-coupled nucleotide excision repair in

mammalian cells: Molecular mechanisms and biological effects. *Cell Res*. 2008;18(1):73-84. doi:10.1038/cr.2008.6.

28. Gillet LC, Shaerer O.D. Molecular mechanism of mammalian global genome nucleotide excision repair. *Chem Rev*. 2006;106(2):253-276. doi:10.1021/cr040483f.

29. Shuck SC, Short EA, Turchi JJ. Eukaryotic nucleotide excision repair: From understanding mechanisms to influencing biology. *Cell Res*. 2008;18(1):64-72. doi:10.1038/cr.2008.2.

30. Sugasawa K, Ng JM, Masutani C, et al. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol Cell*. 1998;2(2):223-232. doi:10.1016/s1097-2765(00)80132-x.

31. de Laat W, Jaspers N, Hoeijmakers J. Molecular mechanism of nucleotide excision repair. *Genes Dev*. 1999;13(7):768-785. doi:10.1101/gad.13.7.768.

32. Kemp MG, Gaddameedhi S, Choi JH, Hu J, Sancar A. DNA repair synthesis and ligation affect the processing of excised oligonucleotides generated by human nucleotide excision repair. *J Biol Chem*. 2014;289(38):26574-26583. doi:10.1074/jbc.M114.597088.

33. Lehmann AR, McGibbon D, Stefanini M. Xeroderma pigmentosum. *Orphanet J Rare Dis*. 2011;6(1):70. doi:10.1186/1750-1172-6-70.

34. Magnaldo T, Sarasin A. Xeroderma pigmentosum: From symptoms and genetics to gene-based skin therapy. *Cells Tissues Organs*. 2004;177(3):189-198. doi:10.1159/000079993.

35. Rothkamm K, Krüger I, Thompson LH, Kru I, Lo M. Pathways of DNA Double-

Strand Break Repair during the Mammalian Cell Cycle Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. *Mol Cell Biol*. 2003;23(16):5706-5715. doi:10.1128/MCB.23.16.5706.

36. Alexander JL, Orr-Weaver TL. Replication fork instability and the consequences of fork collisions from rereplication. *Genes Dev*. 2016;30(20):2241-2252. doi:10.1101/gad.288142.116.

37. Kawashima Y, Yamaguchi N, Teshima R, et al. Detection of DNA double-strand breaks by pulsed-field gel electrophoresis. *Genes to Cells*. 2017;22(1):84-93. doi:10.1111/gtc.12457.

38. Hartlerode A, Odate S, Shim I, Brown J, Scully R. Cell cycle-dependent induction of homologous recombination by a tightly regulated I-SceI fusion protein. *PLoS One*. 2011;6(3). doi:10.1371/journal.pone.0016501.

39. Potts PR, Porteus MH, Yu H. Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. *EMBO J*. 2006;25(14):3377-3388. doi:10.1038/sj.emboj.7601218.

40. Bekker-Jensen S, Mailand N. Assembly and function of DNA double-strand break repair foci in mammalian cells. *DNA Repair (Amst)*. 2010;9(12):1219-1228. doi:10.1016/j.dnarep.2010.09.010.

41. Falck J, Forment J V, Coates J, et al. CDK targeting of NBS1 promotes DNA-end resection, replication restart and homologous recombination. *EMBO Rep*. 2012;13(6):561-568. doi:10.1038/embor.2012.58.

42. Duquette ML, Zhu Q, Taylor ER, et al. CtIP is required to initiate replication-dependent interstrand crosslink repair. *PLoS Genet*. 2012;8(11):e1003050.

doi:10.1371/journal.pgen.1003050.

43. You Z, Shi LZ, Zhu Q, et al. CtIP links DNA double-strand break sensing to resection. *Mol Cell*. 2009;36(6):954-969. doi:10.1016/j.molcel.2009.12.002.

44. Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. *Annu Rev Genet*. 2011;45:247-271. doi:10.1146/annurev-genet-110410-132435.

45. Nimonkar A V, Genschel J, Kinoshita E, et al. BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes Dev*. 2011;25(4):350-362. doi:10.1101/gad.2003811.

46. Hass CS, Lam K, Wold MS. Repair-specific functions of replication protein A. *J Biol Chem*. 2012;287(6):3908-3918. doi:10.1074/jbc.M111.287441.

47. Baumann P, West SC. Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. *Trends Biochem Sci*. 1998;23(7):247-251. doi:10.1016/S0968-0004(98)01232-8.

48. Bzymek M, Thayer NH, Oh SD, Kleckner N, Hunter N. Double holliday junctions are intermediates of DNA break repair. *Nature*. 2010;464(7290):937-941. doi:10.1038/nature08868.

49. Lee SH, Princz LN, Klï¿½gel MF, Habermann B, Pfander B, Biertï¿½mpfel C. Human holliday junction resolvase GEN1 uses a chromodomain for efficient DNA recognition and cleavage. *Elife*. 2015;4(DECEMBER2015):1-24. doi:10.7554/eLife.12256.001.

50. Bhargava R, Onyango DO, Stark JM. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet*. 2016;32(9):566-575.

doi:10.1016/j.tig.2016.06.007.

51. Lee KJ, Saha J, Sun J, et al. Phosphorylation of Ku dictates DNA double-strand break (DSB) repair pathway choice in S phase. *Nucleic Acids Res*. 2015;44(4):1732-1745. doi:10.1093/nar/gkv1499.

52. Saad H, Gallardo F, Dalvai M, Tanguy-le-Gac N, Lane D, Bystricky K. DNA dynamics during early double-strand break processing revealed by non-intrusive imaging of living cells. *PLoS Genet*. 2014;10(3):e1004187. doi:10.1371/journal.pgen.1004187.

53. Williams GJ, Hammel M, Radhakrishnan SK, Ramsden D, Lees-Miller SP, Tainer J a. Structural insights into NHEJ: Building up an integrated picture of the dynamic DSB repair super complex, one component and interaction at a time. *DNA Repair (Amst)*. 2014;17:110-120. doi:10.1016/j.dnarep.2014.02.009.

54. Moldovan G-L, Dejsuphong D, Petalcorin MIR, et al. Inhibition of homologous recombination by the PCNA-interacting protein PARI. *Mol Cell*. 2012;45(1):75-86. doi:10.1016/j.molcel.2011.11.010.

55. Kim Y, Spitz GS, Veturi U, Lach FP, Auerbach AD. Regulation of multiple DNA repair pathways by the Fanconi anemia protein. 2012;(Id):1393-1408. doi:10.1182/blood-2012-07-441212.

56. Clauson C, Schärer OD, Niedernhofer L. Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harb Perspect Biol*. 2013;5(10):a012732. doi:10.1101/cshperspect.a012732.

57. Enoiu M, Jiricny J, Schärer OD. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled

repair and translesion synthesis. *Nucleic Acids Res*. 2012;40(18):8953-8964. doi:10.1093/nar/gks670.

58. Wang W. Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nat Rev Genet*. 2007;8(10):735-748. doi:10.1038/nrg2159.

59. Knipscheer P, Räschle M, Schärer OD, Walter JC. Replication-Coupled DNA Interstrand Cross-Link Repair in Xenopus Egg Extracts. In: Bjergbæk L, ed. *DNA Repair Protocols*. Totowa, NJ: Humana Press; 2012:221-243. doi:10.1007/978-1-61779-998-3_16.

60. Williams HL, Gottesman ME, Gautier J. Replication-independent repair of DNA interstrand crosslinks. *Mol Cell*. 2012;47(1):140-147. doi:10.1016/j.molcel.2012.05.001.

61. Muniandy P a, Liu J, Majumdar A, Liu S, Seidman MM. DNA interstrand crosslink repair in mammalian cells: step by step. *Crit Rev Biochem Mol Biol*. 2010;45(1):23-49. doi:10.3109/10409230903501819.

62. Kee Y, D'Andrea A. Molecular pathogenesis and clinical management of Fanconi anemia. *J Clin Invest*. 2012;122(11):3799-3806. doi:10.1172/JCI58321.anomalies.

63. Moldovan G-L, D'Andrea AD. How the Fanconi Anemia Pathway Guards the Genome. *Annu Rev Genet*. 2009;43(1):223-249. doi:10.1146/annurev-genet-102108-134222.

64. Klein Douwel D, Boonen RACM, Long DT, et al. XPF-ERCC1 Acts in Unhooking DNA Interstrand Crosslinks in Cooperation with FANCD2 and FANCP/SLX4. *Mol Cell*. 2014;54(3):460-471. doi:10.1016/j.molcel.2014.03.015.

65. Kratz K, Schöpf B, Kaden S, et al. Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents. *Cell*. 2010;142(1):77-88. doi:10.1016/j.cell.2010.06.022.

66. Liu T, Ghosal G, Yuan J, Chen J, Huang J. FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. *Science (80-)*. 2010;329(5992):693-696. doi:10.1126/science.1192656.

67. MacKay C, Déclais AC, Lundin C, et al. Identification of KIAA1018/FAN1, a DNA Repair Nuclease Recruited to DNA Damage by Monoubiquitinated FANCD2. *Cell*. 2010;142(1):65-76. doi:10.1016/j.cell.2010.06.021.

68. Khan NE, Rosenberg PS, Alter BP. Preemptive Bone Marrow Transplantation and Event-Free Survival in Fanconi Anemia. *Biol Blood Marrow Transplant*. 2016;22(10):1888-1892. doi:10.1016/j.bbmt.2016.06.018.

69. Zhou W, Otto EA, Cluckey A, et al. FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. *Nat Genet*. 2012;44(8):910-915. doi:10.1038/ng.2347.

70. Smogorzewska A, Desetty R, Saito TT, et al. NIH Public Access. 2011;39(1):36-47. doi:10.1016/j.molcel.2010.06.023.A.

71. Williams HL, Gottesman ME, Gautier J. The differences between ICL repair during and outside of S phase. *Trends Biochem Sci*. 2013;38(8):386-393. doi:10.1016/j.tibs.2013.05.004.

72. Wang AT, Sengerová B, Cattell E, et al. Human SNM1A and XPF – ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes Dev*. 2011;25(17):1859-1870. doi:10.1101/gad.15699211.

73. Walden H, Deans AJ. The Fanconi Anemia DNA Repair Pathway: Structural and Functional Insights into a Complex Disorder. *Annu Rev Biophys*. 2014;(April):1-22. doi:10.1146/annurev-biophys-051013-022737.

74. Berwick M, Satagopan JM, Ben-porat L, et al. NIH Public Access. 2013;67(19):9591-9596. doi:10.1158/0008-5472.CAN-07-1501.Genetic.

75. Auerbach AD. FA and its diagnosis. *Mutat Res*. 2010;668:4-10. doi:10.1016/j.mrfmmm.2009.01.013.Fanconi.

76. Deans AJ, West SC. FANCM Connects the Genome Instability Disorders Bloom's Syndrome and Fanconi Anemia. *Mol Cell*. 2009;36(6):943-953. doi:10.1016/j.molcel.2009.12.006.

77. D'Andrea AD. The Fanconi Anemia and Breast Cancer Suceptibility Pathway. *Nature*. 2010;362(20):1909-1919. doi:10.1056/NEJMra0809889.The.

78. Tsutakawa SE, Classen S, Chapados BR, et al. Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell*. 2011;145(2):198-211. doi:10.1016/j.cell.2011.03.004.

79. Kathera C, Zhang J, Janardhan A, et al. Interacting partners of FEN1 and its role in the development of anticancer therapeutics. *Oncotarget*. 2017;8(16):27593-27602. doi:10.18632/oncotarget.15176.

80. Williams RS, Kunkel TA. FEN nucleases: Bind, bend, fray, cut. *Cell*. 2011;145(2):171-172. doi:10.1016/j.cell.2011.03.039.

81. Keijzers G, Bohr V, Juel Rasmussen L. Human exonuclease 1 (EXO1) activity characterization and its function on FLAP structures. *Biosci Rep*. 2015;1:1-13. doi:10.1042/BSR20150058.

82. Zheng L, Li M, Shan J, Krishnamoorthi R, Shen B. Distinct roles of two Mg2+ binding sites in regulation of murine flap endonuclease-1 activities. *Biochemistry*. 2002;41(32):10323-10331. doi:10.1021/bi025841s.

83. Lee Bi B, Nguyen LH, Barsky D, Fernandes M, Wilson DM. Molecular interactions of human Exo1 with DNA. *Nucleic Acids Res*. 2002;30(4):942-949. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=100345&tool=pmcentrez&rendertype=abstract.

84. Zheng L, Jia J, Finger LD, Guo Z, Zer C, Shen B. Functional regulation of FEN1 nuclease and its link to cancer. *Nucleic Acids Res*. 2011;39(3):781-794. doi:10.1093/nar/gkq884.

85. Parrish JZ, Yang C, Shen B, Xue D. CRN-1, a Caenorhabditis elegans FEN-1 homologue, cooperates with CPS-6/EndoG to promote apoptotic DNA degradation. *EMBO J*. 2003;22(13):3451-3460. doi:10.1093/emboj/cdg320.

86. Balakrishnan L, Bambara R a. Flap Endonuclease 1. *Annu Rev Biochem*. February 2013. doi:10.1146/annurev-biochem-072511-122603.

87. Sun H, He L, Wu H, et al. The FEN1 L209P mutation interferes with long-patch base excision repair and induces cellular transformation. *Oncogene*. 2017;36(2):194-207. doi:10.1038/onc.2016.188.

88. Miętus M, Nowak E, Jaciuk M, Kustosz P, Studnicka J, Nowotny M. Crystal structure of the catalytic core of Rad2: insights into the mechanism of substrate binding. *Nucleic Acids Res*. August 2014:1-14. doi:10.1093/nar/gku729.

89. Guo Z, Qian L, Liu R, et al. Nucleolar Localization and Dynamic Roles of Flap Endonuclease 1 in Ribosomal DNA Replication and Damage Repair □ †.

2008;28(13):4310-4319. doi:10.1128/MCB.00200-08.

90. Chen X, Paudyal SC, Chin R-I, You Z. PCNA promotes processive DNA end resection by Exo1. *Nucleic Acids Res*. 2013;41(20):9325-9338. doi:10.1093/nar/gkt672.

91. Myler LR, Gallardo IF, Zhou Y, et al. Single-molecule imaging reveals the mechanism of Exo1 regulation by single-stranded DNA binding proteins. *Proc Natl Acad Sci*. 2016;113(9):E1170-E1179. doi:10.1073/pnas.1516674113.

92. Chen H, Lisby M, Symington LS. Article RPA Coordinates DNA End Resection and Prevents Formation of DNA Hairpins. *Mol Cell*. 2013;50(4):589-600. doi:10.1016/j.molcel.2013.04.032.

93. Deng SK, Gibb B, de Almeida MJ, Greene EC, Symington LS. RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nat Struct Mol Biol*. 2014;21(4):405-412. doi:10.1038/nsmb.2786.

94. Pokhrel N, Origanti S, Davenport EP, et al. Monitoring Replication Protein A (RPA) dynamics in homologous recombination through site-specific incorporation of non-canonical amino acids. *Nucleic Acids Res*. 2017;45(16):9413-9426. doi:10.1093/nar/gkx598.

95. Nimonkar A V, Ozsoy a Z, Genschel J, Modrich P, Kowalczykowski SC. Human exonuclease 1 and BLM helicase interact to resect DNA and initiate DNA repair. *Proc Natl Acad Sci U S A*. 2008;105(44):16906-16911. doi:10.1073/pnas.0809380105.

96. Jasin M, Rothstein R. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol*. 2013;5(11):a012740.

doi:10.1101/cshperspect.a012740.

97. Sluis TVANDER, Kempinga C, Sijmons RH, Zee ATEGJVANDER, Hofstra RMW. Germline Mutations of EXO1 Gene in Patients With Hereditary. 2001:1580-1587. doi:10.1053/gast.2001.25117.

98. Rein K, Yanez DA, Terr B, et al. EXO1 is critical for embryogenesis and the DNA damage response in mice with a hypomorphic Nbs1 allele. 2018;43(15):7371-7387. doi:10.1093/nar/gkv691.

99. Chan YW, West S, Laboratories CH, Mimms S, En H. GEN1 promotes Holliday junction resolution by a coordinated nick and counter-nick mechanism. 2015;43(22):41-43. doi:10.1093/nar/gkv1207.

100. Liu T, Huang J. Quality control of homologous recombination. *Cell Mol Life Sci*. May 2014. doi:10.1007/s00018-014-1649-5.

101. Constantinou A, Gunz D, Evans E, et al. Conserved Residues of Human XPG Protein Important for Nuclease Activity and Function in Nucleotide Excision Repair *. 1999;274(9):5637-5648.

102. Henneke G, Friedrich-Heineken E, Hübscher U. Flap endonuclease 1: a novel tumour suppresser protein. *Trends Biochem Sci*. 2003;28(7):384-390. doi:10.1016/S0968-0004(03)00138-5.

103. Osawa T, Davies D, Hartley J a. Mechanism of cell death resulting from DNA interstrand cross-linking in mammalian cells. *Cell Death Dis*. 2011;2:e187. doi:10.1038/cddis.2011.70.

104. Jamieson ER, Lippard SJ. Structure, Recognition, and Processing of Cisplatin-DNA Adducts. *Chem Rev*. 1999;99(9):2467-2498.

http://www.ncbi.nlm.nih.gov/pubmed/11749487.

105. Basu A, Krishnamurthy S. Cellular responses to Cisplatin-induced DNA damage. *J Nucleic Acids*. 2010;2010. doi:10.4061/2010/201367.

106. Bunting SF, Callén E, Kozak ML, et al. BRCA1 functions independently of homologous recombination in DNA interstrand crosslink repair. *Mol Cell*. 2012;46(2):125-135. doi:10.1016/j.molcel.2012.02.015.

107. Bhatia V, Herrera-moyano E. The Role of Replication-Associated Repair Factors on R-Loops. 2017:1-12. doi:10.3390/genes8070171.

108. Paulsen RD, Sollier J, Stork CT, Cimprich KA. Short Article Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. 2014:777-785. doi:10.1016/j.molcel.2014.10.020.

109. García-rubio ML, Pérez-calero C, Barroso SI, Tumini E. The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. 2015:1-17. doi:10.1371/journal.pgen.1005674.

110. Skourti-stathaki K, Proudfoot NJ, Livingston DM, et al. BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair Article BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair. *Mol Cell*. 2015;57(4):636-647. doi:10.1016/j.molcel.2015.01.011.

111. Hoeijmakers JHJ. DNA damage, aging, and cancer. *N Engl J Med*. 2009;361(15):1475-1485. doi:10.1056/NEJMra0804615.

112. Aguilera A, Cimprich KA. promote R-loop-induced genome instability. 2015;56(6):777-785. doi:10.1016/j.molcel.2014.10.020.Transcription-coupled.

113. Yamamoto KN, Kobayashi S, Tsuda M, Kurumizaka H, Takata M, Kono K. Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. 2011. doi:10.1073/pnas.1018487108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1018487108.