

Investigating the role of homologous recombination at human centromeres

Lincoln College, University of Oxford



Xanita Saayman

Hilary Term 2022

DPhil in Molecular Cell Biology in Health and Disease

William Dunn School of Pathology

Dedication

This work would not have been possible without the unwavering support and guidance from mentors, friends and family. A special dedication to Max, without whom I would have undoubtedly continued working weekends. And to Stephen and Louretha, for everything you've done for me.

Declaration

I declare that the work presented in this thesis is my own. This thesis was written independently, except where work which has formed part of jointly-authored publications has been included.

Abstract

Centromeres are structurally essential regions of the eukaryotic genome, acting as platforms for kinetochore establishment and microtubule attachment during chromosome segregation. Despite their essential function, centromeres are implicated during genome instability, as DNA breaks and rearrangements originating at centromeres are frequently observed in human disease.

In this thesis, I first investigate the innate DNA fragility of centromeres through two independent approaches. I perform computational analyses of publicly available next-generation sequencing datasets, revealing a striking incidence of spontaneous DNA single-strand breaks at human centromeres. Through the development of a novel single-cell imaging assay to directly assess DNA breaks at repetitive regions of the genome including centromeres, I confirm this observation and further investigate potential sources of spontaneous centromere fragility. I also provide evidence that centromeric DNA breaks are surprisingly refractory to canonical DNA damage signalling mechanisms, raising important questions regarding the detection and repair of DNA breaks at centromeres.

Secondly, building on a growing body of evidence suggesting that centromeres are unusually prone to spontaneous recombination, I examined the role of the repair pathway homologous recombination (HR) at human centromeres. I demonstrate that the central mammalian recombinase RAD51 suppresses DNA break accumulation in a manner dependent on its recombination activity, providing molecular evidence for the recombination processes proposed to occur at human centromeres.

Finally, I show that RAD51 and its canonical regulator BRCA2 are intimately linked to the maintenance of centromere identity, defined by the centromere-specific epigenetic marker CENP-A, providing insight into a potential function of innate centromere fragility.

Based on the results presented in this thesis, I propose that human centromeres are innately susceptible to spontaneous DNA breaks, which are subsequently targeted for recombination-based repair. These findings shed light on several questions ranging from centromere-derived genome instability in human disease to our fundamental understanding of genome evolution.

Contents

1	Introduction	7
1.1	Homologous Recombination (HR)	7
1.1.1	Genetic recombination	7
1.1.1.1	Early models for homologous recombination	8
1.1.1.2	Identifying mediators and regulators of homologous recombination	9
1.1.1.3	Current model of homologous recombination	11
1.1.1.4	Substrates initiating homologous recombination	12
1.1.2	Physiological outcomes of homologous recombination	13
1.2	DNA damage and repair	14
1.2.1	DNA breaks and genome fragility hotspots	14
1.2.2	Sources of DNA breaks in the human genome	14
1.2.3	DNA break repair	16
1.2.3.1	Overview of DNA break repair	16
1.2.3.2	Recognition and signalling of DNA breaks	16
1.2.3.3	DNA break repair mechanisms	18
1.2.4	Cellular outcomes of DNA breaks	19
1.2.5	Detection of DNA breaks	21
1.3	Centromeres	23
1.3.1	Centromere structure and function	23
1.3.2	Centromere evolution	24
1.3.3	CENP-A in the specification of human centromeres	26
1.3.4	CENP-A chromatin dynamics	27
1.3.4.1	Spatial and temporal regulation of CENP-A deposition	27
1.3.5	Centromeres as genome instability hotspots	29
1.3.5.1	non-B-form DNA structures at centromeres	29
1.3.5.2	Atypical chromatin structure at centromeres	32
1.3.6	Homologous recombination at human centromeres	33
1.3.6.1	Evidence for constitutive homologous recombination at centromeres	33
1.3.6.2	“Centromeres convert, but don’t cross”	34
1.4	Research Aim	35
2	Materials and methods	36
2.1	Tables of materials	36
2.1.1	Mammalian cell lines	36
2.1.2	Plasmids	37
2.1.3	Oligonucleotides	38
2.1.3.1	PCR primer sequences	38
2.1.3.2	siRNA sequences	39
2.1.3.3	RT-qPCR primer sequences	39
2.1.3.4	ChIP-qPCR primer sequences	41
2.1.3.5	exo-FISH PNA probe sequences	41
2.1.4	Antibodies	42
2.1.4.1	Primary antibodies	42
2.1.4.2	Secondary antibodies	42
2.1.5	Treatment and control compounds	43
2.1.6	Accession codes	43
2.2	Cloning and bacterial cell culture	44

2.2.1	Plasmid amplification	44
2.2.2	Glycerol stocks	45
2.2.3	Construction of RAD51 variant plasmids	45
2.3	Mammalian cell culture	46
2.3.1	General cell culture	46
2.3.2	Transfections	46
2.3.3	siRNA transfections	46
2.3.4	Cell cycle regulation	46
2.4	Molecular biology techniques	47
2.4.1	Western blotting	47
2.4.2	Cell lysis and DNA extraction	48
2.4.3	Chromatin immunoprecipitation (ChIP)	48
2.4.4	Reverse transcription polymerase chain reaction (RT-PCR)	50
2.4.5	Flow cytometry	51
2.4.6	Cell survival analysis (WST-1)	51
2.5	Microscopy techniques	51
2.5.1	exo-FISH	51
2.5.2	Native immuno-cytofluorescence (ICF)	52
2.5.3	Immunofluorescence (IF)	53
2.5.4	SNAP labelling for CENP-A deposition and stability assays	54
2.6	Data analysis	54
2.6.1	Frequency analysis of arm copy-number alterations (arm-CNAs)	54
2.6.2	Short read sequencing analysis	54
2.6.3	exo-FISH, ICF and IF quantification and statistical analysis	55
3	Investigating innate DNA fragility at human centromeres	57
3.1	Introduction	57
3.1.1	Centromeres as sites of chromosomal breakage in human disease	57
3.1.2	Current understanding of centromeres as fragile sites	58
3.2	Results	59
3.2.1	Investigating spontaneous DNA fragility originating at human centromeres	59
3.2.1.1	Chromosome arm gains and losses as early drivers of carcinogenesis	59
3.2.1.2	Spontaneous DNA breaks at human centromeres	60
3.2.1.3	A novel assay for the detection of DNA breaks at repetitive elements	65
3.2.1.4	Centromere-specific chromatin in spontaneous centromeric DNA breakage	69
3.2.1.5	DNA replication in spontaneous centromeric DNA breakage	71
3.2.2	Characterizing DNA damage signaling at human centromeres	72
3.2.2.1	DNA damage signaling in response to spontaneous centromeric DNA breaks	72
3.2.2.2	DNA damage signaling capacity of centromeres	73
3.3	Discussion	76
3.3.1	Investigating innate fragility at human centromeres	76
3.3.1.1	Potential sources of centromere-derived structural abnormalities	77
3.3.1.2	A potential role for DNA SSBs in driving chromosomal rearrangements	78
3.3.1.3	The relative contribution of centromeric and pericentromeric breaks	78
3.3.1.4	Investigating sources of spontaneous centromeric DNA breaks	79
3.3.2	Investigating DNA damage responses (DDR) at human centromeres	82
3.3.2.1	Evidence for DDR suppression at human centromeres	82
3.3.2.2	Potential mechanisms and implications of centromeric DDR suppression	83
4	Investigating the role of homologous recombination (HR) in the repair of centromere breaks	85
4.1	Introduction	85
4.1.1	Homologous recombination: mechanisms and regulation	85
4.1.2	Investigating homologous recombination at human centromeres	86
4.1.3	Human artificial chromosomes (HACs) as centromere models	87
4.2	Results	88
4.2.1	HR factor localization to centromeres	88
4.2.2	RAD51-mediated protection of human centromeres	92
4.2.3	RAD51 mechanism of protection	95

4.3	Discussion	99
4.3.1	HR factor recruitment to human centromeres	99
4.3.2	HR as a repair mechanism for centromere DNA breaks	100
4.3.3	Additional repair pathways in the repair of centromeric DNA breaks	101
5	Investigating RAD51 function in centromere specification	104
5.1	Introduction	104
5.1.1	CENP-A in the specification of centromeres	104
5.1.2	Alternative pathways for CENP-A deposition	104
5.1.2.1	Transcription in the epigenetic definition of centromere identity	104
5.1.2.2	DNA repair in the epigenetic definition of centromere identity	105
5.2	Results	106
5.2.1	Reduction of CENP-A in the absence of HR recombinase RAD51	106
5.2.2	CENP-A reduction upon RAD51 depletion is independent of active DNA replication	108
5.2.3	CENP-A protein and mRNA levels in the absence of RAD51	109
5.2.4	HeLa CENP-A-SNAP to distinguish CENP-A deposition and stability	112
5.2.4.1	CENP-A-SNAP deposition in the absence of RAD51	113
5.2.4.2	CENP-A-SNAP stability in the absence of RAD51	114
5.2.5	Centromeric transcription in the absence of RAD51	115
5.2.6	Involvement of the canonical RAD51 regulator BRCA2 in centromere specification	118
5.3	Discussion	121
5.3.1	Evidence for the role of HR recombinase RAD51 in promoting CENP-A deposition	121
5.3.2	Specificity of RAD51 at centromeres in CENP-A deposition	121
5.3.3	Potential mechanisms by which RAD51 promotes CENP-A deposition	122
5.3.3.1	HR-dependent pathways linking RAD51 and CENP-A deposition	123
5.3.3.2	HR-independent pathways linking RAD51 and CENP-A deposition	124
5.3.4	Physiological significance of CENP-A reduction in the absence of RAD51-BRCA2	126
6	Discussion	128
6.1	Thesis summary and future perspectives	128
6.1.1	Spontaneous DNA fragility at human centromeres	129
6.1.2	DNA damage signalling at human centromeres	130
6.1.3	RAD51-driven recombination in the repair of centromeric DNA breaks	131
6.1.4	Proposed relationship between DNA repair factors and centromere specification . .	133
6.2	Concluding remarks	135
	Bibliography	136
	Appendices	156
6.3	Sequencing: NGS alignment pipeline [bash]	156
6.4	NGS: per-chromosome enrichment score analysis [python]	158
6.5	Automated microscopy analysis [Macro, Fiji]	160
6.6	Down-stream processing of microscopy data [Python]	165

List of Figures

1.1	Forms of genetic recombination	8
1.2	Early models of genetic recombination	10
1.3	Current model of homologous recombination	12
1.4	DNA break signaling and repair	17
1.5	Outcomes of DNA breaks and repair	20
1.6	Structure of the human centromere	24
1.7	Layered expansion model for centromere evolution	26
1.8	CENP-A chromatin dynamics	28
3.1	Mechanisms of arm-CNAs due to centromeric breakage	58
3.2	Investigating arm-CNAs as drivers of carcinogenesis	60
3.3	Detecting DNA breaks at human centromeres	62
3.4	DNA single strand breaks at human centromeres	63
3.5	Investigating the distribution of DNA breaks and CENP-A across human centromeres	64
3.6	exo-FISH for the detection of centromeric DNA breaks	65
3.7	Validation of exo-FISH	67
3.8	Characterizing spontaneous DNA breaks at centromeres	68
3.9	Investigating the contribution of CENP-A to spontaneous centromeric SSBs	70
3.10	DNA replication as a potential source for spontaneous centromeric SSBs	71
3.11	Investigating DNA damage signaling at human centromeres	73
3.12	DNA damage signaling capacity at human centromeres	75
4.1	Models of homologous recombination at repetitive regions	86
4.2	Human Artificial Chromosomes (HACs) as centromere models	88
4.3	Investigating HR factor localization to HAC centromeres	90
4.4	BRCA2 in the recruitment of RAD51 to centromeres	91
4.5	Optimization for the investigation of HR at centromeres	92
4.6	RAD51 prevents the accumulation of centromeric DNA breaks	94
4.7	Centromeric DDR activation in the absence of RAD51	95
4.8	Generation and validation of RAD51 separation-of-function (SOF) variants	96
4.9	Investigating RAD51 mechanism of centromeric protection	98
5.1	Reduction of CENP-A in the absence of HR recombinase RAD51	107
5.2	Reduction of CENP-A upon RAD51 depletion is independent of active DNA replication	109
5.3	CENP-A protein and mRNA in the absence of RAD51	110
5.4	CENP-A centromeric occupancy in HeLa CENP-A-SNAP	112
5.5	CENP-A deposition in the absence of RAD51	114
5.6	CENP-A stability in the absence of RAD51	115
5.7	Centromeric transcription in the absence of RAD51	117
5.8	Involvement of the canonical RAD51 regulator BRCA2 in centromere specification	120
5.9	Potential mechanisms by which RAD51 promotes CENP-A deposition	123
6.1	Model for RAD51-mediated recombination at spontaneous centromeric DNA breaks	129
6.2	Model for RAD51-mediated recombination in the evolution of centromeric satellite arrays	133

Chapter 1

Introduction

1.1 Homologous Recombination (HR)

1.1.1 Genetic recombination

The term genetic recombination can refer to many phenomena from the rearrangement of genotypes during sexual reproduction to the horizontal exchange of genetic information between prokaryotic organisms. In this work, genetic recombination refers to the transfer of genetic information within a single organism, either between or within chromosomes.

Genetic recombination can occur through either a reciprocal exchange of physical DNA strands, or the unidirectional transfer of genetic information from one locus to another (**Figure 1.1**). The first evidence for genetic recombination via physical exchanges of chromosome segments came as early as 1931 through seminal work on maize meiosis by Barbara McClintock and Harriet Creighton (Creighton and McClintock, 1931). Subsequent research further revealed that genetic information could also be unidirectionally transferred (i.e., without reciprocal physical exchanges) from one locus to another (Lindegren, 1955). From lower to higher eukaryotes, this unidirectional transfer of information was observed both between and within chromosomes, provided that sufficient DNA sequence homology existed between loci, across many species (Jackson and Fink, 1981; Klein and Petes, 1981; Liskay and Stachelek, 1983; Scherer and Davis, 1980). These observations collectively demanded the existence of a highly conserved mechanism driving both physical exchanges as well as unidirectional genetic transfers between homologous sequences throughout the genome.

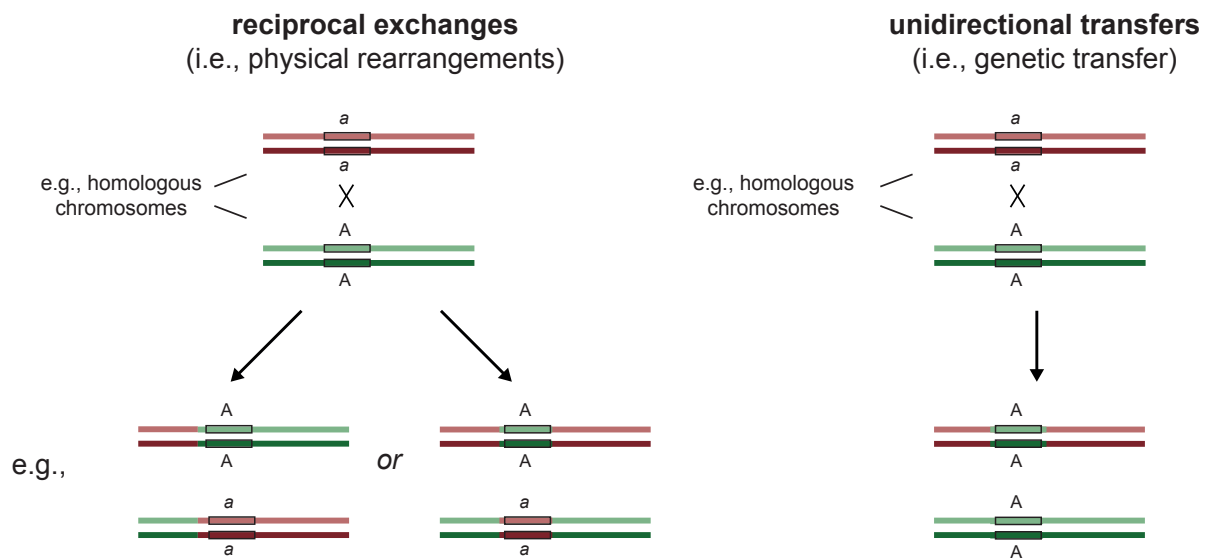


Figure 1.1: **Forms of genetic recombination.** Genetic recombination can occur through either reciprocal physical exchanges or unidirectional genetic transfers. Recombination between homologous chromosomes harboring a/A alleles is depicted here, with the outcome of each type of recombination depicted below. Importantly, recombination can occur both between chromosomes (e.g., sister chromatid) or within chromosomes, provided that sufficient homology exists between recombining loci.

1.1.1.1 Early models for homologous recombination

Models were subsequently proposed attempting to explain the molecular mechanisms by which genetic recombination could occur (**Figure 1.2**). The first model, proposed by Robin Holliday in 1964, posited that two DNA nicks on opposing sister chromatid perform reciprocal strand-invasions, forming a structural intermediate known as a Holliday junction (Holliday, 1964) (**Figure 1.2A**). The Holliday junction can migrate in either direction and be processed to form crossover or non-crossover products depending on how the junction is resolved. Meselson and Radding subsequently modified this model, while maintaining the notion of the Holliday junction, by proposing an asymmetric strand invasion following a single nick on one sister chromatid (Meselson and Radding, 1975) (**Figure 1.2B**). Upon strand invasion, a nick is generated in the recipient strand. Once again, reciprocal strand exchange forms a Holliday junction that can be processed into crossover or non-crossover products. Radding further developed the recombination model by proposing that an asymmetric strand invasion following a DNA nick can displace the recipient strand, forming what is known as a 'D-loop' structure (**Figure 1.2C**). Cleavage of the displaced strand as well as synthesis on the donor strand forms a single Holliday junction. Alternatively, he proposed that extended ssDNA gaps can form similar products following strand exchange and subsequent cleavage of the intact strand. Importantly, the asymmetric invasion posited in this model allows for the unidirectional transfer of genetic information from donor to recipient strand (Radding, 1982).

Finally, the Szostak model proposed that DNA double-stranded breaks (DSB), rather than DNA nicks, initiate recombination (**Figure 1.2D**). In this model, nuclease-mediated DSB resection exposes ssDNA that initiates strand-exchange and repair synthesis, resulting in a double Holliday junction. Differential cleavage of the double Holliday junction subsequently produces crossover and non-crossover products (Szostak *et al.*, 1983). Following the proposal of the Szostak DSB repair model, attention increasingly shifted towards DSBs as initiating substrates, arguably a result of several DSB-induced HR mechanisms observed in nature in both lower and higher eukaryotic cells (e.g., DSB-induced meiotic recombination). In addition, it was experimentally easier to induce and study DSB-initiated HR given the genetic tools available at that time (e.g., site-directed endonucleases and restriction enzymes) (Keeney, 2008; Lee *et al.*, 2015; Liang *et al.*, 1998; Rouet *et al.*, 1994; Song *et al.*, 1985).

1.1.1.2 Identifying mediators and regulators of homologous recombination

Several decades after the first model of recombination was proposed, the RecA/RAD51 family of recombinases was identified as the catalyst of strand exchange during recombination. RecA was first identified in bacteria a DNA damage-induced factor that is essential for genetic recombination, and shown to catalyse the strand exchange reaction conceptualized in the aforementioned models (Clark, 1973; Clark and Margulies, 1965; Shibata *et al.*, 1981). In eukaryotes, *S. cerevisiae* Rad51 and *H. sapiens* RAD51 were subsequently identified as RecA homologs and similarly shown to catalyse homology pairing and strand exchange (Baumann *et al.*, 1996; Benson *et al.*, 1994; Shinohara *et al.*, 1992; Sung, 1994). Highlighting the functional significance of HR, genes encoding the RecA/RAD51 recombinase family exist across all domains of life (i.e., prokaryotes, eukaryotes, archaea), although their essentiality in unperturbed conditions varies significantly (Chintapalli *et al.*, 2013). Generally, RAD51 appears to be essential for cell viability only in eukaryotic vertebrates, leading some to propose that HR becomes increasingly important as genome size and complexity increases (Muris *et al.*, 1993; Sonoda *et al.*, 1998; Symington, 2002; Tsuzuki *et al.*, 1996).

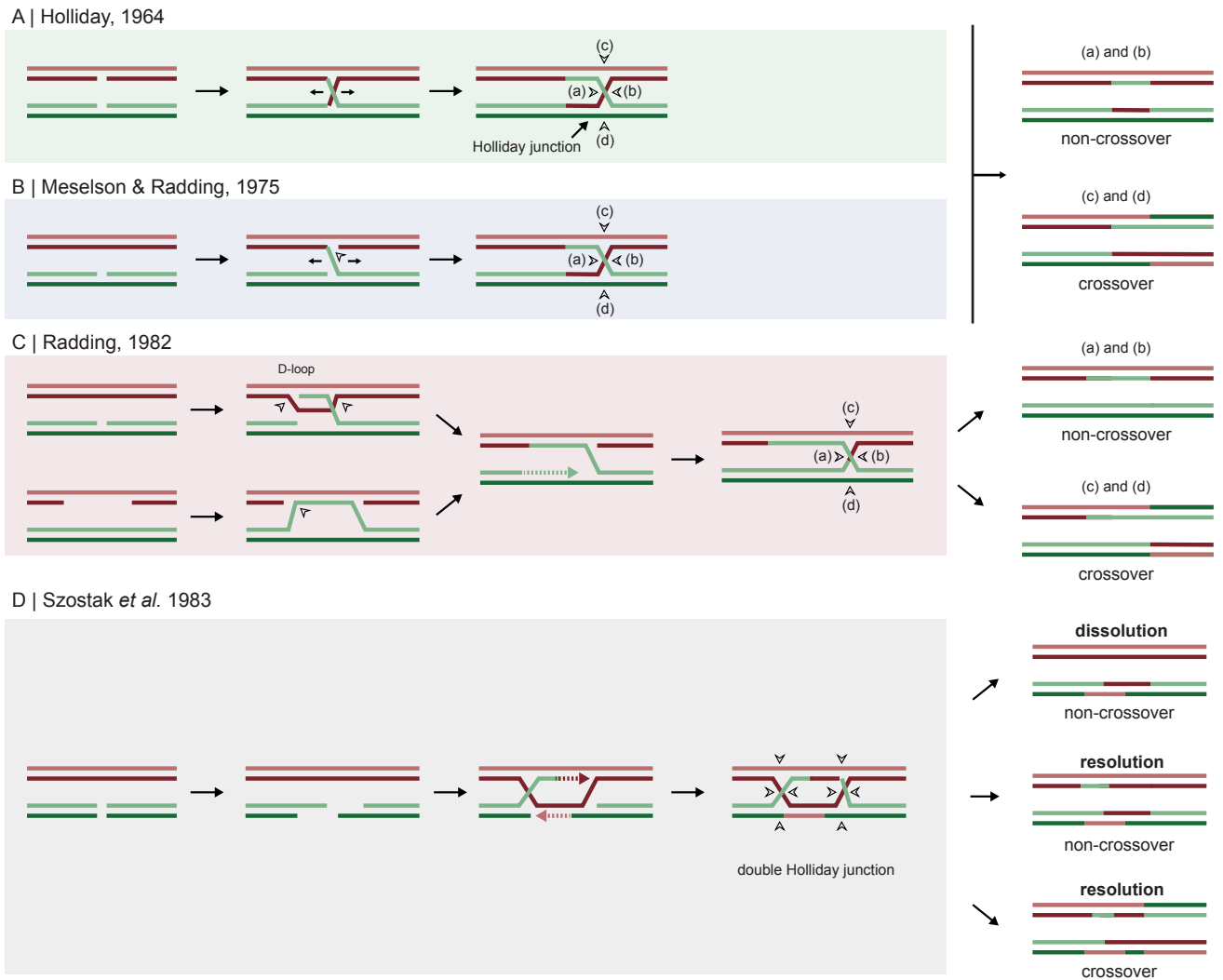


Figure 1.2: Early models of genetic recombination. (A) Holliday, 1964 model for genetic recombination whereby symmetric strand exchanges are initiated by two single-strand DNA nicks. (B) Meselson and Radding, 1975 model whereby asymmetric strand exchanges induce a nick in the recipient strand. (C) Radding, 1982 model whereby asymmetric strand exchanges displace the recipient strand to form a D-loop, enabling repair synthesis on the donor strand. Holliday, Meselson and Radding, and Radding models all entail formation of a single Holliday junction which can be processed into non-crossover or crossover products. (D) Szostak, 1983 model whereby a double-strand DNA break is resected, exposing ssDNA that can perform reciprocal strand exchanges. A double Holliday junction is formed, which can undergo either dissolution or resolution. Differential resolution produces either crossover or non-crossover products. Non-reciprocal exchanges (i.e., unidirectional genetic transfers) can form following SDSA or Holliday junction dissolution.

In canonical DSB-induced HR in mammals, both RAD51 recruitment and RAD51-mediated recombination is largely stimulated by the HR mediator complex composed of breast cancer 1 (BRCA1) and breast cancer 2 (BRCA2) as well as the partner and localiser of BRCA2 (PALB2) proteins (Jensen *et al.*, 2010; Sy *et al.*, 2009; Yuan *et al.*, 1999). However, as further discussed in **section 4.1.1**, RAD51 functional regulation appears to be an intricate and diverse process, with additional RAD51 regulators outside of

the HR mediator complex having been identified in recent years.

1.1.1.3 Current model of homologous recombination

Several decades of research following the discovery of genetic recombination have extensively characterized the molecular mechanisms of HR. Broadly, HR can be broken down into three principal phases: pre-synapsis, synapsis and post-synapsis (described here as in the Szostak model) (**Figure 1.3**). (Symington, 2002). First, single-strand DNA (ssDNA), typically generated by DSB end resection, is bound by Replication Protein A (RPA), which protects ssDNA and resolves secondary DNA structures. RAD51 is subsequently recruited to sites of DNA damage, generally assisted by the HR mediator complex (i.e., BRCA1-BRCA2-PALB2), and loaded onto ssDNA in an ATP-dependent manner, thereby replacing RPA and forming a RAD51-ssDNA filament known as a presynaptic filament.

In the synapsis phase, RAD51 performs homology search for complementary DNA sequences, resulting in a physical connection between the invading DNA substrate and the complementary duplex DNA template (D-loop). In the post-synaptic phase of the canonical DSB repair model, DNA is synthesized using the invading DNA strand as a primer for DNA synthesis as RAD51 is released through the hydrolysis of ATP (Chi *et al.*, 2006; Morrison *et al.*, 1999; Szostak *et al.*, 1983; van Mameren *et al.*, 2009). The invading strand can then be directly displaced in a mechanism known as synthesis-dependent strand-annealing (SDSA), forming non-crossover products. Alternatively, second-end capture of the invading strand can stabilize the D-loop to form a double Holliday junction, which can migrate in either direction to generate different lengths of recombination tracts. Dissolution of double Holliday junctions without cleavage can form non-crossover products. Alternatively, endonucleolytic cleavage, or resolution, of the branched Holliday junctions can produce either crossover or non-crossover products depending on the sites of cleavage.

The importance of distinguishing between HR products is highlighted when considering genome structure following HR. For example, whereas dissolution of Holliday junctions or SDSA transfers genetic information unidirectionally without the physical and reciprocal exchange of DNA, resolution of Holliday junctions produces a physical exchange of DNA. Therefore, only HJ resolution can produce large structural genome rearrangements, especially when HR occurs between non-allelic sequences, making the choice between HR outcomes critical for maintaining genome stability (more in **section 1.2.4**).

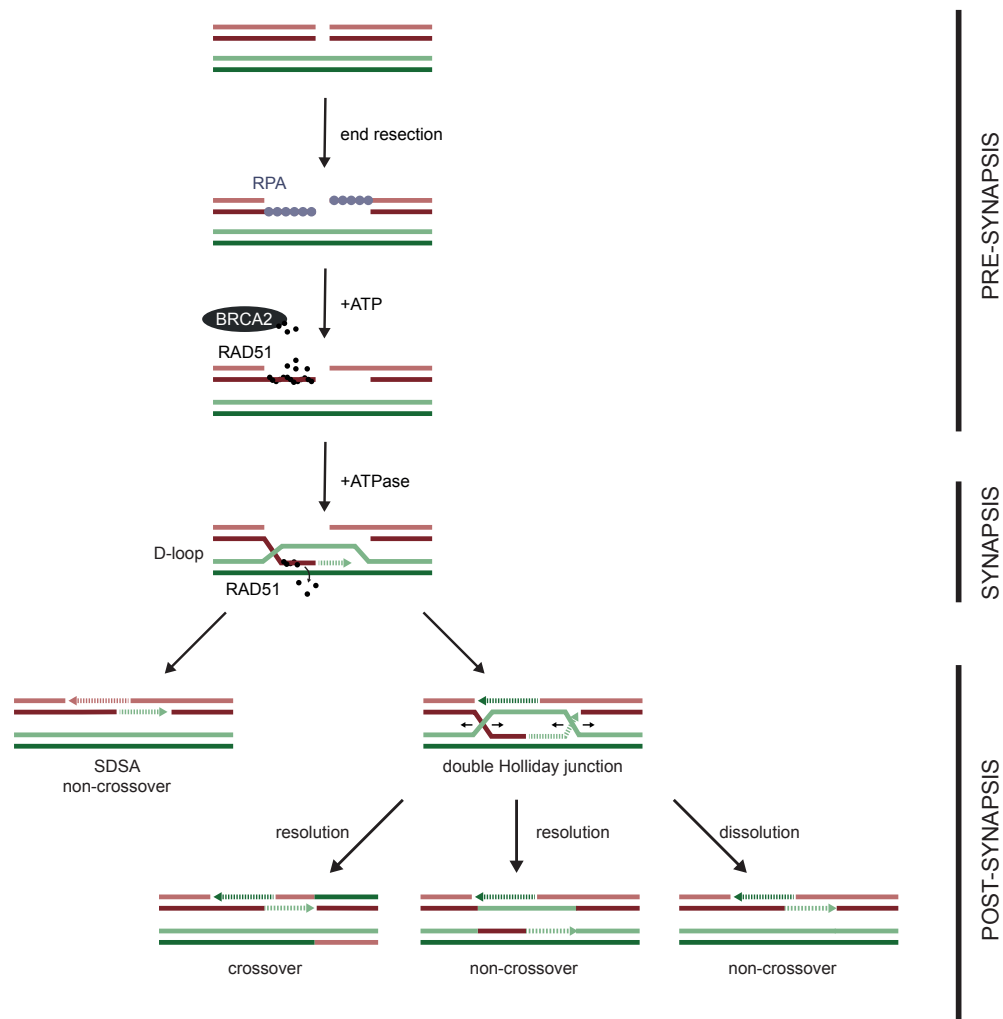


Figure 1.3: Current model of homologous recombination. The current model for homologous recombination can be broken down into three phases: pre-synapsis, synapsis, and post-synapsis. As in the Szostak 1983 model, depicted here, DNA double-strand breaks are resected to expose single-strand DNA that is bound by the protective protein RPA. BRCA2, as part of the HR mediator complex, recruits RAD51 and promotes RAD51 filament assembly to replace RPA in an ATP-dependent manner. During synapsis, RAD51 catalyzes strand exchange to form a 'D-loop' upon displacement of the recipient strand. RAD51 filaments can disassemble with ATP hydrolysis. During post-synapsis, D-loops can either be immediately displaced, resulting in synthesis-dependent strand annealing (SDSA) products, or form a double Holliday Junction (dHJ). dHJs can subsequently undergo dissolution or resolution, with the latter capable of forming both crossover-and non-crossover products.

1.1.1.4 Substrates initiating homologous recombination

Despite the wide acceptance of the Szostak DSB-induced repair model, researchers have begun reconsidering the substrate diversity during HR initiation in recent years. Advances in CRISPR-based technologies have allowed for the direct comparison between DSB-induced and nick-induced HR using canonical CRISPR or CRISPR nickases engineered through a mutated catalytic site (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012). Using these systems, researchers have determined that the rates of nick-induced HR are

comparable to those of DSB-induced HR (Mali *et al.*, 2013). Other studies have estimated that the rates of nick-induced HR are roughly 25% of that of DSB-induced HR and have also characterized several mechanistic differences between the two pathways (Davis and Maizels, 2014; Smith *et al.*, 2009). For example, one major difference is that nick-induced HR occurs more efficiently on actively transcribed strands, conceivably because transcription processes at nicked DNA could uncouple DNA strands, exposing ssDNA with available 3' ends for RAD51 binding (Davis and Maizels, 2014; Nakajima *et al.*, 2018).

1.1.2 Physiological outcomes of homologous recombination

Today, HR is recognized as a DNA repair process that has important functions in maintaining genome integrity by repairing DNA breaks, as fully discussed below (**section 1.2.3**). However, through both physical rearrangements and unidirectional genetic transfers (**Figure 1.1**), HR can have significant ramifications for genome architecture, a feature that has been evolutionarily co-opted on multiple occasions. For example, as described above, HR facilitates the reciprocal exchange of genetic information during meiosis to maintain genetic diversity within sexually dividing populations. Other examples also exist of HR rearranging genome architecture to facilitate sexual reproduction, such as during mating-type switching in *S. cerevisiae* (Haber *et al.*, 1980).

HR is also thought to impact genome architecture at repetitive elements, both for maintaining sequence homogeneity within repeated elements as well as impacting repeat array size. Recombination between DNA repeats is thought to be the major mechanism governing the evolution of multigene families (Edelman and Gally, 1970). Similarly, HR maintains sequence homogeneity between repeats such as rDNA loci, ensuring uniformity in protein transcription machinery (Baltimore, 1981; Egel, 1981; Nagylaki and Petes, 1982). While unidirectional genetic exchanges primarily function to homogenize repeat sequences, unequal reciprocal exchanges can also result in changes in repeat array sizes (Gangloff *et al.*, 1996; Szostak and Wu, 1980). Exemplifying this notion, rDNA arrays only shrink in the absence of active HR, suggesting that array expansions are recombination-dependent (Gangloff *et al.*, 1996).

Importantly, homologous recombination between DNA repeats can also be a significant source of genome instability. As further discussed in **section 1.2.4**, HR between non-allelic repeats, known as non-allelic HR (NAHR), is recognized as one of the main mechanisms leading to recurrent genome rearrangements implicated in human disease (Gu *et al.*, 2008; Liu *et al.*, 2012).

Another example of HR-induced genome instability can be observed in human repeat expansion disorders. Over 50 human disorders are a result of progressive expansion of one of the many short tandem repeats found throughout the genome. As a result, expanded repeats can interfere with protein function through a

diverse range of mechanisms, loosely categorized into loss-of-function effects (e.g., direct gene silencing) or toxic-gain-of-function effects (Malik *et al.* 2021). Repeat expansion has been proposed to occur primarily during DNA replication, either by direct slippage during replication fork progression or replication-associated DNA damage (e.g., due to secondary DNA structures preferentially forming within these repeats). In turn, DNA damage initiates repair pathways, including but not limited to HR, driving repeat instability (Khristick and Mirkin 2020). In the case of HR, non-canonical HR mechanisms such as break-induced replication seem to be primarily responsible for repeat instability (Kononenko *et al.* 2018).

Therefore, while HR plays several key roles in the maintenance of genome structure and fidelity, it must be finely regulated to prevent excessive recombination that may hinder cell survival.

1.2 DNA damage and repair

1.2.1 DNA breaks and genome fragility hotspots

The accurate preservation and faithful transmission of genetic information is arguably the single most important task of the living organism. This is no small feat, given the considerable range and extent of DNA-damaging threats experienced by any given cell. It has been estimated that human cells undergo approximately 70,000 lesions per day, of which the majority (75%) are single-stranded breaks (SSBs) and 0.03-0.06% of these are double-strand breaks (DSBs) (Lindahl and Barnes; Tubbs and Nussenzweig, 2017). While most DNA breaks are promptly repaired by multipart DNA signalling and repair pathways, unrepaired DNA breaks or even DNA repair itself can result in base mutations or structural rearrangements, ultimately leading to genome instability (as further discussed in **section 1.2.4**).

DNA lesions can result from either intrinsic or extrinsic sources. Intrinsic DNA damage is thought to occur more frequently than extrinsic DNA damage, but the types of damage incurred by either source are chemically similar (Jackson and Loeb, 2001). However, while DNA breaks from extrinsic sources tend to be more randomly distributed across the genome, DNA breaks from intrinsic sources tend to reoccur at specific genome locations. This non-random distribution indicates that there are regions of the genome that are unusually susceptible to DNA break formation, also known as genome instability hotspots.

1.2.2 Sources of DNA breaks in the human genome

Single-strand DNA breaks can refer to ssDNA nicks but can also include larger ssDNA gaps formed by, for example, stalled or collapsed DNA replication forks or resected nicks. For simplicity, the term SSBs in this work includes both ssDNA nicks and ssDNA gaps. SSBs constitute the vast majority of DNA lesions experienced per cell per day (Lindahl and Barnes, 2000) and can result from endogenous cellular

processes (e.g., DNA replication, topoisomerases, etc.) or DNA metabolic base damage (e.g., oxidative damage, oestrogens, DNA hydrolysis, etc.). Several cellular processes directly induce SSBs as requisite intermediates. For example, DNA replication of lagging strands generates DNA nicks between intermittent Okazaki fragments. In addition, endogenous structure-specific nucleases such as Topoisomerase I induce transient SSBs to relax positive- or negative- DNA supercoiling during transcription (Pommier *et al.*, 2016). Exemplifying this, SSBs are enriched at sites of active transcription such as at transcription promoters in *S. cerevisiae* and transcription termination sites in humans (Hegedüs *et al.*, 2018; Sriramachandran *et al.*, 2020). SSBs also feature prominently as pathway intermediates during the repair of various lesions including the mismatch-repair pathway (MMR), nucleotide excision repair (NER) and ribonucleotide excision repair (RER), as well as through repair-associated endonucleases targeting structures such as RNA-DNA hybrids (e.g., XPF, XPG) (Cristini *et al.*, 2019; Sassa *et al.*, 2019; Tubbs and Nussenzweig, 2017). Finally, SSBs can also be generated through the transposition of replicative DNA transposases (e.g., DDE transposases and HUH nuclease-like transposase) (Arinkin *et al.*, 2019).

SSBs can also result indirectly from both intrinsic and extrinsic sources. One of the most frequent intrinsic sources of SSBs is oxidative damage produced by metabolic reaction oxygen species (ROS), either through direct disintegration of the oxidized sugar or through the repair of damage bases (Hegde *et al.*, 2008; Pogozelski and Tullius, 1998). In addition, extrinsic sources of DNA damage can directly induce DNA breaks. For example, ionizing radiation induces both SSBs and DSBs at a ratio of 10:1 through radiolysis radicals that attack DNA backbones (Ma *et al.*, 2012; Milligan *et al.*, 1995). Exogenous perturbation of replication fork progression during DNA replication (e.g., disruption of replicative polymerases or drug-induced replication stress) can also aggravate ssDNA gaps (Feng *et al.*, 2011; Klein, 2020).

As opposed to SSBs, DSBs seem to occur relatively infrequently compared to other DNA lesions – estimates of the number of endogenous DSBs in human cells is about 50 breaks per cell per cell cycle, roughly 100-fold less than the number of endogenous SSBs (Vilenchik and Knudson, 2003). However, since SSBs maintain the integrity of a single strand, DSBs are often considered to be more genotoxic due to their potential impact on chromosomal structure. As with SSBs, DSBs can occur as requisite intermediates for physiological processes such as DNA topology management (e.g., transient DSBs induced by Topoisomerase II), genome and immune diversification, stalled replication fork restart, or transcription induction (Khan and Ali, 2017). Indirectly, DSBs can result from processes such as replication, transcription and chromatin looping, or following strand separation of unrepaired SSBs (Canela *et al.*; Hanada *et al.*; Kuzminov, 2001; Peak and Peak; Sax; Thomas and Rothstein). Like SSBs, DSBs in the human genome can also result from extrinsic sources such as ionizing radiation (i.e., X-rays and gamma rays) or ultraviolet radiation.

1.2.3 DNA break repair

1.2.3.1 Overview of DNA break repair

Given the extent of DNA damage experienced by a cell, several mechanisms for the repair of DNA SSBs and DSBs have evolved to maintain genome integrity (**Figure 1.4**). Briefly, small SSBs are rapidly repaired through the combined actions of multiple enzyme complexes, mainly coordinated by PARP1 and XRCC1, whereas larger SSBs are repaired by either translesion synthesis (TLS) or recombination-mediated template switching (TS) mechanisms. Alternatively, DSBs are repaired by either non-homologous end-joining (NHEJ) or homologous recombination (HR).

1.2.3.2 Recognition and signalling of DNA breaks

A critical first step in DNA break repair is the appropriate recognition and signalling of DNA breaks. Fundamental to the activation of the downstream DNA damage response (DDR), DNA break recognition and signalling relies on just a few upstream recognition proteins or complexes and varies according to the form of lesion.

DNA SSBs, particularly DNA nicks, are primarily detected by PARP1, which rapidly binds SSBs (**Figure 1.4A**). In a ‘global’ SSB repair mechanism proposed by Keith Caldecott, SSBs are recognized by PARP1, which modifies surrounding proteins, including histones, with chained branches of poly(ADP-ribose) (Caldecott, 2008; Okano *et al.*, 2003). poly-ADP ribosylation (known as ‘PARylation’) is thought to mimic ssDNA, in turn amplifying the signal and promoting recruitment of various ssDNA-binding DNA repair factors (Wei and Yu, 2016).

Alternatively, DSBs and longer ssDNA gaps are primarily signalled by the phosphatidylinositol-3 kinase (PIKK) family of proteins including ataxia-telangiectasia-mutated (ATM) and ataxia telangiectasia and Rad3-related (ATR), respectively (**Figure 1.4A, B**). ATM is activated in response to DNA damage following DSB recognition by the MRN complex (i.e., MRE11-RAD50-NBS1) (Uziel *et al.*, 2003). Alternatively, ATR is activated in response to extended ssDNA tracts bound by RPA, either due to DSB resection or ssDNA tracts associated with stressed replication forks.

Both ATM and ATR kinases initiate a cascade of signalling events as part of the DNA damage checkpoint, ultimately arresting cell cycle progression to allow sufficient time for the faithful repair of DNA breaks. A notable player in this cascade is the phosphorylation of a H2A.X histone variant at its conserved serine residue Ser-139, referred to as γ H2A.X (Burma *et al.*, 2001; Rogakou *et al.*, 1998; Ward and Chen, 2001; Ward *et al.*, 2004). γ H2A.X spreads to adjacent chromatin, extending up to several dozen megabases around sites of DNA damage to amplify the signal and ensure sufficient activation of the DDR (Rogakou

et al., 1999).

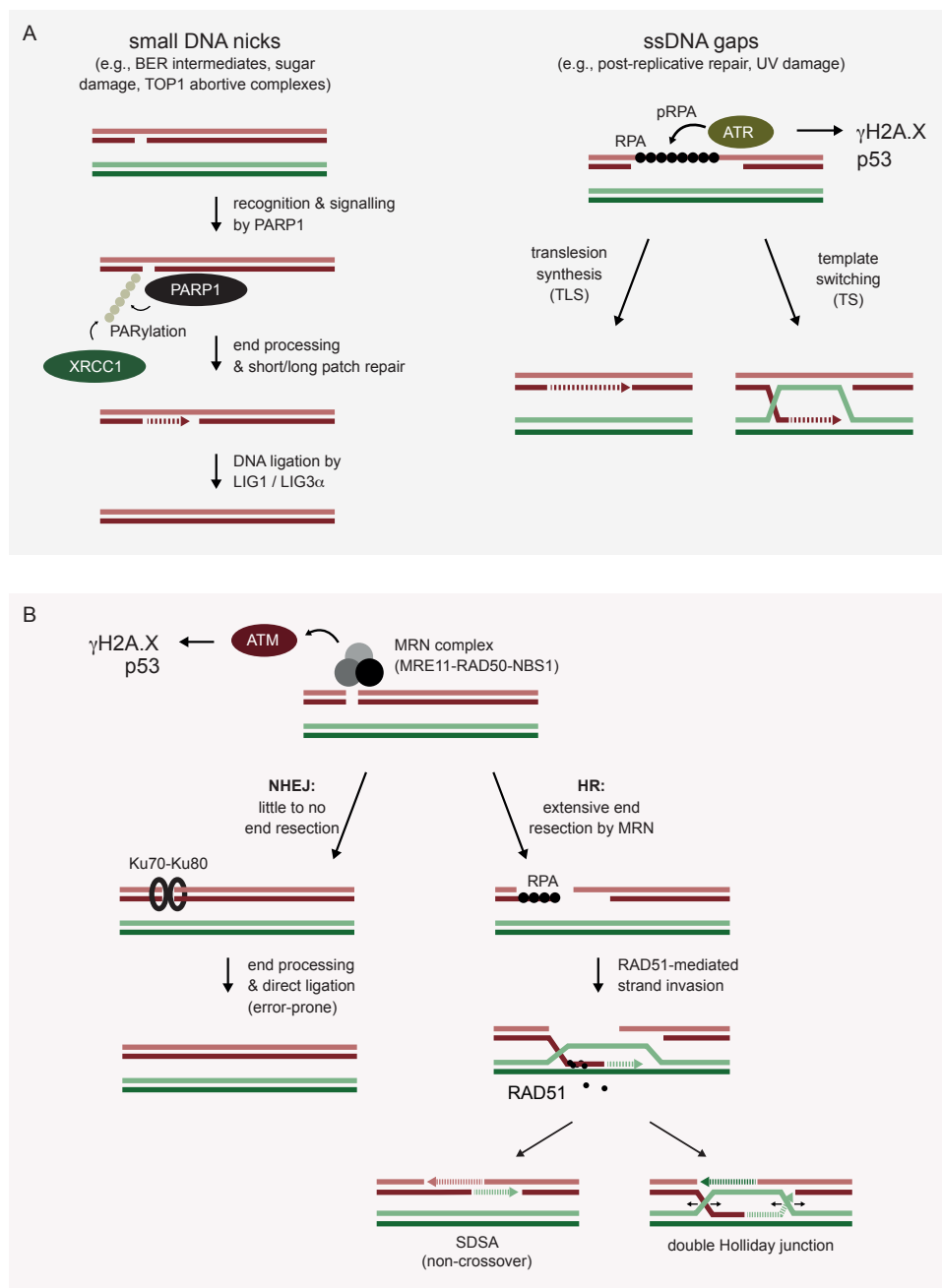


Figure 1.4: DNA break signaling and repair. (A) DNA break signaling and repair in response to small DNA nicks and larger ssDNA gaps. Small DNA nicks are detected and repaired primarily through PARP1 and XRCC1, to facilitate either short-patch and long-patch repair. Alternatively, longer ssDNA stretches coated by RPA are recognized by ATR and can be repaired by the post-replicative DNA damage tolerance pathways (i.e., translesion synthesis (TLS) or template switching (TS)). (B) Double-strand DNA breaks are detected by the MRN complex, which signals damage through activation of ATM. ATM, in turn, amplifies the signal through H2A.X phosphorylation (γ H2A.X) and p53 activation to halt cell cycle progression. DSBs are repaired by either non-homologous end-joining (NHEJ) or homologous recombination (HR).

γ H2A.X subsequently promotes the recruitment of downstream repair factors to reinforce recruitment of the MRN complex (Lukas *et al.*, 2004; Paull *et al.*, 2000; Stucki *et al.*, 2005). In addition, ATR also

phosphorylates the single-strand binding protein RPA in response to DNA damage, aiding checkpoint activation for downstream DNA repair (Byrne and Oakley, 2019; Olson *et al.*, 2006). Among the various checkpoint targets, ATM and ATR also phosphorylate p53, known as a universal guardian of genome integrity. p53 induces transcriptional changes to initiate cellular entry into senescence or apoptosis upon excessive DNA damage, thereby preventing the continued proliferation of excessively damaged cells (Surova and Zhivotovsky, 2013).

1.2.3.3 DNA break repair mechanisms

Following DNA break recognition and signalling, DNA break repair mechanisms have been extensively studied for both SSBs (i.e., DNA nicks and ssDNA gaps) as well as DSBs. As described above, SSB are detected through PARP1 recognition, followed by PARylation of local chromatin proteins. Following SSB signalling, most SSBs have damaged 3' and/or 5' terminal ends that must be processed to conventional 3'OH and 5' phosphate ends for SSB repair to successfully occur. Although DNA end processing is facilitated by a variety of enzymes (e.g., PNKP, APE1, TDP1, APTX, Pol β), XRCC1 plays a prominent role by either directly or indirectly interacting with many SSB repair enzymes. Gap filling can either be short-patch repair or long-path repair, depending on the number of nucleotides that need repairing, and the molecular machinery varies accordingly. In the final step of SSB repair, DNA ligation is orchestrated by LIG1 and LIG3 α in the nuclear genome for long- and short-patch repair, respectively (Caldecott, 2008).

An additional class of SSB repair has also been established for ssDNA gaps, for example behind stressed replication forks upon encountering obstacles such as UV-induced adducts (**Figure 1.4A**). In this context, both leading and lagging strands can re-prime downstream of the lesion to re-initiate DNA synthesis, leaving single-strand DNA gaps behind ongoing replication forks. ssDNA gaps are then repaired by either translesion synthesis (TLS) or template-switching (TS) mechanisms, collectively referred to as the DNA damage tolerance pathways (Elvers *et al.*, 2011; Lopes *et al.*, 2006; Svoboda and Vos, 1995). During TLS, repair synthesis is orchestrated by error-prone TLS polymerases which can directly synthesize DNA over lesions. Alternatively, TS involves strand exchange of the nascent strand to undamaged sister strands for error-free DNA synthesis, which can form a double Holliday junction (Bi, 2015). A direct role for the recombinase RAD51 has been demonstrated during TS at post-replicative lesions, catalysing sister chromatid exchanges and recombination (Piberger *et al.*, 2020; Tirman *et al.*, 2021). Additional support for the role of HR at replication-associated ssDNA gaps comes from recent observations that excessive ssDNA gaps, rather than the previously thought DSBs, underly the chemosensitivity of tumours deficient in HR mediator complex factors (i.e., BRCA1, BRCA2) (Cong *et al.*, 2021; Panzarino *et al.*, 2019; Tirman *et al.*, 2021). However, although ssDNA gap repair via the DNA damage tolerance pathways is often discussed in the context of replication-associated gaps, it does not appear to be exclusive to DNA

replication. In *S. cerevisiae*, for example, UV-induced lesions in non-replicating G2 cells can similarly be repaired by either TLS or TS (Ma *et al.*, 2013). Therefore, TLS- and TS- mediated ssDNA gap filling does not appear to occur exclusively in a DNA replication context.

DSBs are primarily repaired by non-homologous end-joining (NHEJ) or homologous recombination (HR), depending on the extent of DNA end processing, cell cycle stage, chromatin context, and depending on whether it is a single- or double- ended DSB (**Figure 1.4B**) (Brandsma and Gent, 2012). NHEJ is initiated by the rapid binding of the Ku heterodimer composed of Ku70 and Ku80, which activates the catalytic subunit of DNA-PKcs for rapid re-ligation of broken DNA ends (Mahaney *et al.*, 2009). Depending on the extent to end-processing, NHEJ can result in small deletions and is therefore considered to be an error-prone yet rapid DSB repair pathway. Alternatively, HR facilitates repair of DSBs while preserving the original genetic sequence, as fully described in **section 1.1.1.3**.

1.2.4 Cellular outcomes of DNA breaks

Failure to detect or repair DNA breaks can hold significant implications for cellular function and survival (**Figure 1.5A**). For example, unrepaired SSBs can block RNA polymerase progression *in vitro*, interfering with global cell function (Kathe *et al.*, 2004; Zhou and Doetsch, 1993, 1994). SSBs can also be converted to DSBs (Kuzminov, 2001) which can directly lead to chromosome fragment loss upon cell division. Excessive break detection without adequate repair can also be a source of cell toxicity. Prolonged levels of SSBs, for example, excessively activate PARP1, depleting cellular NAD^+ and ATP levels and releasing apoptosis-inducing factors from mitochondria (Heeres and Hergenrother, 2007; Moroni, 2008). Likewise, as mentioned above, excessive DNA damage-induced activation of ATM/ATR can also result in p53-induced senescence or apoptosis, directly inducing cell senescence or death (Surova and Zhivotovsky, 2013).

Counterintuitively, DNA repair itself can also be a potent source of genome instability on at least two fronts (**Figure 1.5B**). First, the repair of both DSBs and SSBs can produce DNA base mutations. Double-strand breaks can induce small base mutations or deletions if repaired by NHEJ due to end-processing prior to ligation (Lieber *et al.*, 2003). Single-strand gaps, on the other hand, can recruit TLS polymerases which are more error-prone than replicative DNA polymerases, or be targeted by mutagenic cytidine deaminases which preferentially act on ssDNA (Sale *et al.* 2013)(Roberts *et al.*, 2012). Second, DNA repair can also drastically impact genome stability by generating genome rearrangements such as inversions, insertions, and translocation.

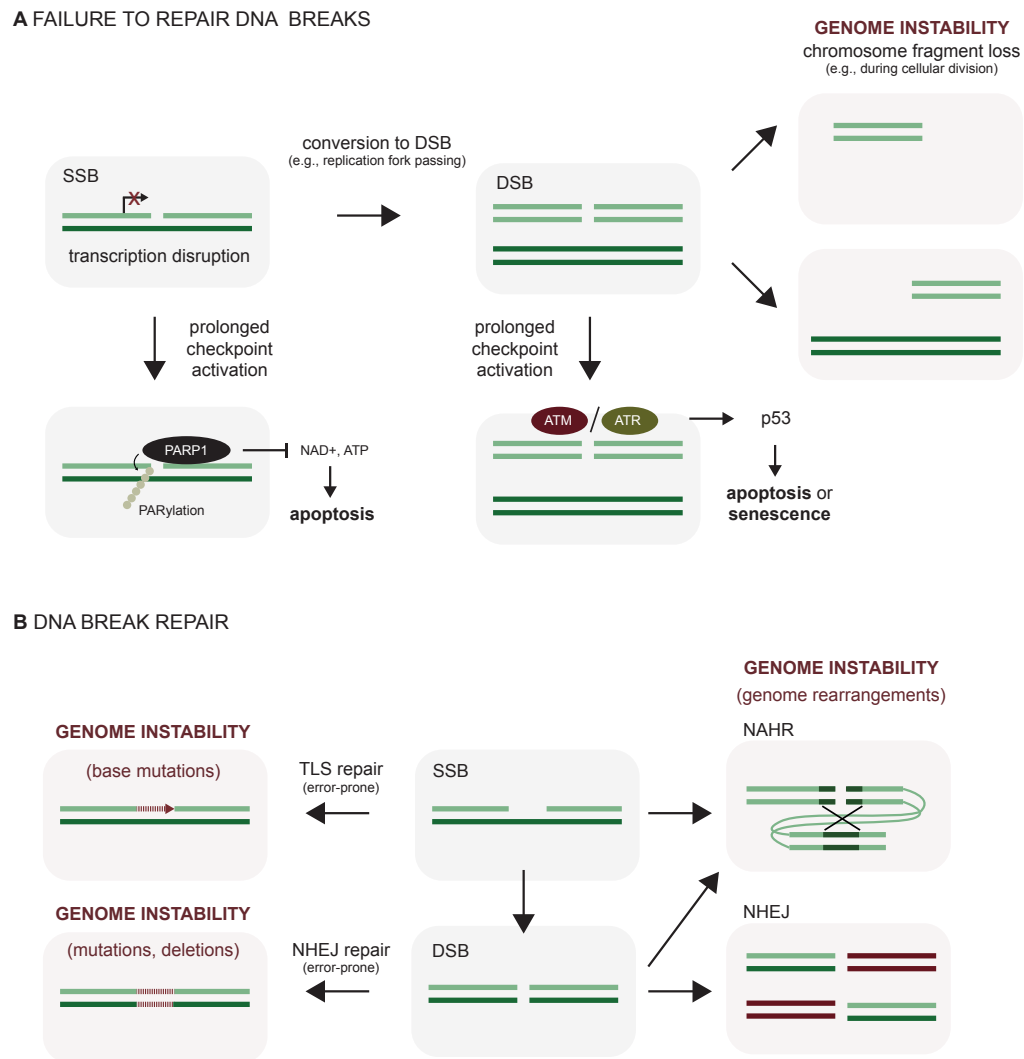


Figure 1.5: Outcomes of DNA breaks and repair. (A) Cellular outcomes upon failure to repair breaks. Local changes, such as transcriptional disruptions in response to SSBs, can hinder cell function. SSBs can also be converted to DSBs upon strand separation, such as during DNA replication. Upon cell division, unrepaired DSBs can fragment the genome and induce significant genome instability. Alternatively, detection of DNA breaks without repair can cause prolonged checkpoint activation through either PARP1 or p53, directly initiating apoptotic or senescence cascades. (B) DNA break repair itself can also act a source of genome instability. Both SSBs and DSBs can result in base mutations through error-prone repair pathways, or even structural rearrangements. Structural rearrangements can result from homologous recombination repair between non-allelic sites (NAHR), non-homologous end-joining of ectopic loci (NHEJ) or replicative mechanisms that produce complex genome rearrangements (not depicted here).

Mechanisms by which genome rearrangements arise in response to DNA break repair include non-allelic homologous recombination (NAHR), replicative mechanisms, and non-homologous end-joining (NHEJ) (Gu *et al.*, 2008; Liu *et al.*, 2012). NAHR, or recombination-based repair between non-allelic loci, was the first mechanism to be discovered and accounts for most recurrent genome rearrangements (Ou *et al.*, 2011). Such recurrent genome rearrangements are based on pre-existing genome architecture, typically originating at repetitive regions (e.g., centromeres) as well as fragile sites (e.g., FRA3B, FRA3C and

FRA3D) in cell lines with complex karyotypes (Darai-Ramqvist *et al.*, 2008). NHEJ and replicative mechanisms (e.g., fork stalling and template switching, FoSTeS) were later identified and underlie incidences of non-recurrent genome rearrangements. As described above, NHEJ-mediated repair can form small insertions, deletions, and mutations due to end processing prior to ligation. However, NHEJ-mediated repair can also produce translocations if the repair substrates are breaks originating from different genome loci. DNA replication-based mechanisms such as FoSTeS also produce non-recurrent and highly complex rearrangements and are initiated by re-priming following replication fork stalling (Lee *et al.*, 2007). Notably, NAHR and NHEJ can both be initiated by DSBs whereas NAHR can conceivably also be initiated by SSBs, as HR can be instigated by both DSBs and SSBs (**section 1.1.1.4**). The relative contribution of SSBs and DSBs to these chromosomal rearrangements is currently unknown, although DSBs are more commonly proposed to drive genome rearrangements.

Overall, the inability to faithfully repair DNA breaks can underlie a range of pathologies related to genome instability including carcinogenesis and premature aging (Boveri T; Hanseemann; Strehler). However, as described above, DNA breaks may also serve positive physiological functions such as in DNA topology management, repair pathways, and generating genetic and immune diversity through programmed DSBs (e.g., meiotic recombination and class switch recombination) (Bassing and Alt, 2004; Taccioli *et al.*, 1993). In terms of cellular responses and outcomes, the discrepancies between such ‘pathological’ and ‘physiological’ DNA breaks remain unclear.

1.2.5 Detection of DNA breaks

Our understanding of DNA breaks and their impact on cellular physiology has evolved over the last decades and relies heavily on the available techniques to experimentally detect DNA breaks. Until the advent of recent sequencing-based detection methods, research has largely depended on molecular biology techniques and microscopy techniques. These techniques can be loosely categorized into global or local detection methods. Global detection methods, capable of detecting genome-wide DNA fragmentation, is useful as an indication of total DNA damage between experimental conditions. However, locus-specific DNA detection methods are critical for our understanding of genome fragility hotspots and as readouts for more intricate experimental approaches.

Globally, DNA fragments can be detected using techniques such as pulse-field gel electrophoresis and comet assays (Grimaldi *et al.*, 2002). Pulse-field gel electrophoresis (PFGE) has been used to detect global DNA fragmentation as fragmented DNA migrate further on gels (Kawashima *et al.*, 2017). Alternatively, comet assays have been a gold-standard for the detection of both DNA single- and double-strand breaks in eukaryotic cells for many years (Collins, 2004). A spherical mass of DNA is generated by lysing agarose-embedded cells, and gel electrophoresis smears fragmented DNA. Modifications such as alkaline

denaturation or lesion-specific enzymes increase sensitivity and allow for the detection of specific types of DNA lesions. Notably, these techniques typically only provide ensemble datasets from whole cell populations.

Locally, DNA breaks can be detected by PCR or microscopy-based methods including fluorescent *in situ* hybridization (FISH) and terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) assays (Grimaldi *et al.*, 2002). With PCR-based detection methods, DNA lesions interrupt DNA amplification, thereby inversely correlating the quantity of PCR product and the number of intact templates, which can be subsequently detected on agarose gels using electrophoresis or through quantitative PCR (qPCR). FISH-based assays use fluorescent- or radio- labelled probes complementary to specific sequences and can be coupled with comet assays described above to detect DNA locus-specific DNA fragmentation. DNA breakage detection-FISH (DBD-FISH) performs an additional alkaline unwinding step on comet assays to improve FISH staining and distinguish between DNA SSBs and DSBs (Fernandez *et al.*, 2001). Similarly, the TUNEL assay and modern variants of it identify both SSBs and DSBs by detecting DNA fragmentation (Gavrieli *et al.*, 1992; Kordon *et al.*, 2020). Briefly, TdT enzymes incorporate fluorescent nucleotide analogues onto free 3'-OH of DNA strand. However, TUNEL is typically used for tissue apoptosis staining and is limited in its sensitivity and specificity. Although these assays provide single-cell and locus-specific data, these techniques typically only enable comparative assessments of DNA breaks between experimental conditions. Namely, these assays cannot be used to assess baseline levels of DNA breaks at specific loci, largely due to inherent differences in background fluorescence staining.

More commonly, molecular markers have been heavily exploited to detect locus-specific DNA breaks in single cells using the recruitment or activation of DNA damage signaling or repair proteins as a proxy for DNA breaks (Figuerola-Gonzalez and Perez-Plasencia, 2017). Common markers including Ku, γ H2A.X, XRCC1, PARP, PARylation and phosphorylated RPA. Recruitment of these markers to DNA damage loci can be detected by microscopy or even flow cytometry when assessing whole-cell activation of DNA damage signaling (i.e., as with γ H2A.X). For example, γ H2A.X has become the gold standard for an activated DNA damage response as a proxy for DNA breaks in recent years, on the grounds of it being responsive to many forms of DNA damage as well as providing spatially sensitive information (Bhagal *et al.*, 2009; Iacovoni *et al.*; Mah *et al.*, 2010; Szilard *et al.*). This approach is useful for the simultaneous assessment of locus-specific as well as global DNA damage on a single-cell level, although this approach necessitates a homogenous and robust activation of DNA damage signaling and repair pathways.

Finally, more recent advancements in sequencing technologies have enabled DNA break detection to accelerate massively in terms of both scope and precision. As recently reviewed (Bouwman and Crosetto, 2018), a range of sequencing-based technologies developed over the last decade enable the assessment of

DNA breaks genome-wide and in an unbiased fashion, leading to key advancements in our understanding of the mechanisms and distributions of DNA fragility (Saayman and Esashi, 2021). Notably, these techniques can in principle distinguish between SSBs and DSBs and provide a quantitative snapshot of DNA breaks even in baseline conditions, enabling the detection of innate DNA fragility hotspots.

1.3 Centromeres

1.3.1 Centromere structure and function

Centromeres are essential regions of the eukaryotic genome, acting as structural platforms for kinetochore establishment and microtubule attachment during both meiotic and mitotic chromosome segregation (McKinley and Cheeseman, 2016). Centromeres and flanking peri-centromeres constitute up to 6.2% of the human genome (Altemose *et al.*, 2021). With the exception of some organisms such as *S. cerevisiae*, which have point centromeres that are bound by sequence-specific centromere proteins, most plant and animal centromeres are composed of tandem repeat sequences known as satellite repeats (Fitzgerald-Hayes *et al.*, 1982; Melters *et al.*, 2013). In humans specifically, 171-bp monomeric sequences known as alpha satellites are repeated in head-to-tail tandem fashion (Willard, 1989; Wu and Manuelidis, 1980).

Human alpha satellite repeat arrays are further organized on a higher level (**Figure 1.6**). The most basic multimer structure is a dimer containing two units of alpha satellite monomers, which is found across all human chromosomes and is the most functionally active in terms of centromere protein binding (Henikoff *et al.*, 2015). Beyond dimers, higher-order repeats (HOR) can include dimers but are typically much larger. Each HOR is ~97-100% identical in sequence, whereas the constituent monomers within HORs are only ~70% identical (Rudd *et al.*, 2006; Willard, 1985; Willard and Waye, 1987). With the exception of a subset of HORs that are near-identical between chromosomes (e.g., chromosomes 13 and 21, chromosomes 14 and 22), HOR structures and sequences tend to differ between chromosomes, such that each chromosome has a unique HOR. Curiously, although each centromere contains between 1 and 9 HOR arrays, only one is functionally active in terms of inner kinetochore protein binding (Altemose *et al.*, 2021; McNulty and Sullivan, 2018). Alpha satellites not found in dimers or HORs, also referred to as monomeric alpha satellites, are located towards centromere peripheries, show higher levels of sequence divergence, and do not form functional centromeres.

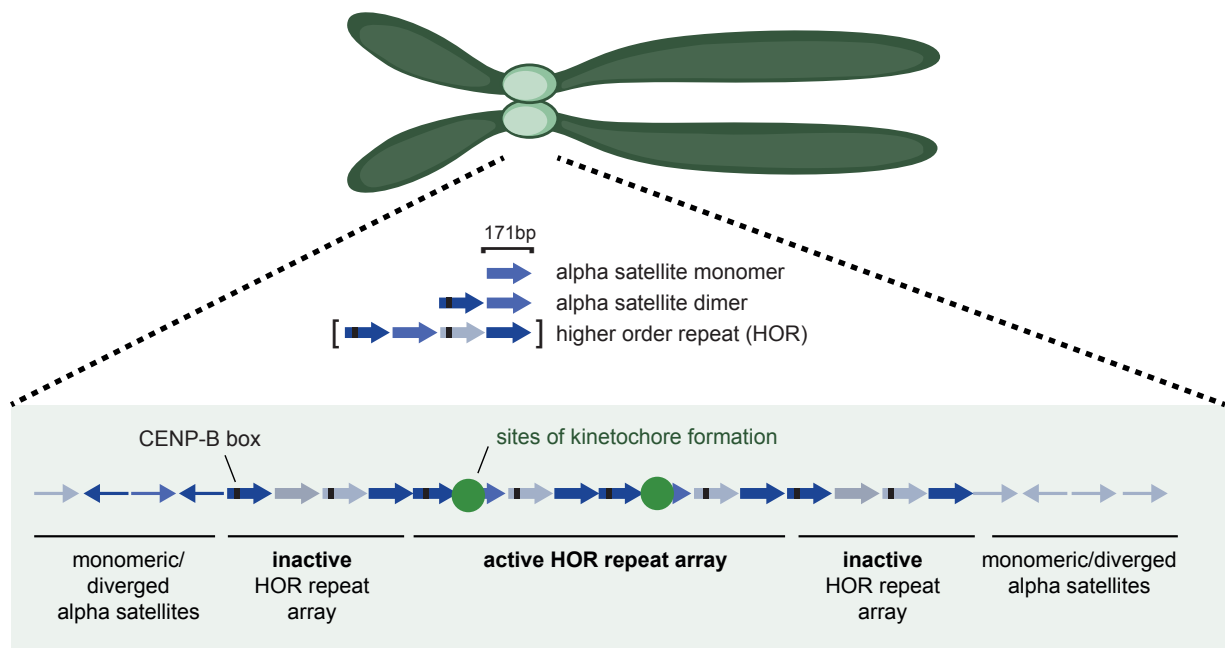


Figure 1.6: **Structure of the human centromere.** The basic unit of human centromeres is a 171-bp monomeric alpha satellite. Alpha satellites can form functionally active dimers, which may also go on to form higher-order repeats (HOR). Each centromere contains between 1 and 9 HORs, although only one is functionally active in terms of kinetochore formation and microtubule attachment. The CENP-B box is bound by the only sequence-specific centromere protein CENP-B. HORs are found within centromere cores, whereas increasingly divergent and monomeric alpha satellites are found toward centromere peripheries.

Pericentromeres flank centromeres and are composed of several different satellite repeats as well as other repeats and genes. Repeats within pericentromeres include simple 5-80bp satellites I, II and III, as well as beta and gamma satellites (Thakur *et al.*, 2021). Non-satellites regions between arrays of centromeres and pericentromeres are known as ‘centric transition’ regions, comprised of coding genes as well as long tracts of segmental duplications (Altemose *et al.*, 2021).

1.3.2 Centromere evolution

Despite their highly conserved function, centromeres are surprisingly among the most rapidly evolving regions of the genome (Henikoff *et al.*, 2001; Melters *et al.*, 2013; Miga, 2019; Warburton and Willard, 1995). The DNA sequences of the repeated monomers are highly divergent between closely related species and even between chromosomes within species, and arrays expand and contract frequently (Alexandrov *et al.*, 2001; Lee *et al.*, 2005; Talbert *et al.*, 2018; Willard, 1985). The centromere ‘paradox’ describes this dichotomy between rapidly evolving centromere DNA sequences and highly conserved centromere functionality (Henikoff *et al.*, 2001). This paradox is partially resolved by the observation that kinetochore formation and microtubule attachment are almost universally determined by the presence of the

centromere-specific histone variant CENP-A that can, to a large degree, be propagated in a sequence-independent manner (Earnshaw and Rothfield, 1985; Van Hooser *et al.*, 2001; Warburton *et al.*, 1997). As a result, evolution of the underlying sequence of centromeric repeats is likely not constrained. Nonetheless, it is clear that the repetitive structure of centromeres, especially across regional centromeres such as those of human chromosomes, is highly conserved (Melters *et al.*, 2013).

Several models have been proposed to explain the rapid evolution of satellite sequences, ranging from concerted evolution to a complex selection process known as centromere drive (Henikoff *et al.*, 2001; Malik and Bayes, 2006; Thakur *et al.*, 2021). A model known as the layered expansion model has more recently been proposed and is supported by genomic analyses of HOR evolution (Altemose *et al.*, 2021; Miga and Alexandrov, 2021) (**Figure 1.7**). The layered expansion model posits that new alpha satellite repeats periodically emerge and expand within centromeres. Upon expansion, older repeats are symmetrically displayed and become increasingly divergent with mutations, structural rearrangements, transposable element insertions, and expansion of non-alpha satellite arrays (Logsdon *et al.*, 2021). In support of this model, the inner regions of centromeres that are functionally active tend to contain the most homogenous and evolutionarily youngest alpha satellite HORs, whereas older monomeric satellite repeats become progressively less ordered with lower sequence homology towards centromere peripheries (Altemose *et al.*, 2021; Schueler *et al.*, 2005). The mechanisms driving the expansion of functionally active satellite arrays remain unclear, although early models proposed that frequent unequal crossovers could explain both array expansions as well as homogenization of satellite sequences (Smith, 1976). Curiously, the decay of sequence homogeneity within peripheries of core centromere suggests that the mechanism driving array expansion and homogenization acts primarily within functionally active regions of the centromeres.

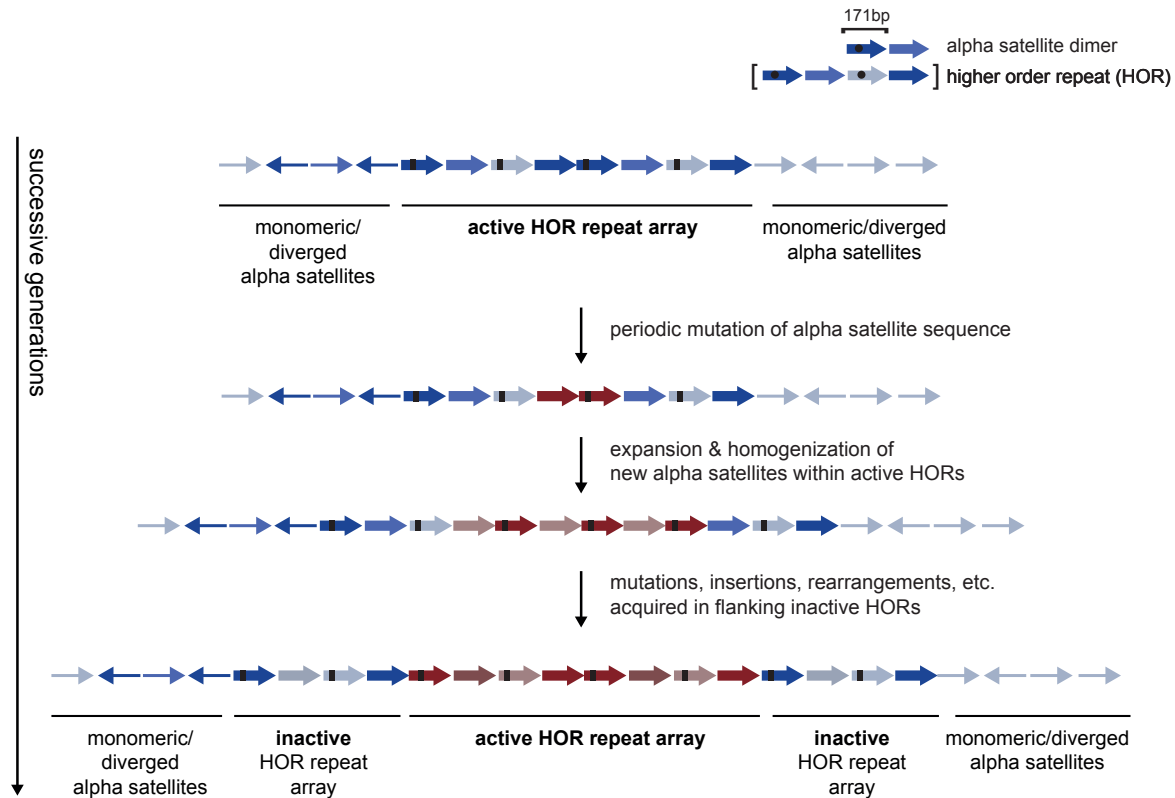


Figure 1.7: **Layered expansion model for centromere evolution.** CENP-A is contained within centromeric chromatin, interspersed between canonical H3-containing nucleosomes. Upon DNA replication, CENP-A is equally partitioned among sister chromatid. In late mitosis or early G1, CENP-A deposition to centromeric chromatin is strictly regulated both spatially and temporally. HJURP, the dedicated CENP-A chaperone, interacts with the MIS18 complex which in turn directly binds to CENP-C complexed to pre-existing CENP-A. Deposition of new CENP-A within CENP-A-containing chromatin thereby creates a self-propagating system for maintaining centromere identity in actively replicating cells.

1.3.3 CENP-A in the specification of human centromeres

In most eukaryotes, centromere function is spatially defined by a specialized nucleosomes containing the histone H3.3 variant centromere protein A (CENP-A). CENP-A was originally identified in humans as a centromere-specific variant of histone H3 (Earnshaw and Rothfield, 1985; Palmer *et al.*, 1987), and was subsequently found to direct kinetochore assembly. Indeed, targeting CENP-A to non-centromeric loci is sufficient to form functional kinetochores in a variety of organisms (Heun *et al.*, 2006; Ho *et al.*, 2014; Logsdon *et al.*, 2019; Mendiburo *et al.*, 2011; Van Hooser *et al.*, 2001). CENP-A initiates the recruitment of both inner and outer kinetochore proteins, and chromosome segregation is significantly impaired in the absence of CENP-A (Fachinetti *et al.*, 2013; Liu *et al.*, 2006; Regnier *et al.*, 2005). In accordance with the layered expansion model, CENP-A binding is primarily found within the evolutionarily youngest HORs of centromeres (Altemose *et al.*, 2021).

Centromere functionality appears to be defined by CENP-A positioning alone, largely independent of

the underlying DNA sequence. Supporting this notion, the only sequence-specific centromere binding protein is CENP-B, which binds a 17-bp motif present in one of each of the satellite array dimers described above (Earnshaw and Rothfield, 1985; Masumoto *et al.*, 1989; Muro *et al.*, 1992; Tanaka *et al.*, 2001). Although CENP-B may play a role in stabilizing centromere assemblies and promoting *de novo* centromere formation (Hoffmann *et al.*, 2016; Ohzeki *et al.*, 2002), the absence of the CENP-B box from human Y chromosomes and neocentromeres as well as the viability of CENP-B knockout mice suggest that sequence-specific centromere binding is not canonically required (Hudson *et al.*, 1998; Kapoor *et al.*, 1998; Koren *et al.*, 2010; Perez-Castro *et al.*, 1998; Tyler-Smith *et al.*, 1993).

In support of an epigenetic definition model for human centromeres, ectopic sites known as neocentromeres can become functional centromeres when native satellite centromeres are inactivated (Depinet *et al.*, 1997; Vance *et al.*, 1997; Voullaire *et al.*, 1993). Neocentromeres form outside of canonical satellite-containing DNA yet can recruit centromere and inner kinetochore proteins to orchestrate chromosome segregation. The mechanisms governing the site of neocentromerization remain poorly understood, with little common ground in chromatin features (i.e., heterochromatin levels, transcription levels) enabling neocentromerization having been identified (Scott and Sullivan, 2014).

1.3.4 CENP-A chromatin dynamics

Only a small minority of centromeric chromatin consists of CENP-A-containing nucleosomes. CENP-A comprises 0.1% of total H3 variants in the genome, and 3-4% of centromeric nucleosomes contain CENP-A instead of H3.3, resulting in an average of only ~200 CENP-A molecules per chromosome centromere (Bodor *et al.*, 2014; Nechemia-Arbely *et al.*, 2017). Following the discovery that centromere functionality is almost entirely defined by CENP-A positioning, a surge of research has been conducted to elucidate the mechanisms regulating CENP-A epigenetic inheritance.

1.3.4.1 Spatial and temporal regulation of CENP-A deposition

CENP-A-containing nucleosomes exhibit a striking stability within centromeric chromatin, perhaps due to stabilization by CENP-A-binding centromeric components such as CENP-C (Bodor *et al.*, 2013; Falk *et al.*, 2015; Jansen *et al.*, 2007). It is only upon DNA replication, when CENP-A is equally partitioned between sister chromatid, that the density of CENP-A-containing nucleosomes is effectively halved. An equivalent number of CENP-A-containing nucleosomes must therefore be deposited at centromeres following each cell cycle to ensure the propagation of centromere identity. Given that there are only ~200 CENP-A molecules per centromere, the deposition of new CENP-A within chromatin is a tightly regulated process in both spatial and temporal terms (**Figure 1.8**).

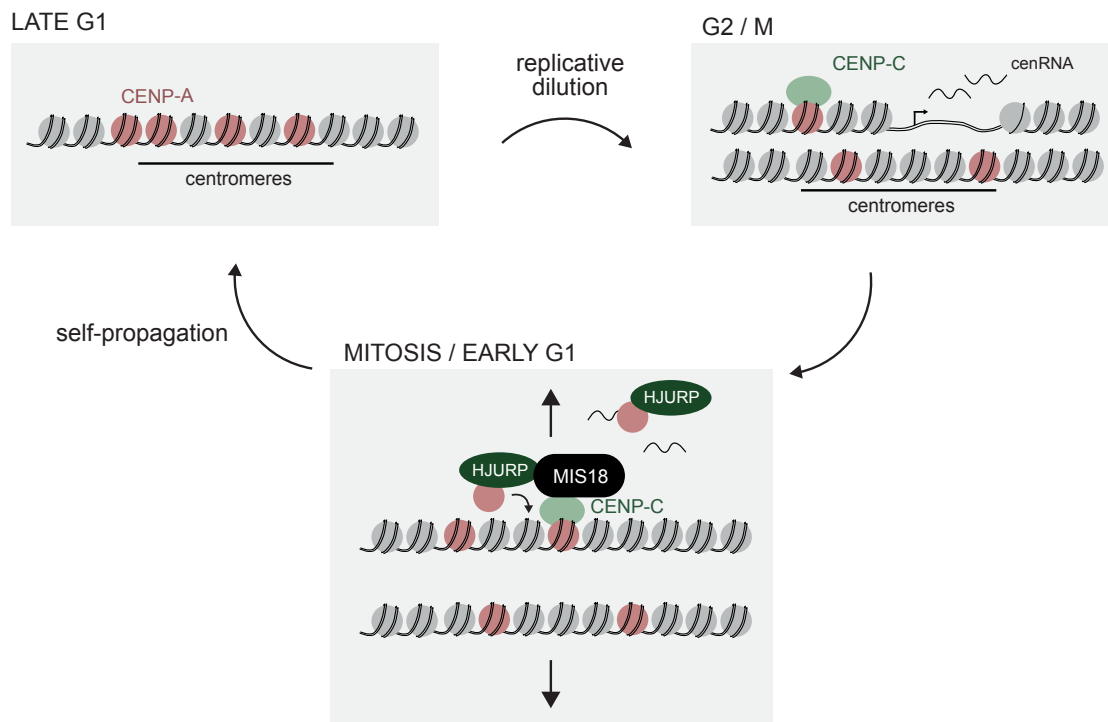


Figure 1.8: **CENP-A chromatin dynamics.** CENP-A is contained within centromeric chromatin, interspersed between canonical H3-containing nucleosomes. Upon DNA replication, CENP-A is equally partitioned among sister chromatid. In late mitosis or early G1, CENP-A deposition to centromeric chromatin is strictly regulated both spatially and temporally. HJURP, the dedicated CENP-A chaperone, interacts with the MIS18 complex which in turn directly binds to CENP-C complexed with pre-existing CENP-A. Deposition of new CENP-A within CENP-A-containing chromatin thereby creates a self-propagating system for maintaining centromere identity in actively replicating cells. Emerging evidence suggests that centromeric transcription or centromeric RNA facilitate CENP-A deposition, as depicted here. Different models have been proposed explaining this relationship, such as that centromeric transcription establishes a permissive chromatin environment for CENP-A deposition, or that centromeric non-coding RNAs complex with CENP-A and CENP-C to facilitate CENP-A deposition. However, the role for centromeric transcription in the deposition of CENP-A remains contentious.

Unlike canonical histones where deposition is coupled to DNA replication, it is not until late mitosis or early G1 when a new round of CENP-A is deposited at centromeres in mammalian cells (Jansen *et al.*, 2007; Mellone *et al.*, 2011). A dedicated histone chaperone called Holliday Junction Recognition Protein, or HJURP, stabilizes soluble CENP-A:H4 dimers and directs CENP-A deposition at centromeric chromatin (Dunleavy *et al.*, 2009; Foltz *et al.*, 2009; Shuaib *et al.*, 2010). Highlighting its functional significance, HJURP localization alone is entirely sufficient for CENP-A deposition (Barnhart *et al.*, 2011). HJURP recruitment, in turn, is facilitated by the MIS18 complex, composed of MIS18 α , MIS18 β and the MIS18-binding protein 1 (M18BP1) (Fujita *et al.*, 2007; Hayashi *et al.*, 2004; Maddox *et al.*, 2007; Wang *et al.*, 2014). The MIS18 complex is directed to centromeres through its interaction with CENP-C, which directly binds existing CENP-A (Carroll *et al.*, 2010; Moree *et al.*, 2011) (**Figure 1.8**). The spatial

regulation of CENP-A deposition is therefore orchestrated through the combined actions of HJURP and the MIS18 complex, generating a feed-forward loop in which CENP-A deposition is directed to pre-existing CENP-A-containing chromatin. Ultimately, this mechanism enables the self-propagating inheritance of centromere identity during active cell division. Temporally, the cell-cycle regulated factors PLK1 and CDK collectively prevent CENP-A deposition outside of early G1 through regulating HJURP and MIS18 complex function (McKinley and Cheeseman, 2014; Silva *et al.*, 2012).

Despite this self-propagating mechanism, the discovery of CENP-A-containing nucleosomes at non-centromeric sites suggests that CENP-A-containing chromatin cannot be the only chromatin environment permissive for CENP-A deposition (Bodor *et al.*, 2014). Indeed, the observation that neocentromeres formation in the absence of pre-existing CENP-A chromatin requires that alternative CENP-A deposition mechanisms exist. As further discussed in **section 5.1.2**, several alternative pathways for the spatial specification of CENP-A deposition have been proposed including centromeric transcription and DNA damage and repair.

1.3.5 Centromeres as genome instability hotspots

Centromeres are structurally critical to chromosome segregation but may also act as potential threats to genome instability (Barra and Fachinetti, 2018; Black and Giunta, 2018). Supporting the role of (peri)centromeric instability during carcinogenesis, several studies have identified chromosome breaks and rearrangements originating at centromeric or pericentromeric regions in human tumours (Adeyinka *et al.*, 2007; Alsop *et al.*, 2006; Balachandar *et al.*, 2007; Hermesen *et al.*, 2005; Martínez *et al.*, 2012; Smedley *et al.*, 2000). The methods employed in these studies typically lack the spatial resolution to distinguish between pericentromeric and centromeric regions, and so some ambiguity remains regarding the exact source of these rearrangements. Exemplifying this ambiguity, mapping the distribution of breaks in 26 carcinoma cell lines, Alsop *et al.* identified the highest incidence of breaks in regions surrounding centromeres, both close to centromeres as well as within centromeres (Alsop *et al.*, 2006). Regardless of this ambiguity, (peri)centromeres appear to act as potent genome instability hotspots, with DNA breaks at or near centromeres occurring frequently and with substantial consequences to genome stability.

Despite several indications that genome rearrangements frequently originate at or near centromeres, very little is known about the nature or source of DNA instability driving such rearrangements. However, several hypotheses have been proposed based on centromeric DNA structure and chromatin.

1.3.5.1 non-B-form DNA structures at centromeres

For decades there have been indications that centromeric DNA is inherently prone to forming non-B-form DNA structures. For example, *in silico* analyses of satellite DNA across eukaryotes revealed either dyad

symmetries (likely to form non-B-conformations) or DNA-bending protein binding as universal features of centromeric DNA (Kasinathan and Henikoff, 2018). *In vitro*, centromeric satellite repeats have also been shown to exhibit several types of non-B-DNA structures including DNA hairpins and i-motifs (Gallego *et al.* 1997; Nonin-Lectome and Leroy 2001; Jonstrup *et al.* 2015; Garavis *et al.* 2015; Chardon *et al.* 2022). Further, in human tumour cells, genome-wide mapping of Z-DNA, or left-handed double helical structures, revealed that 46 of 186 Z-DNA hotspots were located within centromeres (Li *et al.* 2009), confirming the propensity of centromeric DNA to form non-B-form DNA structures *in vivo*.

Until recently, however, it was unclear how frequently these non-B-form DNA structures form in healthy (i.e., non-cancerous cells) and whether they play a role in centromere functionality. A recent study investigated the prevalence of such structures and provided evidence for a physiological function of non-canonical DNA conformations at centromeres (Chardon *et al.* 2022). More specifically, DNA hairpins and loops were shown to be enriched in centromeric DNA, and intriguingly mediated by binding of CENP-B. In the absence of CENP-B, cells accumulated centromeric breaks in mitosis, suggesting that non-B-form DNA structures are both prevalent in most cells and physiologically important for protecting centromeres against mechanical stress during mitotic division (Chardon *et al.* 2022).

However, non-B-form DNA structures are typically associated with genome instability, and few studies have directly addressed the potential contribution of these centromeric non-B-form structures to centromere-derived instability. Centromeric non-B-form DNA conformations can conceivably result in DNA damage both through their interactions with molecular processes (e.g., DNA replication) as well as through structure-specific nucleases. Although little is known about how centromeric non-B-form conformations affect DNA stability, some general lessons can be derived from non-B-form DNA which promote genome instability in other regions of the genome (Bacolla *et al.*, 2006; Wang and Vasquez, 2006).

Hairpins, for example, can result in replication stalling, slippage and misalignments but can also act as structure-specific cleavage hotspots. Replication is slower at hairpin-forming DNA sequences (Lindsey and Leach, 1989), and DNA polymerases alpha, delta and epsilon stall at hairpin structures (Kamath-Loeb *et al.*, 2001). However, hairpin structures such as CTG repeats also show instability in non-proliferating tissues (Monckton *et al.*, 1999) and non-dividing cells (Kennedy and Shelbourne, 2000), providing evidence for replication-independent sources of hairpin instability.

Independently of replication, there are several structure-specific nucleases that can recognize hairpins and directly initiate DNA breaks. Topoisomerase II, for example, has site-specific cleavage activity at DNA hairpins (Froelich-Ammon *et al.*, 1994). In addition, *S. cerevisiae* Rad50 and Mre11 also show cleaving activity at hairpin structures when not annealed to a complementary DNA strand (Trujillo and Sung,

2001). Finally, the DNA mismatch repair (MMR) enzymes can recognize mismatches within imperfectly paired hairpins, although MSH2 catalytic activity may be impaired in these settings (Owen *et al.*, 2005).

Specifically at centromeres, studies have shown that non-B-form DNA conformations may similarly act to both impede replication fork progression as well as act as endonucleolytic targets. Evidence for the role of secondary structures perturbing replication fork progression comes from Li *et al.* In this study, researchers demonstrate that the structure-specific nuclease DNA2 binds centromeric DNA and cleaves centromeric hairpins and DNA loops *in vitro* (Budd *et al.*, 1995; Lee *et al.*, 2013; Li *et al.*, 2018; Lin *et al.*, 2013). In the absence of DNA2, cells display reduced centromeric replication fork progression and mitotic errors, presumably through the accumulation of DNA secondary structures at centromeres. Therefore, an accumulation of DNA secondary structures has the potential to induce replication fork stalling or even collapse, resulting in DNA breaks and mitotic errors. Notably, the data from Li *et al.* also highlight that DNA secondary structures can be directly targeted by enzymes such as DNA2 as another potential source of DNA fragility at centromeres.

Along with these findings, the observation that heterochromatic regions of human centromeres are late replicating has led to the proposal that certain portions of centromeres are inherently difficult to replicate, rendering them prone to breakage (Erliandri *et al.*, 2014; Ten Hagen *et al.*, 1990). However, it remains unclear whether this delayed replication is due to inherent difficulties in DNA replication. Indeed, it is equally plausible that secondary DNA structures render centromeres prone to damage only when the replication program is perturbed through genetic- or drug-induced means. In support of this notion, Crosetto *et al.* 2013 performed genome-wide DSB mapping (BLESS) under drug-induced replication stress conditions. They identified regions that were particularly prone to DSB formation upon replication stress treatment and showed that alpha satellites were enriched within these regions (Crosetto *et al.*, 2013). Therefore, centromeres may be some of the first regions of the genome to experience breaks in response to replication stress. However, in *D. melanogaster* and *S. cerevisiae*, centromeres replicate in early or mid S-phase, suggesting that late replication is not a universal feature of centromeres (Ahmad and Henikoff, 2001; McCarroll and Fangman, 1988). In addition, it appears that centromeric DNA in *Xenopus laevis* does not have inherently altered replication patterns from non-centromeric DNA (Aze *et al.*, 2016). In this study, the researchers also found that centromere DNA is spontaneously enriched for DNA repair factors such as MRE11/RAD50, Ku80, PARP1 and the MSH2-6 complex, suggesting that additional repair systems may be in place to ensure that centromere replication patterns aren't perturbed in physiological conditions.

In addition to non-B-form DNA, there is also a notable enrichment of other non-canonical structures such as RNA-DNA hybrids both at human and maize centromeres (Kabeche *et al.*, 2018; Liu *et al.*, 2021).

While native RNA-DNA hybrids are transient and appear to be mostly innocuous, the detection of RNA-DNA hybrids by microscopy at mitotic centromeres in unperturbed conditions suggests unusually high levels or innate stabilization of centromeric RNA-DNA hybrids. During replication, an accumulation of RNA-DNA hybrids can physically impede replication fork progression, resulting in a stalled fork or incomplete replication. Independently of replication, RNA-DNA hybrids may also promote the formation of single-stranded breaks (SSBs) or even DSBs by endonucleolytic cleavage on the exposed single-stranded DNA (Gan *et al.* 2011; Lindahl 1993; Promonet *et al.*, 2020; Sollier *et al.* 2014; Tuduri *et al.* 2009). In support of RNA-DNA hybrids as sources of genome instability at centromeres, Giunta *et al.* found that depletion of CENP-A throughout synchronized S-phase populations resulted in an accumulation of RNA-DNA hybrids at centromeres, which subsequently promoted sister-chromatid exchanges between centromeres (Giunta *et al.*, 2021). Although this does not indicate whether spontaneously formed RNA-DNA hybrids can pose any threat to centromere stability, it is evident that excessive RNA-DNA hybrids can indeed be genotoxic within centromere regions.

1.3.5.2 Atypical chromatin structure at centromeres

Centromeres act as structural platforms on which kinetochore structures assemble, as designated by CENP-A localization (Barnhart *et al.*, 2011; Hori *et al.*, 2013; Mendiburo *et al.*, 2011). Although centromeric chromatin only contains relatively few CENP-A-containing nucleosomes interspersed with canonical H3 histones, the chromatin structure at centromeres remains unique. In yeast, centromeric chromatin exhibits non-canonical structures (Bloom and Carbon, 1982; Takahashi *et al.*, 1992) and in humans, DNA is less tightly wrapped around CENP-A-containing nucleosomes within both natural centromeres and neocentromeres (Hasson *et al.*, 2013). A recent cryo-EM study reconstituted tri-nucleosomes containing a central CENP-A-containing nucleosome surrounded by two canonical H3 nucleosomes and found that CENP-A nucleosomes adopt an untwisted architecture within the tri-nucleosome, creating an outward facing DNA path and exposing CENP-A-containing chromatin (Takizawa *et al.*, 2020). This atypical chromatin structure may render centromeric DNA more susceptible to endonucleolytic attack or interfere with processes such as transcription or DNA replication.

The presence of CENP-A further initiates the recruitment of hundreds of proteins as part of the functional kinetochore plate (Barnhart *et al.*, 2011; Hori *et al.*, 2013; Mendiburo *et al.*, 2011). Massive centromeric protein complexes may theoretically produce mechanical stresses not experienced by the rest of the genome, rendering centromeres susceptible to breakage. Indeed, some preliminary work provides supports a role for mechanical force generated during chromosome segregation in causing centromeric DNA damage. In Guerrero *et al.*, researchers showed that spontaneously formed micronuclei were generated in mitosis and had markers of DNA damage colocalizing with centromeres, suggesting that the mechanical force

during chromosome segregation results in DNA damage (Guerrero *et al.*, 2010).

1.3.6 Homologous recombination at human centromeres

As described above, centromeres and pericentromeres appear to frequently break during carcinogenesis, although the source of this instability remains unclear. Given the atypical DNA and chromatin structure at centromeres, this observation raises important questions as to how DNA integrity is maintained at these structurally unique and highly repetitive regions. As discussed below, several lines of evidence indicate an active role for homologous recombination as a constitutive repair pathway at centromeres, suggesting that HR may have an essential function in protecting human centromeres from spontaneous DNA damage. Indeed, a recent report revealed that CRISPR-induced DNA damage at human centromeres was sufficient to activate RAD51-mediated repair (Yilmaz *et al.*, 2021), but it remains unclear whether HR repairs centromere DNA breaks in physiological settings (i.e., in normally dividing cells).

1.3.6.1 Evidence for constitutive homologous recombination at centromeres

Centromeres have been designated recombination ‘cold-spots’ for many decades due to a paucity of observable chromosome exchanges during meiotic recombination (Beadle, 1932; Mather, 1939). However, recent evidence suggests that recombination may in fact occur relatively frequently in mitotic cells (i.e., in the absence of programmed DSBs during meiosis). As described above, HR intermediates can be differentially processed (**Figure 1.3**), with reciprocal exchanges between sister chromatids being referred to as sister chromatid exchanges (SCEs). Alternatively, during SDSA and dissolution of Holliday junctions, the use of homologous sequences as repair templates without reciprocal exchange leads to non-reciprocal exchanges of genetic information. Evidence for recombination occurring at centromeres comes from both the observation of frequent SCEs as well as frequent non-reciprocal exchanges of genetic information through genetic studies.

First, it has been reported that centromeres undergo frequent SCEs during mitotic recombination (Giunta and Funabiki, 2017; Jaco *et al.*, 2008). This observation is primarily based on the chromosome-orientation fluorescent in situ hybridization (CO-FISH) technique. Briefly, metaphase spreads are made using mitotic cells and newly replicated strands are degraded, followed by hybridization with centromere specific PNA probes to differentially label sister chromatid and detect SCEs. Using CO-FISH, rates of SCEs at murine centromeres are estimated to be ~175-fold more frequent than on chromosome arms and ~27-fold more frequent than at telomeres (Jaco *et al.*, 2008). A similarly high rate of centromeric SCEs is observed in human cell lines. Cancer cells (HeLa, MCF7 and HCT116) have up to ~20%+ of mitotic chromosomes exhibiting centromeric SCEs (Giunta and Funabiki, 2017). Although this level is notably lower in non-cancerous hTERT-RPE1 cells (~5%), centromeric SCEs still appear to be a frequent occurrence even in

unperturbed growth conditions. Given that SCEs are only one of two major HR outcomes, this rate is probably an under-representation of the actual rate of recombination occurring at centromeres. As further described below (**section 1.3.6.2**), there is also evidence suggesting that SCEs are actively suppressed at centromeres (Onaka *et al.*, 2016; Zafar *et al.*, 2017), supporting the idea that SCE rates underestimate the actual rate of recombination.

Second, as described above, centromeres are among the most rapidly evolving regions of the genome, with alpha satellite sequences substantially diverging between closely related species and even between chromosomes within species (Alexandrov *et al.*, 2001; Balzano and Giunta, 2020; Talbert *et al.*, 2018; Willard, 1985). Across the many models explaining the rapid evolution of alpha satellite arrays, one common feature is their assumption that centromere repeat arrays undergo frequent expansions and contractions as well as homogenization, whereby a nucleotide mutation event in one array can quickly spread to surrounding arrays. Centromere expansions and contractions, as well as homogenization, can be explained by repeated homology-dependent exchange events driven by homologous recombination (Balzano and Giunta, 2020; Smith, 1976). Therefore, centromere evolutionary trajectories support the notion of active HR at centromeres, without necessarily involving SCEs. At least two studies provide some experimental evidence for non-SCE HR events between centromeric repeats. First, mapping non-reciprocal genetic exchanges using centromeric markers in maize led to the observation that non-reciprocal genetic exchanges occur at high frequencies within centromere cores, despite a total suppression of crossover events (Shi *et al.*, 2010). Second, studies in fission yeast revealed that Rad51 promotes non-crossover recombination events between inverted repeats at centromere peripheries (Onaka *et al.*, 2016).

1.3.6.2 “Centromeres convert, but don’t cross”.

An important distinction should be made between centromeric recombination rates (i.e., including both crossover and non-crossover reactions) and centromeric crossover rates, as these appear to be two independently regulated mechanisms. Supporting evidence for this notion comes from work done in *S. pombe*, which harbour inverted repeat sequences flanking centromeres (Onaka *et al.*, 2016; Zafar *et al.*, 2017). In *S. pombe*, the rate of crossover events between inverted centromeric repeats is notably lower than that between chromosome arms, despite a high rate of total recombination (i.e., non-crossover events). Therefore, the authors propose that crossovers are suppressed specifically at centromeres. The authors provide evidence that centromere proteins Mhf1/CENP-S and Mfh2/CENP-X suppress crossover formation without affecting non-crossover recombination rates, highlighting the dissimilitude between centromeric recombination and centromeric crossover suppression mechanisms. Importantly, the disruption of crossover suppression in the absence of Mhf1/CENP-S and Mfh2/CENP-X resulted in an increase in centromeric gross chromosomal rearrangements, suggesting that crossover suppression may be in place to

prevent whole-arm translocations due to illegitimate or excessive crossovers. During *S. cerevisiae* meiosis, the conserved Ctf19/CCAN kinetochore subcomplex was also shown to suppress crossover formation independent of DNA break incidence (Vincenten *et al.*, 2015).

In humans, depletion of CENP-A, CENP-C, CENP-T and CENP-W similarly resulted in increased crossover exchanges (i.e., sister-chromatid exchanges), as detected by CO-FISH (Giunta and Funabiki, 2017). It remains unclear whether the depletion of these factors results in an increase in recombination rates due to elevated levels of DNA damage or due to elevated downstream crossover rates. Following on this finding, the researchers established that concomitant with the increase in centromeric SCEs in the absence of CENP-A, frequent whole-arm translocations were observed (Giunta *et al.*, 2021), once again implicating excessive centromeric SCEs with centromeric gross chromosomal rearrangements.

Collectively, recombination appears to occur frequently at centromere repeats, of which a small indeterminate fraction of events results in crossovers. Crossover suppression at centromeres may play a critical role in the prevention of chromosomal rearrangements. However, many outstanding questions remain regarding the molecular players involved in both centromeric recombination and crossover suppression.

1.4 Research Aim

Centromeres and/or the surrounding pericentromeres can act as potent sources of genomic instability and are thereby frequently associated with human disease (**section 1.3.5**). However, our understanding of both the source and nature of (peri)centromeric fragility in humans remains remarkably poor.

In this work, I set out to investigate centromeres as potential genome instability hotspots, including both how DNA instability is manifested at centromeres as well as the protective mechanisms in place to protect such fragile regions of the genome. In **Chapter 3**, I begin by investigating the prevalence of DNA damage as well as DNA damage signaling at human centromeres in unperturbed human cells. With this objective in mind, I develop and validate a new experimental technique with which to detect DNA damage at repetitive regions, including centromeres. In **Chapter 4**, I begin to address key questions regarding how centromeres are protected from DNA damage. Given previous indications that centromeres undergo frequent recombination (**section 1.3.6**), I specifically investigate the role of the repair pathway homologous recombination through the central mammalian recombinase RAD51. Finally, in **Chapter 5**, I follow up on observations of a peculiar connection between DNA damage repair proteins and centromere specification and explore the potential biological significance of this relationship.

Chapter 2

Materials and methods

2.1 Tables of materials

2.1.1 Mammalian cell lines

Table 2.1: Mammalian cell lines

Name	Origin	Growth Conditions
hTERT-RPE1	ATCC	1:1 Dulbecco Modified Eagle Medium / Nutrient Mixture F12, 10% fetal bovine serum (FBS), 1% penicillin-streptomycin (15140122, Life Technologies), 0.123% sodium bicarbonate (S8761, Sigma-Aldrich)
hTERT-RPE1 Flp-In T-REx	John Pines	1:1 Dulbecco Modified Eagle Medium / Nutrient Mixture F12, 10% FBS, 1% penicillin-streptomycin, 0.123% sodium bicarbonate, 100 µg/ml Zeocin (ant-zn-1, InvivoGen), 10 µg/ml Blasticidin (ant-bl-1, InvivoGen)
hTERT-RPE1 Flp-In T-REx FLAG	Section 2.3.2	1:1 Dulbecco Modified Eagle Medium / Nutrient Mixture F12, 10% FBS, 1% penicillin-streptomycin, 0.123% sodium bicarbonate, 200 µg/ml G418 (G8168, Sigma), 10 µg/ml Blasticidin
hTERT-RPE1 Flp-In T-REx FLAG-RAD51-WT	Section 2.3.2	1:1 Dulbecco Modified Eagle Medium / Nutrient Mixture F12, 10% FBS, 1% penicillin-streptomycin, 0.123% sodium bicarbonate, 200 µg/ml G418 (G8168, Sigma), 10 µg/ml Blasticidin

Continued on next page

Table 2.1: Mammalian cell lines (Continued)

Name	Origin	Growth Conditions
hTERT-RPE1 Flp-In T-REx FLAG-RAD51- K133R	Section 2.3.2	1:1 Dulbecco Modified Eagle Medium / Nutrient Mixture F12, 10% FBS, 1% penicillin-streptomycin, 0.123% sodium bicarbonate, 200 µg/ml G418 (G8168, Sigma), 10 µg/ml Blasticidin
hTERT-RPE1 Flp-In T-REx FLAG-RAD51- T131P	Section 2.3.2	1:1 Dulbecco Modified Eagle Medium / Nutrient Mixture F12, 10% FBS, 1% penicillin-streptomycin, 0.123% sodium bicarbonate, 200 µg/ml G418 (G8168, Sigma), 10 µg/ml Blasticidin
hTERT-RPE1 Flp-In T-REx FLAG-RAD51-II3A	Section 2.3.2	1:1 Dulbecco Modified Eagle Medium / Nutrient Mixture F12, 10% FBS, 1% penicillin-streptomycin, 0.123% sodium bicarbonate, 200 µg/ml G418 (G8168, Sigma), 10 µg/ml Blasticidin
HeLa CENP-A-SNAP	Lars Jansen	Dulbecco Modified Eagle Medium supplemented, 10% FBS, 1% penicillin-streptomycin, 5 µg/ml Blasticidin

2.1.2 Plasmids

Table 2.2: Plasmids

Name	Purpose and construction
pcDNA 3.1+	PCR amplification of NeomycinR cassette for switching resistance cassette from hygR to neoR in pDEST-FLAG_FRT_TO plasmid series.
pOG44	Co-transfection into Flp-In T-REx cell lines for integration of pDEST-FLAG_FRT_TO (neoR) plasmids expressing RAD51 variants.
pDEST-FLAG_FRT_TO	Starting plasmid for generation of pDEST-FLAG_FRT_TO (neoR) RAD51 separation-of-function plasmid series.
pDEST- FLAG_FRT_TO(neoR)	Negative control for hTERT-RPE1 RAD51 separation-of-function variants, previously generated by lab (Wassing <i>et al.</i> , 2020), resistance cassette switched from hygR to neoR (section 2.2.3).

Continued on next page

Table 2.2: Plasmids (Continued)

Name	Purpose and construction
pDEST- FLAG_FRT_TO(neoR) RAD51-WT	Construction of hTERT-RPE1 RAD51 separation-of-function variants, previously generated by lab (Wassing <i>et al.</i> , 2020), resistance cassette switched from hygR to neoR (section 2.2.3).
pDEST- FLAG_FRT_TO(neoR) RAD51-K133R	Construction of hTERT-RPE1 RAD51 separation-of-function variants, previously generated by lab (Wassing <i>et al.</i> , 2020), resistance cassette switched from hygR to neoR (section 2.2.3).
pDEST- FLAG_FRT_TO(neoR) RAD51-T131P	Construction of hTERT-RPE1 RAD51 separation-of-function variants, previously generated by lab (Wassing <i>et al.</i> , 2020), resistance cassette switched from hygR to neoR (section 2.2.3).
pDEST- FLAG_FRT_TO(neoR) RAD51-II3A	Construction of hTERT-RPE1 RAD51 separation-of-function variants. Site-directed mutagenesis of R130A, R303A, K131A from pDEST-FLAG_FRT_TO RAD51-WT, resistance cassette switched from hygR to neoR (section 2.2.3).

2.1.3 Oligonucleotides

2.1.3.1 PCR primer sequences

Table 2.3: PCR primer sequences

Target	Construction	Sequence
5'_pcDNA_neoR	NeomycinR cassette	5' TATTCTCTAGAAAGTATAGGAACTTCCTTTTCGCAT- GATTGAACAAGATGG 3'
3'_pcDNA_neoR	NeomycinR cassette	5' GGCGTCCCGGAAAACG 3'
R130A_F	pDEST- RAD51-II3A	5' AGAAATGTTTGGAGAATTCGCAACTGGGAAGAC- CCAGATC 3'
R130A_R	pDEST- RAD51-II3A	5' GATCTGGGTCTTCCCAGTTGCGAATTCTCCAAA- CATTTCT 3'

Continued on next page

Table 2.3: PCR primer sequences (Continued)

Target	Construction	Sequence
R303A_F	pDEST- RAD51-II3A	5' AACCAACCAGATTGTATCTGGCGAAAGGAA- GAGGGGAAACC 3'
R303A_R	pDEST- RAD51-II3A	5' GGTTTCCCCTCTTCCTTTTCGCCAGATACAATCTG- GTTGTT 3'
K313A_F	pDEST- RAD51-II3A	5' GGGGAAACCAGAATCTGCGCAATCTACGACTCTCC- CTG 3'
K313A_R	pDEST- RAD51-II3A	5' CAGGGAGAGTCGTAGATTGCGCAGATTCTGGTTTC- CCC 3'

2.1.3.2 siRNA sequences

Table 2.4: siRNA sequences

Target	Final conc.	Sequence
siCENP-A	20 nmol	5' GGACUCUCCAGAGCCAUGAUU 3'
siBRCA2 #1	20 nmol	5' CAACAAUUACGAACCAAAC 3'
siBRCA2 #2	20 nmol	5' CUGAGCAAGCCUCAGUCAA 3'
siRAD51 #1	20 nmol	5' GACUGCCAGGAUAAAGCUU 3'
siRAD51 #2	20 nmol	5' GUGCUGCAGCCUAAUGAGA 3'
siRAD51 3' UTR #1	20 nmol	5' GACUGCCAGGAUAAAGCUU 3'
siRAD51 3' UTR #2	20 nmol	5' GUGCUGCAGCCUAAUGAGA 3'

2.1.3.3 RT-qPCR primer sequences

Table 2.5: RT-qPCR primer sequences

Target	Input	Forward sequence	Reverse sequence	Source
5S rDNA	1-90ng	5' CCGGACCC- CAAAGGCG- CACGCTGG 3'	5' TGGCTGGCGTCT- GTGGCACCCGCT 3'	(Nakano <i>et al.</i> , 2008)
D17Z-1	1-90ng	5' CTTTGGATG- GAGCAGGTTTGA- GAC 3'	5' CCGTTTAGTTAG- GTGCAGTTATCC 3'	(McNulty <i>et al.</i> , 2017)
Alpha / alphoid	1-90ng	5' CTCACA- GAGTTGAAC- CTTCC 3'	5' GAAGTTTCTGA- GAATGCTTCTG 3'	(Chan <i>et al.</i> , 2012)
TERRA-17p	1-90ng	5' GGGACAGAAGTG- GATAAGCTGATC 3'	5' GATCCCACT- GTTTTTATTACT- GTTCCCT 3'	(Beishline <i>et al.</i> , 2017)
TERRA-Xq	1-90ng	5' CCCCTTGC- CTTGGGAGAA 3'	5' GAAAGCAAAAGC- CCCTCTGA 3'	(Beishline <i>et al.</i> , 2017)
MALAT-1	1-90ng	5' GGTCTGTGCTA- GATCAAAAGG 3'	5' AAAGCAAG- GTCTCCCCACAAG 3'	(Merta <i>et al.</i> , 2020)
NEAT-1	1-90ng	5' GTGGTAGGAAAT- GCAGGTTGA 3'	5' CCAAGCAA- CAACTTAAC- CAACA 3'	(Merta <i>et al.</i> , 2020)
RAD51	100ng	5' TCTCTGGCAGT- GATGTCCTGGA 3'	5' TAAAGGGCG- GTGGCACTGTCTA 3'	(Chappell <i>et al.</i> , 2015)
BRCA2	100ng	5' CCAAGTGGTC- CACCCCAAC 3'	5' CACAATTAGGA- GAAGACATCA- GAAGC 3'	(Rabiau <i>et al.</i> , 2011)

Continued on next page

Table 2.5: RT-qPCR primer sequences (Continued)

Target	Input	Forward sequence	Reverse sequence	Source
CENP-A	100ng	5' ACAAGGTTGGC-TAAAGGA 3'	5' ATGCTTCTGCT-GCCTCTT 3'	(Li <i>et al.</i> , 2011)
GAPDH	100ng	5' CTGTTGCTG-TAGCCAAATTCGT 3'	5' ACCCACTCCTC-CACCTTTGAC 3'	(Chappell <i>et al.</i> , 2015)

2.1.3.4 ChIP-qPCR primer sequences

Table 2.6: ChIP-qPCR primer sequences

Target	Dilutions	Forward sequence	Reverse sequence	Source
tetO	1:1	5' CCACTCCCTATCAGT-GATAGAGAA 3'	5' TCGACTTCTGTT-TAGTTCTGTGCG 3'	(Erliandri <i>et al.</i> , 2014)
kanR	1:1	5' AACATCGCATC-GAGCGAGCA 3'	5' AAGCACGAG-GAAGCGGTCAG 3'	N/A
endo. cen (total alphoid)	1:10	5' CTCA-GAAACTTCTTTGT-GATGTGT 3'	5' TATTCCCTTTTG-GAACGAAGGC 3'	(Erliandri <i>et al.</i> , 2014)
GAPDH	1:1	5' CTCATGAC-CACAGTCCATGCC 3'	5' GCCATC-CACAGTCTTCTGGGT 3'	(McNulty <i>et al.</i> , 2017)
160/161	1:1	5' ACGTGGCTTCCT-CACTAAGC 3'	5' GGACTCCTC-CATTCACCAGC 3'	Nieduszynski lab

2.1.3.5 exo-FISH PNA probe sequences

Table 2.7: exo-FISH PNA probe sequences

Target	Catalog No.	Sequence	Source
CENPB-RC-Cy3	F3009	5' ATTCGTTGGAAACGGGA 3'	PNABio
TelC-Cy5	F1003	5' CCCTAACCCTAACCCTAA 3'	PNABio

2.1.4 Antibodies

2.1.4.1 Primary antibodies

Table 2.8: Primary Antibodies

Target	IF Dilution	ChIP Input	Catalog No.	Source
eYFP (GFP)	-	2µg	G1544	Sigma
BRCA2	1:1000	2µg	OP95	Millipore
RAD51	1:2000	2µg	ab176458	Abcam
CENP-A	1:1000	2µg	ab13939	Abcam
IgG (mouse)	-	2µg	10400C	Invitrogen
Lamin-A	1:20 000	-	L1293	Sigma
Topoisomerase IIα	1:1000	-	D10G9	CST
γH2A.X	1:1000-10:000	-	05-636	Millipore
CENP-B	1:1000	-	IHC-00064	Bethyl Laboratories
Pol II (pan)	1:1000	-	N/A	Proudfoot lab
Pol II-S2P	1:1000	-	N/A	Proudfoot lab

2.1.4.2 Secondary antibodies

Table 2.9: Secondary Antibodies

Target	Dilution	Catalog No.	Source
anti-rabbit Alexa Fluor 555	1:1000	A21430	Thermo Fisher Scientific
anti-rabbit Alexa Fluor 647	1:1000	A27040	Thermo Fisher Scientific
anti-mouse Alexa Fluor 488	1:1000	A11017	Thermo Fisher Scientific
anti-mouse Alexa Fluor 555	1:1000	A21425	Thermo Fisher Scientific
anti-mouse Alexa Fluor 647	1:1000	A21237	Thermo Fisher Scientific
anti-human Alexa 647	1:1000	109-605-088	Jackson ImmunoResearch

2.1.5 Treatment and control compounds

Table 2.10: Treatment and control compounds

Compound	Catalog No.	Concentration
Dimethyl sulfoxide (DMSO)	D2650, Sigma	100% at corresponding volume
5,6-dichlorobenzimidazole-1- β -D-ribofuranoside (DRB)	D1916, Sigma	75 μ M
S-trityl-L-cysteine (STLC)	164739, Sigma	5 μ M
RAD51 inhibitor (B02)	533525, Calbiochem	20 μ M
Etoposide (ETOP)	E1383, Sigma	25 μ M
Doxycycline hyclate (DOX)	D9891, Sigma	1 μ g/ml

2.1.6 Accession codes

Table 2.11: Accession codes

Dataset	Source	SRA
HCT116 GLOE-seq	Sriramachandran <i>et al.</i> 2020	SRR9676440
HCT116 END-seq	Canela <i>et al.</i> 2019	SRR8870099
HCT116 input	ENCODE	SRR504927 / SRR577511

Continued on next page

Table 2.11: Accession codes (Continued)

Dataset	Source	SRA
BLESS - APH	Crosetto <i>et al.</i> 2013	SRR695321
BLESS + APH	Crosetto <i>et al.</i> 2013	SRR695402
HeLa input	ENCODE	SRR5331490
CENP-A CUT&RUN	Thakur <i>et al.</i> 2018	SRR5886259
IgG CUT&RUN	Thakur <i>et al.</i> 2018	SRR5886261
γ H2A.X ChIP-seq - APH	Lyu <i>et al.</i> 2019	SRR6986055
γ H2A.X ChIP-seq + APH	Lyu <i>et al.</i> 2019	SRR6986058
γ H2A.X ChIP-seq input	Lyu <i>et al.</i> 2019	SRR6986070
sBLISS siCTRL	Hazan <i>et al.</i> , 2019	SRR10045032
sBLISS siRAD51	Hazan <i>et al.</i> , 2019	SRR10045033

2.2 Cloning and bacterial cell culture

2.2.1 Plasmid amplification

DB3.1 competent cells were thawed on ice and transformed by mixing ~ 10 ng plasmid per 100 μ l DB3.1 and tapping gently. After 30 min incubation on ice and a 30 second heat-shock at 42°C, cells were left on ice for 2 minutes. Transformed cells were then allowed to recover for 1 hour at 37°C with gentle perturbation. Finally, pellets formed by gentle centrifugation (200 rcf, 4 minutes) were resuspended in 100 μ l filter-sterilized super optimal broth (SOB) containing 20 mM glucose. Cells were plated on LB agar plates containing 100 μ g/ml ampicillin and incubated at 37°C. Colonies formed over the following 24-48 hours were re-streaked to form single colonies. Single colonies were subsequently inoculated in 5-30 ml LB broth containing 100 μ g/ml ampicillin for plasmid extraction the following day.

Plasmids were extracted from ~ 5 -30 ml bacterial overnight cultures using the Qiagen miniprep kit (Qiagen, 27106) or Qiagen plasmid Plus midiprep kit (Qiagen, 12943) according to manufacturer's specifications.

2.2.2 Glycerol stocks

Bacterial glycerol stocks for long-term storage at -80°C were formed by mixing 500 µl saturated bacterial overnight cultures with 500 µl 50% glycerol, mixing thoroughly and placing in cryovials.

2.2.3 Construction of RAD51 variant plasmids

pDEST-FLAG-FRT-TO plasmids expressing FLAG, RAD51-WT, RAD51-K133R and RAD51-T131P were generated by previous lab members (Wassing *et al.*, 2020). RAD51-II3A point mutations (R130A, R303A, K131A) were introduced using sequential site-directed PCR mutagenesis of pDEST-FLAG-RAD51-WT.

For site-directed mutagenesis, pDEST-FLAG-RAD51-WT plasmids were PCR-amplified using PfuUltra II (Agilent, 600670) and primers listed in **section 2.1.3.1**. Per PCR reaction, a final concentration of 1X PfuUltra reaction buffer, 100 mM dNTPs, 5-30 ng primer DNA template, 10 µM of forward and reverse primers and 0.4 µl PfuUltra High Sensitivity polymerase were added to a final reaction volume of 20 µl. Following initial denaturation (95°C, 2 min), 20 cycles of denaturation (95°C, 20 seconds), annealing (50-60°C, 20 seconds) and extension (72°C, 105-135 seconds) were completed prior to a final extension (72°C, 3 min). DpnI (NEB, R0176S) digestion was subsequently performed using 2 units DpnI in 1X CutSmart buffer per 30 µl PCR product for 1 hour at 37°C. DpnI-digested PCR products were transformed into DB3.1 for plasmid amplification **section 2.2.1**. Amplified colonies were submitted for Sanger sequencing after each round of mutagenesis to ensure the desired sites were mutated.

To enable transfection of the RAD51 separation-of-function variants into the hTERT-RPE1 Flp-In T-REx cell line, the hygromycin resistance cassette was switched for a neomycin resistance cassette for all pDEST-FLAG-FRT-TO plasmids. First, the neomycin resistance cassette was PCR-amplified from the pcDNA 3.1+ vector backbone, with primers containing XbaI and PfoI recognition sequences. Both plasmids and the PCR-amplified neomycin resistance cassette were then digested with 10 units restriction enzymes XbaI (ThermoScientific, ER0685) and PfoI (ThermoScientific, ER1751) with 4 units Calf Intestinal Alkaline Phosphatase per 1 µg plasmid DNA for 1 hour at 37°C. Digests were loaded onto a 1% agarose gel and extracted using the Qiagen Gel Extraction kit. Ligations were then performed with 1:3 vector:insert molar ratio using a Takara DNA Ligation Kit (Takara Bio, 6023) according to manufacturer's specifications. Ligation products were then used for plasmid amplification **section 2.2.1**.

2.3 Mammalian cell culture

2.3.1 General cell culture

hTERT-immortalized retinal pigmented epithelial cells (hTERT-RPE-1) were cultured in standard growth conditions (5% CO₂, 37°C) in complete media (1:1 Dulbecco Modified Eagle Medium / Nutrient Mixture F12 supplemented with 10% fetal bovine serum (FBS) (F9665, Merck), 1% penicillin-streptomycin (15140122, Life Technologies) and 0.123% sodium bicarbonate (S8761, Sigma-Aldrich)). hTERT-RPE1 Flp-In T-REx were grown under similar conditions, including 100 µg/ml Zeocin (ant-zn-1, InvivoGen) and 10 µg/ml Blasticidin (ant-bl-1, InvivoGen).

HeLa CENP-A-SNAP cells were cultured in standard growth conditions (5% CO₂, 37°C) in complete media (Dulbecco Modified Eagle Medium supplemented with 10% fetal bovine serum (FBS) (F9665, Merck), 1% penicillin-streptomycin (15140122, Life Technologies) and 5 µg/ml blasticidin for the maintenance of CENP-A-SNAP).

2.3.2 Transfections

hTERT-RPE1 Flp-In T-REx cells expressing RAD51 separation-of-function variants were generated by transfecting cells with 500 ng pDEST-FRT plasmids containing FLAG, FLAG-RAD51-WT, FLAG-RAD51-K133R, FLAG-RAD51-T131P or FLAG-RAD51-II3A with 2 µg pOG44 using Lipofectamine LTX, according to manufacturer's specifications. Selection was done in 500 µg/ml G418 (G8168, Sigma-Aldrich), and maintained in 200 µg/ml G418 and 10 µg/ml Blasticidin. To induce variant expression, cells were incubated with complete media containing 1 µg/ml doxycycline (D9891, Sigma Aldrich) for at least 24 hours.

2.3.3 siRNA transfections

Lipofectamine RNAiMAX (13778075, ThermoFisher) was used for siRNA-mediated depletion according to manufacturer's recommendation. A final concentration of 20 nM siRNA was used for siRAD51, siBRCA2, siCENP-A or the universal negative control siMisNeg (SIC001, Merck) using the sequences listed in **section 2.1.3.2**. Cells were further incubated for 24 hours before changing media to fresh complete media and harvesting at the indicated times for analysis.

2.3.4 Cell cycle regulation

To enrich for mitotic cells, cells were incubated with 5 µM S-trityl-L-cysteine (STLC) (164739, Sigma-Aldrich) for the indicated period (usually ~3-6 hours). Mitotic cells were specifically collected by mitotic

shake-off, by banging dishes against a hard surface and collecting dislodged cells in the supernatant.

To induce quiescence (i.e., non-replicating conditions), cells were treated with serum-starved media (media with usual additives except for 0.5% FBS instead of 5% FBS). Following 24 hours serum starvation, cell proliferation was effectively halted. When assessing the proportion of proliferating cells, samples were incubated in 10 μ M 5'ethynyl-2'-deoxyuridine (EdU) to label DNA replication and harvested for flow cytometry analysis **section 2.4.6** at the indicated times.

2.4 Molecular biology techniques

2.4.1 Western blotting

To verify protein depletion by western blot, cells were harvested by trypsinization at the indicated time points by centrifugation (500g, 5 min) and washed twice with ice-cold PBS. After the second wash, cell pellets were weighed to estimate pellet mass and resuspended in NETN150 buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.5% NP-40, 10 mM Benzamidine HCl, 10 mM NaF, 1 mM sodium glycerophosphate, 1 mM dithiothreitol, 25U/ml Benzonase nuclease supplemented with protease inhibitor cocktail (P2714, Sigma-Aldrich)) at a volume of 50-100 μ l per 10 mg cells. Cells were then left on ice for 30 minutes with occasional agitation by vortex. Pellets were then centrifuged for >15 minutes at 16,100g to pellet debris, and the supernatant was collected as the whole cell extract (WCE). Protein concentrations were measured by the Bradford protein concentration assay (BioRad, #5000006) according to manufacturer's recommendations. WCEs were diluted to 1-2 mg/ml protein, and ~25 μ g protein was loaded on an SDS-PAGE gel (ThermoFisher, MP0322BOX) along with 2.5 μ l protein ladder protein ladder (Geneflow, S6-0024). The SDS-PAGE gel was run at 80-120V for 2-4 hours, or until the desired resolution was reached.

Gels were then removed and transferred to 0.2 μ M or 0.45 μ M nitrocellulose membranes, depending on the protein of interest, for 10V overnight, 70V for 3 hours, or 90V for 1 hour. For transfer, the gel and membrane were surrounded by two layers of Whatman paper on either end in transfer buffer (100 mM Tris, 767 mM glycine, 10% methanol). The membrane was stained with Ponceau (0.1% Ponceau in 5% acetic acid) to visualize protein loading, and then blocked in PBST (PBS + 0.05% Tween-20) with 5% milk for 30-60 minutes. Membranes were then incubated with primary antibodies (dilutions listed in **section 2.1.4.1**) at 4°C overnight. The next day, membranes were washed thrice for 5 minutes in PBST, incubated with secondary antibodies (dilutions listed in **section 2.1.4.2**) for 1 hour at room temperature and washed thrice in PBST again. Finally, membranes were reacted with a 1:1 ECL Western Blotting detection solution and developed using a Curix 60 tabletop processor.

2.4.2 Cell lysis and DNA extraction

To extract DNA from cells directly, cells were harvested by trypsinization, washed in PBS, counted, and lysed in cell lysis solution (10 mM Tris-Cl pH 8.0, 100 mM EDTA, 0.5% w/v SDS), adding ~200 µl per 1M cells. RNaseA (Sigma, R6513) was added at ~50 µg/ml and incubated at room temperature for ~30 minutes. Proteinase K (Roche, 03115887001) was subsequently added (1:20 of stock), and samples were incubated at 50°C for at least 1 hour and allowed to cool to room temperature before proceeding.

For DNA extraction, an equivalent volume 25:24:1 phenol-chloroform-isoamyl solution (P3803, Sigma) was added to each sample in the chemical hood and mixed by vortexing for 2 minutes at max speed. Samples were centrifuged for 2 minutes at max speed, and the top layer was subsequently transferred to a new tube. These steps were then repeated using just chloroform (288306, Sigma), once again using a chemical hood, and moving the top layer to a new tube. At this point, 1 µl glycogen (10901393001, Roche) could be added to each sample and mixed vigorously to aid pellet formation. To each sample, 2X volume ice-cold 100% EtOH was added, mixed vigorously, and samples were incubated for 20+ mins at -80°C. Samples were then centrifuged for 30 minutes at max speed (4°C) and pellets were washed once with ice-cold 70% EtOH, centrifuging for 10 minutes at max speed (4 °C) to form DNA pellets. Once any residual EtOH was removed, pellets were air-dried for ~20 minutes and resuspended in dH₂O to desired volume.

2.4.3 Chromatin immunoprecipitation (ChIP)

For chromatin immunoprecipitation (ChIP) experiments, cells were seeded in 15-cm dishes at densities to reach desired confluence at the point of harvesting. Where indicated, siRNA transfections were performed as described above (**section 2.3.2**). At the indicated time points, cells were harvested by trypsinization, washed twice in ice-cold PBS, and counted. If needed, cell aliquots were collected at this time for subsequent protein analysis by Western blotting (**section 2.4.1**). Cells were then fixed for 8 minutes, RT in ice-cold PBS containing 1% formaldehyde (Sigma, F1635) at a concentration of ~7 million cells per ml, mixing initially by gentle vortex throughout. Fixation was then quenched by adding 125 mM glycine, mixing thoroughly, and incubating for 5 minutes at room temperature. Samples were then washed twice in ice-cold PBS. Samples were resuspended in cell lysis buffer (10 mM PIPES pH 7.5, 85 mM KCl, 0.5% NP-40, 1X Sigma protease inhibitor cocktail, 10 mM benzamidine hydrochloride) to a final concentration of 5 million cells per ml and incubated for 10 minutes on ice to isolate nuclei. Nuclei were then collected by centrifugation (3 min, 1000g, 4°C) and resuspended in ChIP buffer (20 mM Tris-Cl pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 1X Sigma protease inhibitor cocktail, 10 mM benzamidine hydrochloride) at ~1 ml ChIP buffer per 10 million cells. Samples were then sonicated for 5 cycles of 10 minutes (30 seconds on, 30 seconds off) to shear

chromatin using an ultra sonicator water bath (Bioruptor, Diagenode). Note that sonication times should be optimized for a given cell line and treatment condition.

Cellular debris was then pelleted by centrifugation (15 min, 16100g, 4°C) and supernatants were collected for further processing. Western Blot aliquots (~10 µl) were collected at this point as the 'input' sample. While resting on ice, aliquots (30-60 µl) of the sheared chromatin were analyzed to ensure successful shearing. First, aliquoted quality-control samples were incubated with 300 mM NaCl, ~720 µg/ml RNaseA (Sigma, R6513) and 4 µl Proteinase K (Roche, 03115887001) in nuclease free TE buffer for 3 hours at 65°C with moderate agitation on a mixer (~600rpm). DNA was then extracted according to the standard protocol (**section 2.4.2**), and concentrations were measured by Nanodrop, ideally ranging between 50-200 ng/µl. 500 ng of DNA was then run on a 1.5% agarose gel to verify successful sonication (~300-500bp).

Using the concentrations calculated from the quality-control samples above, ~40 µg sheared chromatin was then incubated with 2.0 µg primary antibodies (listed in **section 2.1.4.1**) overnight at 4°C with gentle agitation, preferably on a rotator wheel. Protein G Dynabeads (ThermoFisher, 1003D) (~20 µl per sample) were washed twice in ChIP buffer using magnetic stands and blocked overnight in ChIP buffer with 5 mg/ml BSA, at 4°C with gentle agitation.

The following day, Protein G Dynabeads were washed twice in ChIP buffer and resuspended in the original volume. Samples incubated with primary antibodies were then added to 20 µl of Protein G beads and further incubated at 4°C for 2 hours with gentle agitation. Beads were then washed twice in low-salt wash buffer (20 mM Tris-Cl pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 1X Sigma protease inhibitor cocktail, 10 mM benzamidinium hydrochloride), twice in high-salt wash buffer (20 mM Tris-Cl pH 8, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 1X Sigma protease inhibitor cocktail, 10 mM benzamidinium hydrochloride), once in LiCl wash buffer (20 mM Tris-Cl pH 8, 250 mM LiCl, 1 mM EDTA, 1% sodium deoxycholate, 1X Sigma protease inhibitor cocktail, 10 mM benzamidinium hydrochloride), and once in TE buffer (10 mM Tris-Cl pH 8, 1 mM EDTA). Finally, beads were resuspended in 200 µl Elution buffer (10 mM Tris-Cl pH 8, 1 mM EDTA, 1% SDS) and incubated at 65°C for 20 minutes with strong agitation (1000 rpm). Supernatants were then collected, and Western blot aliquots were taken at this point to verify immunoprecipitation efficiency. Crosslinks were then reversed overnight by incubating samples in 300 mM NaCl, ~720 µg/ml RNaseA (Sigma, R6513) and 1:20 Proteinase K (Roche, 03115887001) in nuclease free TE buffer at 65°C with moderate agitation on a mixer (~600rpm). DNA was extracted (**section 2.4.2**), and DNA samples were subsequently assessed by qPCR.

For qPCR analysis, optimization curves were first performed on an input sample to determine optimal DNA dilutions for linear amplification of each qPCR primer (listed in **section 2.1.3.4**). Following dilution, 3 µl of DNA was assessed using SensiFAST SYBR No-ROX kit (BIO-98005, Meridian Bioscience) on the Rotorgene Q Real-Time PCR System (Qiagen) using Q-Rex software. Amplification conditions included an initial melting at 95°C for 30 seconds, followed by 40 cycles of amplification (95°C 5 seconds, 62°C 30 seconds). Technical triplicates were performed for each reading. Arbitrary thresholds were chosen to determine Ct, and relative DNA levels were calculated as $2^{-\Delta \Delta C_t}$ where Ct values are subtracted from two controls, as specified in figure legends. Where indicated, qPCR products were run on a 2% agarose gel along with a 100-bp molecular ladder to visualize product sizes.

2.4.4 Reverse transcription polymerase chain reaction (RT-PCR)

RNA was extracted from mitotic or quiescent populations using TRI reagent solution (AM9738, Thermo Fisher Scientific) according to manufacturer's recommendations. Extracted RNA was tested by Nanodrop to verify nucleic acid quality and diluted to a consistent concentration between samples. If phenol contaminations were evident by an elevated A280 peak, nucleic acids were re-precipitated. Briefly, 0.1X volume 3M sodium acetate was added to contaminated samples along with 2.5X volume 100% Ethanol (EtOH). Following robust mixture, samples were incubated at -80°C for >30 mins, and centrifuged at 12,000g, 4°C for 20 mins. Pellets were vigorously washed once in cold 70% EtOH and centrifuged again at 12,000g, 4°C for 10 minutes. After removing supernatant and air-drying for 10 minutes, samples were resuspended in 25 µl nuclease-free dH₂O and diluted to a consistent concentration between samples.

Contaminating genomic DNA was then removed by two successive rounds of DNase treatment using the TURBO DNA-free kit (AM1907, Invitrogen). Up to 2 µg RNA was then reverse-transcribed to cDNA using a high-capacity cDNA reverse transcription kit (4368814, Applied Biosystems), including a non-reverse transcribed (RT-) sample to detect any signal contributions from contaminating gDNA. The equivalent of 1-100 ng cDNA input (as specified in **section 2.1.3.3** or indicated in figure legends) was then assessed by qPCR using SensiFAST SYBR No-ROX kit (BIO-98005, Meridian Bioscience) on the Rotorgene Q Real-Time PCR System (Qiagen) using Q-Rex software. Amplification conditions included an initial melting at 95°C for 30 seconds, followed by 40 cycles of amplification (95°C 5 seconds, 62°C 30 seconds). Technical triplicates were performed for each reading. Arbitrary thresholds were chosen to determine Ct, and relative RNA levels were calculated as $2^{-\Delta C_t(\text{treatment-control})}$, with respective controls indicated in each figure legend.

2.4.5 Flow cytometry

First, cells were harvested by trypsinization and pelleted by centrifugation (500g, 5 min). Cell pellets were then washed twice in PBS and fixed in 70% EtOH overnight at 4°C. The following day, fixed cells were washed twice in PBS, followed by a resuspension in PBSTri-BSA (PBS, 0.1% Triton X-100, 1% BSA) and incubation on ice for 15 minutes. Cells were then washed thrice in PBST-BSA (PBS, 0.1% Tween 20, 1% BSA). To label EdU as a marker of cell proliferation, the Click-iT reaction was used to label any incorporated EdU in Click-iT reaction buffer (PBS pH 7.2, 2 mM CuSO₄, 10 mM sodium ascorbate (A4034, Sigma-Aldrich), 10 µM Alexa Fluor Azide (A20012, ThermoFisher Scientific)) at room temperature. Any remaining Click-iT solution was washed out with three washes in PBST-BSA, and cells were finally resuspended in DAPI staining solution (PBS, 0.1% BSA, 100 µl RNaseA and 2 µg/ml DAPI) prior to flow cytometry. Samples were analyzed on the Cytoflex LX using CytExpert (v.2.3.0.84) and data processed using FlowJo (v10.6.2) (<https://www.flowjo.com/>). Single cells were selectively analyzed by SSC-A, FSC-A and FSC-W using standard gating strategies, as described in text and EdU+ proportions gated as shown in **Figure 4.5D**.

2.4.6 Cell survival analysis (WST-1)

To determine cell survival in the hTERT-RPE1 RAD51 variant lines, cells were seeded at 1K per well in a 96-well plate in technical triplicate, with or without 1 µg/ml doxycycline, where indicated. The following day, endogenous RAD51 was depleted using the protocol described above (**section 2.3.3**), and cells received fresh media with or without 1 µg/ml doxycycline the following day. Fresh media was given every second day. 96 hours following siRNA transfections, cells were incubated with complete media containing 10% WST (5015944001, Roche) for 1 hour at 37°C before reading plate absorbance at 450 and 650 nm. Cell survival was calculated based on 450 nm readings subtracted from the background (650 nm) and converted to a percentage of the negative control.

2.5 Microscopy techniques

2.5.1 exo-FISH

At the indicated time points, cells were harvested by trypsinization, washed in 1X PBS and counted. An equivalent number of cells (usually ~100K) were aliquoted for downstream processing, keeping cell numbers consistent between experimental conditions. Cells were then resuspended in 0.56% KCl (~20K cells / ml) and incubated at room temperature for 20 minutes. Cells were then spun down (4 min, 500g), resuspended in ~100 µl PBS gently but thoroughly and fixed in 3:1 methanol:acetic acid for 20 minutes, adding fixative in gently while rocking samples to prevent cell aggregation. Fixed cells were

then pelleted again (4 min, 500g), resuspended gently in ~60 µl remaining fixative and spread onto glass slides (12392138, Thermo Fisher Scientific) by dropping ~10 µl cell solution onto a ~20 µl drop of water. Since the density of cells can influence the FISH signal intensity, a consistent number of cells (~20-50K) were spread homogenously on each slide and should be optimized for each cell line. Slides were dried overnight at room temperature in the dark. Visual inspection by light microscopy should be performed to ensure cell spreading is homogenous before proceeding with FISH processing.

The following day, slides were rehydrated in PBS for 5 minutes at room temperature in coplin jars and treated with 0.5 mg/ml RNaseA (R6513, Sigma-Aldrich) for 10 minutes at 37°C in a humidified chamber. Slides were then incubated with 50-200 mU/µl Exonuclease III (M1811, Promega) in 1X Exonuclease III buffer (50 mM Tris-Cl, 5 mM MgCl₂, 5 mM DTT, pH 8) for 30 minutes at 37°C in a humidified chamber. Slides were serially dehydrated with 70%, 95% and 100% EtOH, 5 minutes each, and air-dried. Slides were then treated with 200 nM CENPBR-Cy3 (F3009, PNABio) as well as 200 nM TelC-Cy5 (F1003, PNABio) (where indicated) for 90-120 minutes at room temperature. Before treatment, PNA probes were dissolved and mixed vigorously in hybridization solution (10 mM Tris-Cl pH 7.2, 70% formamide, 0.5% blocking solution (blocking solution: 10% blocking reagent (w/v, 11096176001, Roche) dissolved in maleic acid buffer (100 mM maleic acid, 150 mM NaCl, pH 7.5))). Slides were washed for 15 minutes in hybridization wash buffer #1 (10 mM Tris-HCl pH 7.2, 70% formamide, 0.1% BSA), followed by 3 5-minute washes in hybridization wash buffer #2 (0.1M Tris-HCl pH 7.2, 0.15 M NaCl, 0.08% Tween-20). Finally, slides were once again serially dehydrated in 70%, 95% and 100% EtOH, 5 minutes each, before air-drying and mounting on ProLong mounting media containing DAPI (P10144, ThermoFisher Scientific).

Slides were finally imaged with the Olympus FV1000 Fluoview Laser Scanning Microscope with Becker and Hickel FLIM system and FV10-ASW software (v4.2), and cell aggregates were avoided when imaging. To ensure unbiased imaging and quantification, cells were selected by DAPI staining only and quantification was automated. Image processing and quantification are described below (**section 2.6.3**). For the validation of exo-FISH using restriction enzymes *in vitro*, cells were treated with 25 mU/µl (total 5U) BsmAI (R0529S, NEB) or Nt.BsmAI (R0121S, NEB) in 1X CutSmart Buffer (NEB) for 1 hour at 37°C in a humidified chamber, in between the RNaseA digestion and the ExoIII treatment. The negative control was treated with 1X CutSmart Buffer alone. Where indicated, cells were treated with 25 µM Etoposide (E1383, Sigma-Aldrich) or the corresponding volume of 100% DMSO (D2650, Sigma-Aldrich).

2.5.2 Native immuno-cytofluorescence (ICF)

Mitotic cells (enriched as in **section 2.3.4**) were collected by mitotic shake-off, washed in PBS, counted, and diluted to a consistent cell concentration in 0.56% KCl (usually 100-250K cells/ml). Cells were then

incubated for 10-15 minutes at room temperature with gentle rocking. Cells were spun down onto glass slides (~10K cells) using a cyto-spinner. Slides were then incubated in KCM buffer (10 mM KCl, 10 mM Tris-Cl pH 8.0, 1 mM EGTA, 0.1% Triton X-100, pH 7.5) for 10 minutes. For immunofluorescence staining, slides were incubated with primary antibodies (as specified in **section 2.1.4.1**) for 1 hour at 37°C in KCM with 3% BSA, followed by a 15-minute wash in KCM buffer. Slides were then incubated with secondary antibodies for 1 hour at 37°C in KCM with 3% BSA, washed in KCM for 15 minutes, and fixed with 3.5% PFA in PBS for 20 minutes. After washing slides three times with PBS, with the second wash containing 500 ng/ml DAPI (564907, BD Biosciences), slides were finally mounted using ProLong mounting media containing DAPI (P10144, ThermoFisher Scientific). Slides were imaged with the Olympus FV1000 Fluoview Laser Scanning Microscope with Becker and Hickel FLIM system and FV10-ASW software (v4.2), and cell aggregates were avoided when imaging. To ensure unbiased imaging and quantification, cells were selected by DAPI staining only and quantification was automated. Image processing and quantification are described below (**section 2.6.3**).

2.5.3 Immunofluorescence (IF)

First, 12-mm coverslips were pre-treated with ~1 ml poly-L-lysine for 15 minutes and air-dried for ~15 minutes. Cells were harvested at the indicated times, washed in PBS, counted, resuspended to a consistent cell concentration in PBS, and seeded at ~50K cells per pre-treated coverslip for 15 minutes. Cells were left at 4°C for 10 minutes until firmly attached to coverslips and fixed in 4% PFA in PBS (28906, Pierce) in PBS at room temperature for 10 minutes. PFA solution was added very gently to avoid washing away cells. Coverslips were then quenched in ~1 ml 0.1M Tris-Cl pH 7.5 for 5 minutes, washed in PBS and stored for up to 2 weeks in PBS containing 0.04% NaN₃ at 4°C.

Slips were permeabilized in ~500 µl PBS-TX (PBS, 0.1% (v/v) Triton X-100) for 10 minutes, incubated in IF blocking buffer (PBS, 2% (v/v) FBS, 2% (w/v) BSA, 0.1% (v/v) Triton X-100, 0.04% (w/v) NaN₃) for 30 minutes at 37°C, and incubated with primary antibodies (as listed in **section 2.1.4.1**) diluted in IF blocking buffer for 1 hour at 37°C, both in a humidified chamber. Slides were then washed thrice with PBS-TX, incubated with secondary antibodies (as listed in **section 2.1.4.2**) diluted in IF blocking buffer for 1 hour at 37°C and washed again thrice in PBS-TX before mounting using ProLong mounting media containing DAPI (P10144, ThermoFisher Scientific).

Slides were imaged with the Olympus FV1000 Fluoview Laser Scanning Microscope with Becker and Hickel FLIM system and FV10-ASW software (v4.2), and cell aggregates were avoided when imaging. To ensure unbiased imaging and quantification, cells were selected by DAPI staining only and quantification was automated. Image processing and quantification are described below (**section 2.6.3**).

2.5.4 SNAP labelling for CENP-A deposition and stability assays

For CENP-A-SNAP labelling in the loading assay, HeLa CENP-A-SNAP cells were seeded at 20K cells per 12-mm poly-L-lysine-pre-treated coverslips. On the following day, siRNA treatments were performed as described above (**section 2.3.3**). A day later, slides were quenched with complete media containing 2 μ M SNAP blocking reagent (BTP) (S9106S, NEB) for 30 minutes at 37°C, followed by three washes in 37°C PBS and another 30-minute incubation in complete media. Cells were then washed thrice in 37°C PBS and allowed to grow for 48 hours before pulse-labelling CENP-A-SNAP. For pulse-labelling, cells were incubated with complete media containing 2 μ M TMR-Star (S9105S, NEB) for 15 minutes at 37°C, followed by PBS and media washes as described in the quenching step. Slides were then fixed and processed for immunofluorescence as described above (**section 2.5.3**).

For CENP-A-SNAP labelling in the stability assay, a single pulse using 2 μ M TMR-Star was performed as described above directly before siRNA treatment (**section 2.2.3**). Slides were collected every 24 hours for 72 hours and fixed for immunofluorescence as described above (**section 2.5.3**).

Slides were imaged with the Olympus FV1000 Fluoview Laser Scanning Microscope with Becker and Hickel FLIM system and FV10-ASW software (v4.2), and cell aggregates were avoided when imaging. To ensure unbiased imaging and quantification, cells were selected by DAPI staining and quantification was automated. Image processing and quantification are described below (**section 2.6.3**).

2.6 Data analysis

2.6.1 Frequency analysis of arm copy-number alterations (arm-CNAs)

First, all ‘early’ driver event, as identified in Gerstung *et al.*, were manually aggregated from 38 different tumor types, including base mutations (e.g., KRAS), chromosome arm gain/loss (e.g., -17p), whole chromosome gain/loss (e.g., loss 2) and small structural variants (e.g., homdel 9p21.3). Each type of mutation was then counted and divided by the total number of tumor types to calculate the proportion of tumor types exhibiting that driver event as an ‘early’ event. Finally, the total number of events within each subclass were divided by the total number of driver events, indicating the relative contribution of each subclass to early carcinogenesis.

2.6.2 Short read sequencing analysis

Custom scripts are provided in the Appendix. First, short-read sequencing data were trimmed using Trimmomatic (v. 0.39) (Bolger *et al.*, 2014), ensuring read lengths were consistent between samples to be compared, and read quality was verified using FASTQC (v.0.11.5). Reads were then aligned

to the CHM13v1.0 assembly (Nurk *et al.*, 2021) using BWA-ALN (Li, 2013) with high stringency (no mismatches allowed, -n 0). BWA-SAMSE and SAMtools (v1.8) (Li *et al.*, 2009) were used to convert aligned reads into BAM format. PCR duplicates were removed using SAMtools. To ensure multi-mappers were only aligned once, BAM files were further filtered using SAMtools with FLAG score -F 2308. Using deepTools (v. 3.5.1) (Ramírez *et al.*, 2016), read scores across chromosomal coordinates were generated after normalizing to RPKM for comparisons between datasets. To calculate enrichment scores across centromeres of each chromosome (e.g., **Figure 3.4B**), the sums of RPKM-normalized read scores over CHM13v1.0 centromeric coordinates (Nurk *et al.*, 2021) (available on <https://github.com/marbl/CHM13>) were compared to that of the relevant negative control or input sequence. To calculate enrichment distributions within certain types of repeats (e.g., **Figure 3.5A, B**), percent of total dataset read scores mapping to a given coordinate was compared to its size (i.e., percentage genome occupancy). Python (v.3.8.8) (Van Rossum and Drake Jr, 1995), Matplotlib (v.3.4.2) (Hunter, 2007), NumPy (v.1.19.1) (Harris *et al.*, 2020), pandas (v1.2.3) (McKinney, 2010) and pyGenomeTracks (v3.6) (Lopez-Delisle *et al.*, 2021; Ramírez *et al.*, 2018) were all used for data processing and visualization.

2.6.3 exo-FISH, ICF and IF quantification and statistical analysis

Automated quantification scripts are provided in the Appendix. First, cells were segmented according to DAPI staining to separate cells as regions of interest (ROI). cenFSH or telFISH foci were automatically detected within each ROI using Fiji (v. 2.0.0-rc-69/1.52p) (Schindelin *et al.*, 2012) based on an arbitrarily determined threshold (usually ~400 arbitrary units). A 10x10-12x12 pixel box was generated around each focus and saved, after which the medial signal across all foci of a given cell (i.e., ROI) was calculated to create a ‘median focus’ per cell. The median value of the perimeter readings was then used to estimate the background signal, and this was subtracted from the median focus. Representative median foci were plotted using seaborn (v.0.11.1) (Waskom, 2021). For beeswarm plotting in Prism (www.graphpad.com), the sum of fluorescence signal within the background-subtracted median focus was calculated to encapsulate both focus intensity as well as size.

For native ICF and IF, the same approach as described above was used except CENP-B foci were automatically detected using an arbitrarily chosen threshold and used to define the 10x10-12x12 pixel box. The background-subtracted signal intensity of CENP-B as well as the secondary signal (e.g., CENP-A / γ H2A.X / TOP2 α / Pol II / Pol II – S2P) was then calculated within the CENP-B-defined box.

Student’s two-sided paired t-tests were used to compare the medians/averages of three biological replicates between experimental conditions using GraphPAD Prism 8 for Mac OSX (v.8.4.3) (www.graphpad.com). p value $\leq 0.05 = *$; $\leq 0.01 = **$; $\leq 0.001 = ***$; $\leq 0.0001 = ****$. Since the sample size (n=3) for each statistical test was too small to perform normality tests, parametric tests were chosen. No adjustments

for multiple comparisons were made since any data with multiple comparisons only contained positive controls as additional comparisons. Python (v.3.8.8) (Van Rossum and Drake Jr, 1995), Matplotlib (v.3.4.2) (Hunter, 2007), NumPy (v.1.19.1) (Harris *et al.*, 2020), pandas (v1.2.3)(McKinney, 2010) and seaborn (v.0.11.1) (Waskom, 2021) were all used for data processing and visualization.

Chapter 3

Investigating innate DNA fragility at human centromeres

3.1 Introduction

3.1.1 Centromeres as sites of chromosomal breakage in human disease

As discussed in **section 3.2.1.3**, centromeres and surrounding pericentromeres can act as potent sources of genome instability (Barra and Fachinetti, 2018; Black and Giunta, 2018). Perhaps one of the best illustrations of (peri)centromeres as sources of genome instability is the high incidence of chromosome arm copy-number alterations (arm-CNAs) observed in carcinogenesis. Chromosome arm-CNAs are a gain or loss of an entire chromosome arm, and one of the most common forms of structural rearrangements in tumour cells as identified by two pan-cancer genomic analyses (Beroukhi *et al.*, 2010; Kim *et al.*, 2013).

arm-CNAs can conceivably result from at least three mechanisms during cell division (**Figure 3.1**). First, unrepaired DSBs at or near centromeres can result in a loss of a chromosome arm upon cell division. Second, DSBs at or near centromeres can initiate NHEJ-mediated translocations, resulting in whole-arm chromosomal translocations and eventually arm-CNAs. Finally, non-allelic homologous recombination (HR) repair between non-allelic repeats can result in unbalanced arm translocations, ultimately forming arm-CNAs. Importantly, all three mechanisms rely on an initiating double-stranded or single-stranded DNA break. NHEJ requires two DSB ends that can be directly ligated. HR, on the other hand, can be initiated by many different forms of DNA damage including DNA DSBs and SSBs (Li and Heyer, 2008; Maizels and Davis, 2018) (**section 1.1.1.4**).

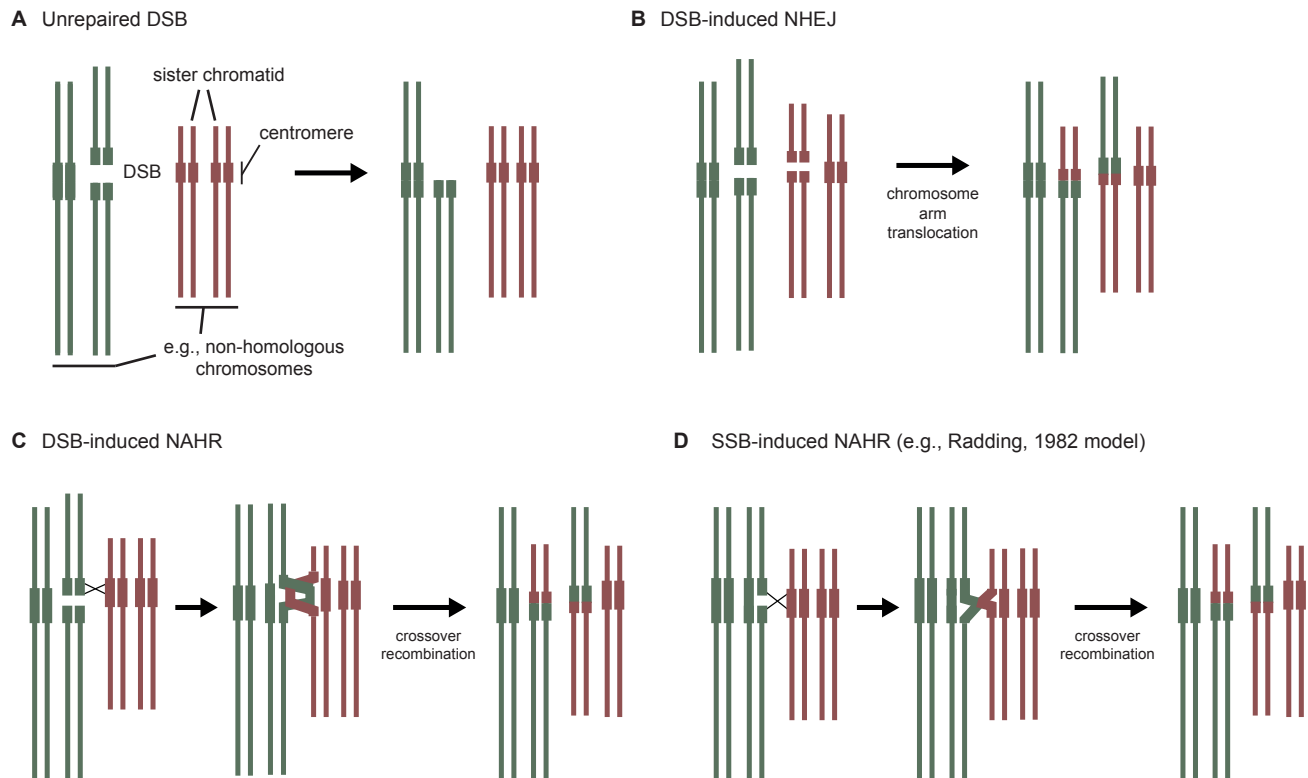


Figure 3.1: Mechanisms of arm-CNAs due to centromeric breakage. A depiction of three potential mechanisms by which chromosome arm copy-number alterations (CNAs) can conceivably form. Each panel contains two non-homologous chromosomes, with the centromeres marked in bold. (A) Unrepaired double-strand breaks (DSBs) at or near centromeres can produce an acrocentric fragment that is lost upon cell division. (B) Two DSBs can be incorrectly ligated to induce chromosome arm translocations. (C, D) DSBs or single-strand breaks (SSBs) can promote homologous recombination (HR). If between non-allelic sequences, non-allelic HR (NAHR) can result in chromosomal arm rearrangements, once again forming arm-CNAs upon cell division.

3.1.2 Current understanding of centromeres as fragile sites

While genomic rearrangements are often observed originating at (peri)centromeres, our understanding of the relevance, nature, and source of centromeric DNA instability is still in its infancy. For example, while arm-CNAs are frequently observed in tumor samples, it remains unclear whether arm-CNAs are an active driving force of genome instability or an inadvertent outcome of global genome instability.

As comprehensively discussed in **section 1.3.5**, there is no clear consensus on what induces (peri)centromere-derived rearrangements, but several broad hypotheses exist based on either the atypical DNA or chromatin structure at (peri)centromeres. First, secondary DNA structures at centromeres can theoretically interfere with replication progression as well as act as targets for structure-specific endonucleases (**section 1.3.5.1**). Second, the specialized chromatin environment at centromeres may adopt unusual structures, rendering centromeres inherently more susceptible to DNA breakage (**section 1.3.5.2**). Despite some

preliminary data supporting the feasibility of these hypotheses, it remains largely unclear why centromeres are inherently more fragile than the rest of the genome, and frequent drivers of genome instability (e.g., through arm-CNA formation).

In this chapter, I set out to first investigate the prevalence of arm-CNAs, a signature outcome of centromere instability, in early carcinogenesis, as a proxy for arm-CNAs potential role in driving human disease. Second, I set out to determine whether centromeres are indeed more prone to DNA damage than the rest of the genome, in the form of single- or double-stranded DNA breaks. Third, I worked to investigate potential sources of DNA instability acting at human centromeres. Through these studies, important questions are raised about the DNA damage signaling responses acting at human centromeres, which are further explored.

3.2 Results

3.2.1 Investigating spontaneous DNA fragility originating at human centromeres

3.2.1.1 Chromosome arm gains and losses as early drivers of carcinogenesis

It has been known for many years that chromosome arm gains and losses, as well as DNA breaks originating at centromeres, occur frequently in tumours (Barra and Fachinetti, 2018; Black and Giunta, 2018; Kim *et al.*, 2013). However, an outstanding question is whether such arm-CNAs contribute to the onset of cancer or whether they are an indirect result of a hyper-mutagenic cancer environment. For some insight into this question, I analysed data from ‘*The evolutionary history of 2,658 cancers*’ as part of the Pan-Cancer Analysis of Whole Genomes (PCAWG) (Gerstung *et al.*, 2020). The authors of this paper computationally reconstructed the life history of mutational processes across 2,658 cancer samples to identify the early somatic drivers of carcinogenesis, integrating both gene mutations and structural copy-number changes. Through this analysis, mutation events were classified as ‘early’, ‘intermediate/-variable’ and ‘late’. Compiling all ‘early’ driver events across 38 tumour types in 2,658 cancer samples, I first calculated the proportion of tumour samples containing each event as an early driver. These were further classified as chromosome arm loss/gain events (e.g., loss 17p, gain 3q), base mutations (e.g., TP53, APC), small structural variations (e.g., -19q13.43, + 7p11.2) and whole chromosome loss/gains (e.g., -6, +7) (**Figure 3.2A**). While several known drivers such as *TP53* and *KRAS* mutations were identified as frequent early mutation events, there was also a striking frequency of chromosome arm loss/gain. Loss of 17p, 9p and 3p were particularly frequent events, with the loss of 17p being an early driver event in more than 60% of all tumour samples. Compiling the relative frequency of each class (**Figure 3.2B**), it

became apparent that the most frequent class of mutation event was in fact chromosome arm loss/gain. Chromosome arm copy-number alterations therefore appear to be not only highly prevalent in tumours with an already mutated genome, but frequent drivers of early carcinogenesis. This ultimately suggests that chromosome arm copy-number alterations occur frequently and spontaneously (i.e., without an already perturbed genome), and highlights the importance of understanding the inherent DNA fragility at human centromeres.

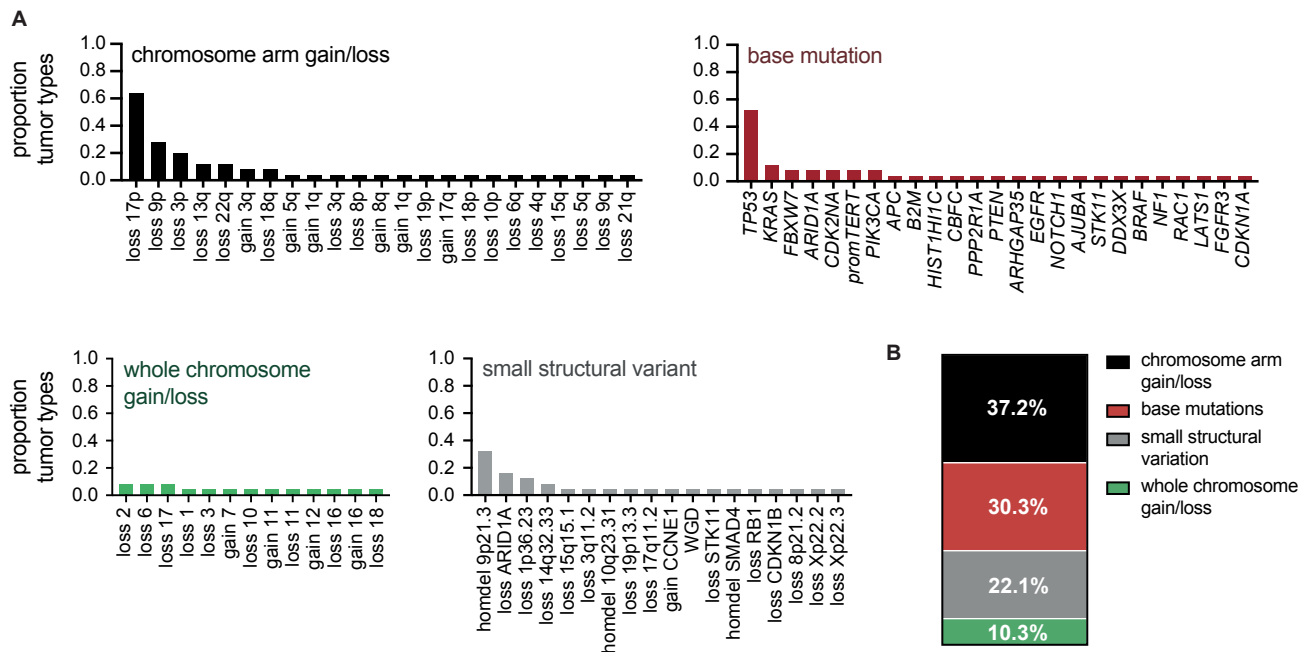


Figure 3.2: **Investigating arm-CNAs as drivers of carcinogenesis.** (A) Analyzing publicly available data (Gerstung *et al.*, 2020), somatic cancer mutations identified as ‘early’ were aggregated and classified as chromosome arm loss/gain, base mutation, whole chromosome gain/loss or small structural variant. Within each class, the proportion of tumor types exhibiting a particular early mutation is depicted. (B) Total proportions of early mutations corresponding to each class described in (A).

3.2.1.2 Spontaneous DNA breaks at human centromeres

Given that chromosome arm-CNAs are recurrent events during early carcinogenesis, I wanted to investigate what drives such potent DNA instability apparent at human centromeres. As described above, arm-CNAs can result from at least three mechanisms, all entailing DNA breaks as initiating substrates. Therefore, I wanted to determine if, and under which conditions, DNA breaks occur at human centromeres. In recent years, next-generation-sequencing (NGS)-based methods such as GLOE-seq and END-seq (and other methods recently reviewed elsewhere (Bouwman and Crosetto, 2018)), have been developed to precisely label broken DNA ends *in vitro* and identify them at high resolution. Although repetitive regions such as alpha satellite arrays are typically excluded from NGS analyses due to the lack of an appropriate reference sequence, the recent completion of a full human sequence reference now

enables the alignment of short-read sequencing dataset to centromeric arrays (Nurk *et al.*, 2021).

To address this question, I analyzed publicly available datasets mapping the genome-wide distribution of either SSBs (GLOE-seq) (Sriramachandran *et al.*, 2020) or DSBs (END-seq) (Wong *et al.*, 2021) (**Figure 3.3A**). Briefly, GLOE-seq identifies 3'hydroxy (OH) ends on DNA through the ligation of biotinylated adaptors. Biotin capture, adaptor ligation and NGS analysis enables the detection of free 3' ends of SSBs as well as 3' ends of DSBs, with the proximity of end detection distinguishing between SSBs and DSBs. Alternatively, END-seq relies on direct *in situ* labeling of blunted DSBs with biotinylated linkers, thereby selectively identifying DSBs following biotin capture, adaptor ligation and NGS analysis (**Figure 3.3A**). As a marker for active centromeres, data from a CENP-A localization study was also included (Thakur and Henikoff, 2018).

To this end, I developed a customized pipeline to enable NGS analysis at human centromeres (**Figure 3.3B**). In this pipeline, reads from candidate datasets are first aligned to the newly released T2T-CHM13 reference genome (Nurk *et al.*, 2021). To prevent artificial amplification of the signal across repetitive elements, any alignments that can map to multiple locations with equal confidence, as would be the case at repetitive elements, are only reported once. However, because multi-mapping alignments are retained to assess potential enrichment across repetitive elements, this pipeline circumvents conventional quality filters (i.e., MAPQ filtering). Therefore, to ensure that alignments have high alignment scores, only high-stringency alignments (i.e., 0 mismatch tolerance) are reported. Second, the reads are normalized to the total number of sequencing reads (reads per thousand million, RPKM) to correct for differences in sequencing depth and directly compare read densities between datasets. The resulting read densities are then plotted to visualize chromosomal patterns or further processed to calculate centromeric 'enrichment scores' for a more quantitative comparison between datasets (as depicted in (**Figure 3.3B**)). To calculate centromeric enrichment scores, the sum of uniquely mapped reads within each centromere are compared to an input sample of the corresponding cell line.

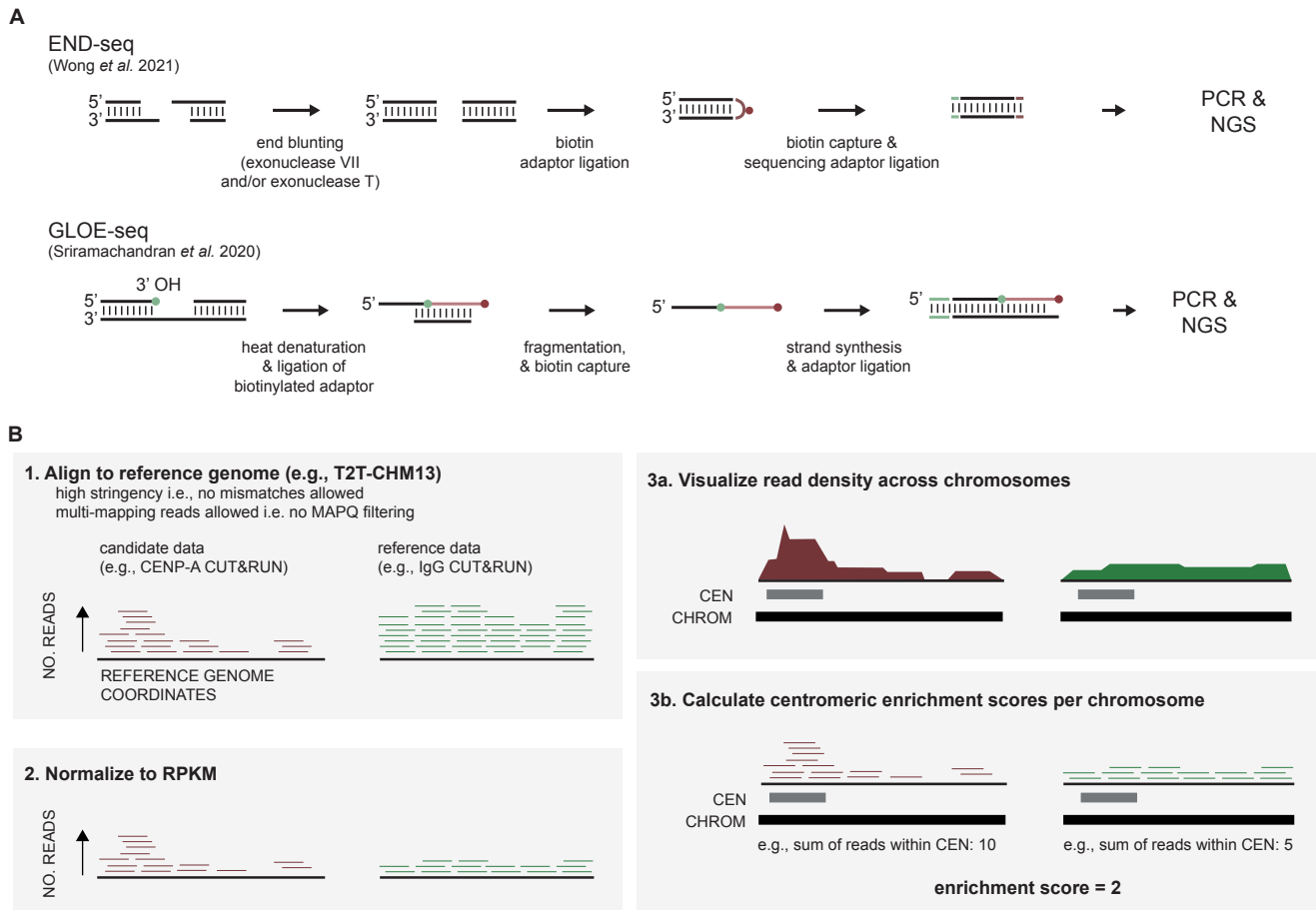


Figure 3.3: **Detecting DNA breaks at human centromeres.** (A) Experimental schematic of GLOE-seq (Sriramachandran *et al.*, 2020) and END-seq (Wong *et al.*, 2021) for the detection of DNA single-strand breaks and DNA double-strand breaks genome-wide, respectively. (B) Computational pipeline for the analysis of next-generation sequencing (NGS) datasets to T2T-CHM13. After read alignment and normalization, NGS read densities can be visualized or further processed to calculate centromeric enrichment scores per chromosome.

Surprisingly, this analysis revealed that SSBs, but not particularly DSBs, were strikingly enriched across centromeric satellite arrays (**Figure 3.4A**). Calculating centromeric enrichment scores for a more quantitative and genome-wide comparison, it was evident that SSBs were enriched at centromeres across all chromosomes, whereas DSBs were not (**Figure 3.4B**).

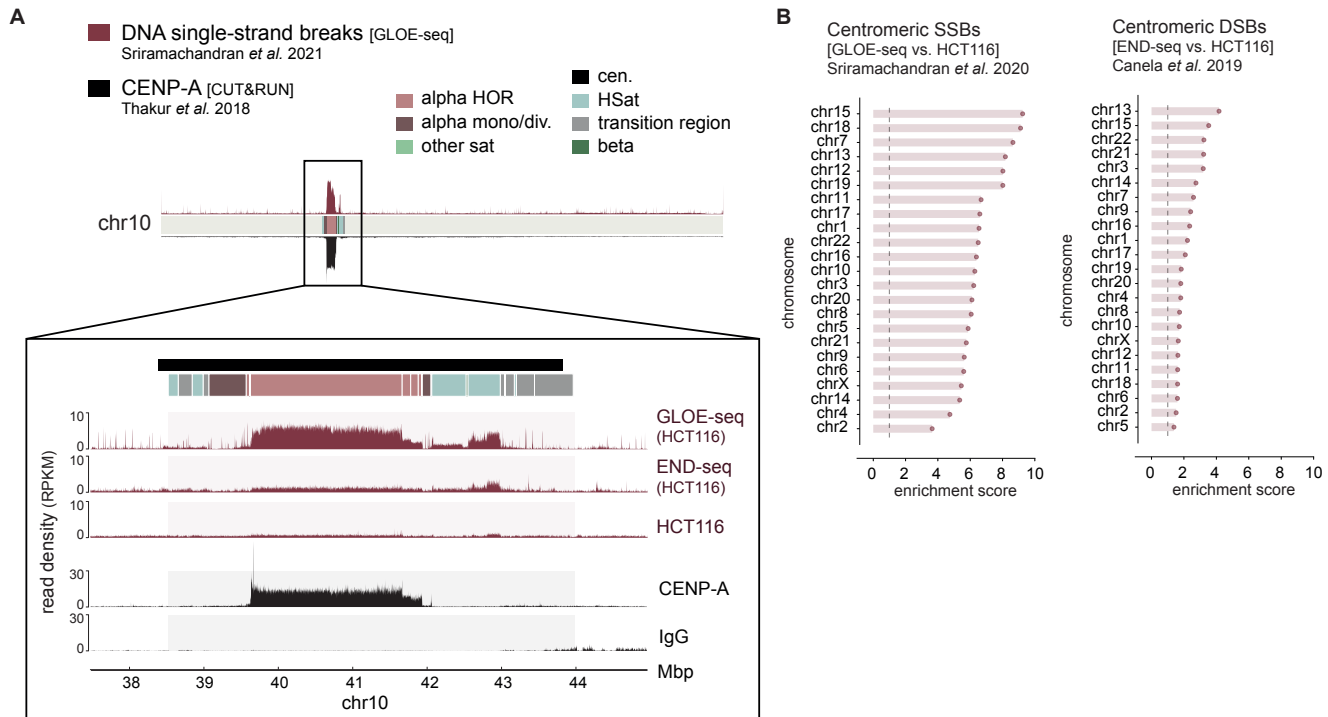


Figure 3.4: DNA single strand breaks at human centromeres. (A) Alignment of short-read sequencing data detecting HCT116 DNA single-strand breaks (GLOE-seq) (Sriramachandran *et al.*, 2020), HCT116 DNA double-strand breaks (END-seq) (Canela *et al.*, 2019) and CENP-A localization (CUT&RUN) (Thakur and Henikoff, 2018) across the complete sequence of chromosome 10 of the T2T-CHM13 reference assembly. Centromeric regions (black bar) are composed of alpha satellites within higher order repeats (HORs), monomeric / diverged alpha satellites, human satellites (HSat), beta satellites, transition regions and other satellites. Inset of chr10 centromere depicts the alignment of GLOE-seq and END-seq with the corresponding input cell line, as well CENP-A CUT&RUN with the corresponding negative control (IgG). Read density is calculated as the reads per kilobase per million mapped reads (RPKM), smoothed using a sliding window for visualisation. (B) Enrichment scores following GLOE-seq or END-seq across human autosomal centromeres. Enrichment scores were calculated as the sums of mapped reads within centromeres relative to that of the corresponding input control, after normalizing to read depth. Chromosomes are organized by descending enrichment score.

A direct comparison between DNA break and CENP-A localization genome-wide revealed a similar pattern of enrichment across chromosomes, with an additional accumulation of DNA breaks across other satellite DNAs, rDNA loci (i.e., p arms of chr13, chr14, chr15, chr21, chr22) and telomeres (**Figure 3.5A**). To determine precisely which satellite DNAs are susceptible to DNA breakage, I extracted the genomic coordinates of all human satellite annotations from RepeatMasker and asked whether DNA breaks were enriched across each given satellite annotation. To this end, the percentage of GLOE-seq reads aligning to a given satellite was quantified relative to the percentage genome occupancy of that satellite (i.e., the cumulative size of the satellite annotation) (**Figure 3.5B**). As a control, the same analysis was performed using a CENP-A localization dataset, and CENP-A was predominantly enriched at centromeric alpha satellites. Similarly, DNA breaks appeared to enrich predominantly at alpha satellite

repeats, although there was also a notable enrichment at HSat1/2 (HSATI, HSATII) repeats, which are found in pericentromeric regions. There was no enrichment of DNA breaks evident at any other type of human satellite DNA.

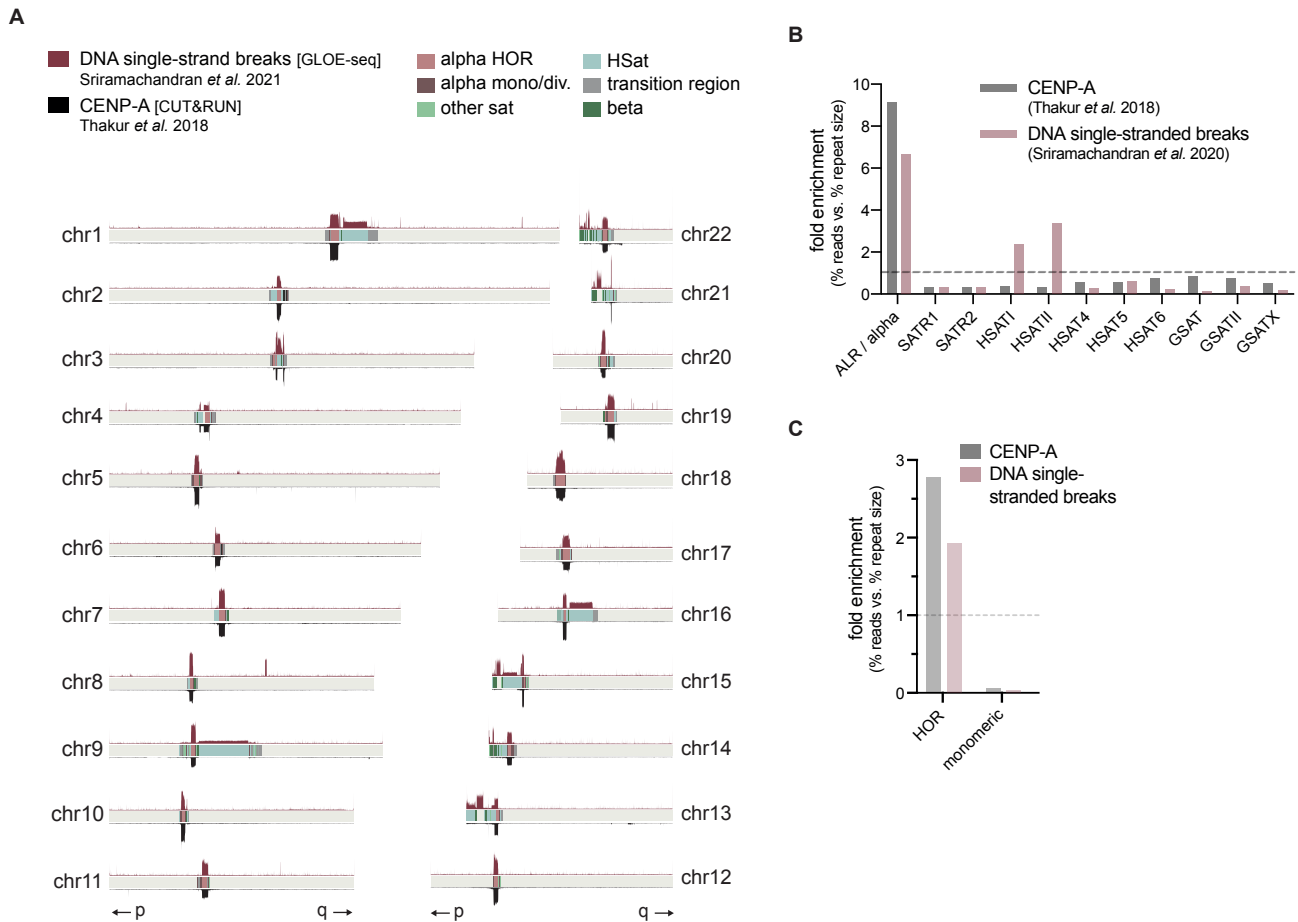


Figure 3.5: Investigating the distribution of DNA breaks and CENP-A across human centromeres. (A) Same as in **Figure 3.4A** but displaying SSB positioning (GLOE-seq) (above) and CENP-A positioning (CUT&RUN) (below) across all autosomal chromosomes of T2T-CHM13. p and q chromosome arms indicated below. (B) Fold enrichment of GLOE-seq and CENP-A CUT&RUN reads across annotated satellite repeats of the T2T-CHM13 reference sequence. Fold enrichment was calculated by calculating the percentage of sequencing reads falling within the coordinates of an annotated repeat type relative to the cumulative size of each repeat type, as a percentage of the full genome. (C) Fold enrichment of GLOE-seq and CENP-A CUT&RUN reads as calculated in (B) but over centromeric alpha satellites (ALR/alpha) annotated as being either HOR-associated or monomeric / diverged.

Upon closer inspection, the centromeric DNA breaks appeared to coincide specifically with alpha satellites found within active centromere HORs, rather than individual alpha satellite monomers located towards the periphery of active centromeres (**Figure 3.5A**). To verify this observation quantitatively, I used T2T-CHM13 centromere annotations to extract the genomic coordinates of alpha satellites classified as either ‘monomeric’ or part of HOR arrays. Once again normalizing to the cumulative size of monomeric/HOR alpha satellites, enrichment scores were calculated for both GLOE-seq and CENP-A datasets (**Figure**

3.5C). Indeed, there was a notable enrichment of reads within HOR alpha satellite arrays, with an under-enrichment across monomeric satellite repeats.

3.2.1.3 A novel assay for the detection of DNA breaks at repetitive elements

Given the unexpected extent of DNA break enrichment at centromeres, I wanted to validate and further study this phenomenon in an independent assay without the possible cell-to-cell variation of ensemble datasets. As discussed in **section 1.2.4**, several molecular biology and microscopy techniques exist for the detection of DNA breaks on both a global and locus-specific level. However, these techniques are either based on ensemble datasets (i.e., without single cell data), or they are only able to compare the extent of DNA breaks between experimental conditions. Therefore, I sought to find a technique with which to verify a potential baseline enrichment of DNA breaks at centromeres.

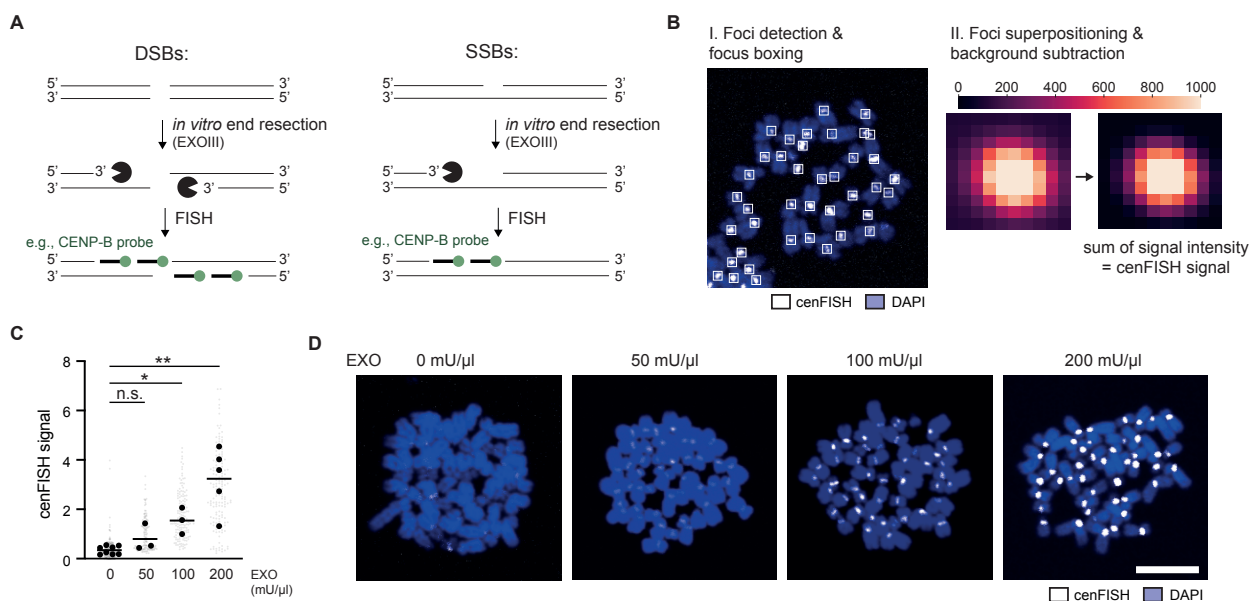


Figure 3.6: exo-FISH for the detection of centromeric DNA breaks. (A) Schematic of the detection of DNA single- and double-stranded breaks at repetitive elements using exo-FISH. (B) Depiction of exo-FISH quantification. Foci are automatically detected and super-positioned for each cell, generating a ‘median’ cell focus. FISH signal intensity is calculated as the sum of the FISH fluorescence signal within the median focus following a perimeter-estimated background subtraction. (C, D) Quantification and representative images of pooled biological replicates depicting the correlation between EXO concentration and FISH signal intensity using a centromere probe (cenFISH). cenFISH and DAPI signals are shown in white and blue, respectively. Scale bar represents 10 μm. FISH signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median value of a cell, with the median value of each biological repeat indicated in darker hues. At least 20 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (*p<0.05).

To this end, I developed a method with which to directly detect DNA breaks at defined repetitive loci of the genome in single cells. This method (hereafter named exo-FISH) relies on an *in vitro* end resection by Exonuclease III (ExoIII) on fixed cells. ExoIII digests DNA from 3’ ends, using either

single- or double-strand DNA breaks as an initial substrate, thereby exposing single-strand DNA. This single-stranded DNA can then be hybridized with a fluorescently labelled complementary probe using fluorescence in-situ hybridization (FISH) (**Figure 3.6A**). Any change in the number of DNA breaks would thereby initiate more ExoIII digestion, resulting in an increase in the exposed ssDNA and ultimately a higher FISH signal intensity. To quantify FISH signal intensities, I developed an automated microscopy quantification pipeline based on previous work (Bodor *et al.*, 2012), which will be adopted throughout this study (**Figure 3.6B**). First, foci are automatically detected using a signal threshold determined by eye and consistently applied throughout all image quantification. A 10x10-12x12 pixel box is then drawn around each focus. For each cell, all foci are super-positioned to generate a ‘median’ focus per cell. The background signal intensity is estimated based on the perimeter signal readings of the median focus, and the estimated background is subtracted from each pixel. Finally, a single signal value for each cell is calculated based on the sum of the fluorescence signal within the background-subtracted median focus, thereby encapsulating both focus size as well as intensity.

To avoid any potential artefacts due to the inherent DNA instability of cancerous cell lines, exo-FISH was optimized and primarily used in the diploid, non-transformed, hTERT-immortalized retinal pigment epithelial-1 (hTERT-RPE1) cell line. First, the concentration of ExoIII was optimized in hTERT-RPE1 to identify a concentration at which cenFISH foci in untreated controls (i.e., 0 mU/μl EXO) were still detectable without saturating the signal in ExoIII-treated samples. Although there was some experimental variation in the FISH signal intensity between biological replicates, a comparison of exoFISH signals between different experiments revealed that the median cenFISH signal per cell generally correlated with increasing ExoIII concentrations (**Figure 3.6C, D**). This confirmed that using ~100-200 mU/μl ExoIII allows for the detection of centromeric foci in non-treated conditions while still maintaining dynamic range for the detection of any changes in FISH signal intensity.

To validate that exo-FISH is able to detect both DSBs and SSBs at human centromeres, hTERT-RPE1 cells were collected in mitosis and fixed before inducing either DSBs or SSBs *in vitro* using the restriction enzyme BsmAI or the DNA nicking enzyme Nt.BsmAI, respectively. BsmAI and Nt.BsmAI generate DNA breaks within human centromeres every 895 bps, on average. In the absence of ExoIII treatment, neither BsmAI nor Nt.BsmAI pre-treatment had any effect on the centromeric FISH signal. In the presence of ExoIII treatment, pre-treatment with either BsmAI or Nt.BsmAI significantly increased the signal intensity of the centromeric FISH probe (**Figure 3.7A-D**). The verification using BsmAI was also performed in interphase cells to verify that exo-FISH could be applied to non-mitotic cells (**Figure 3.7E, F**). The increase in signal intensity selectively upon ExoIII treatment therefore confirmed that exo-FISH can be used to detect DNA breaks at centromeres.

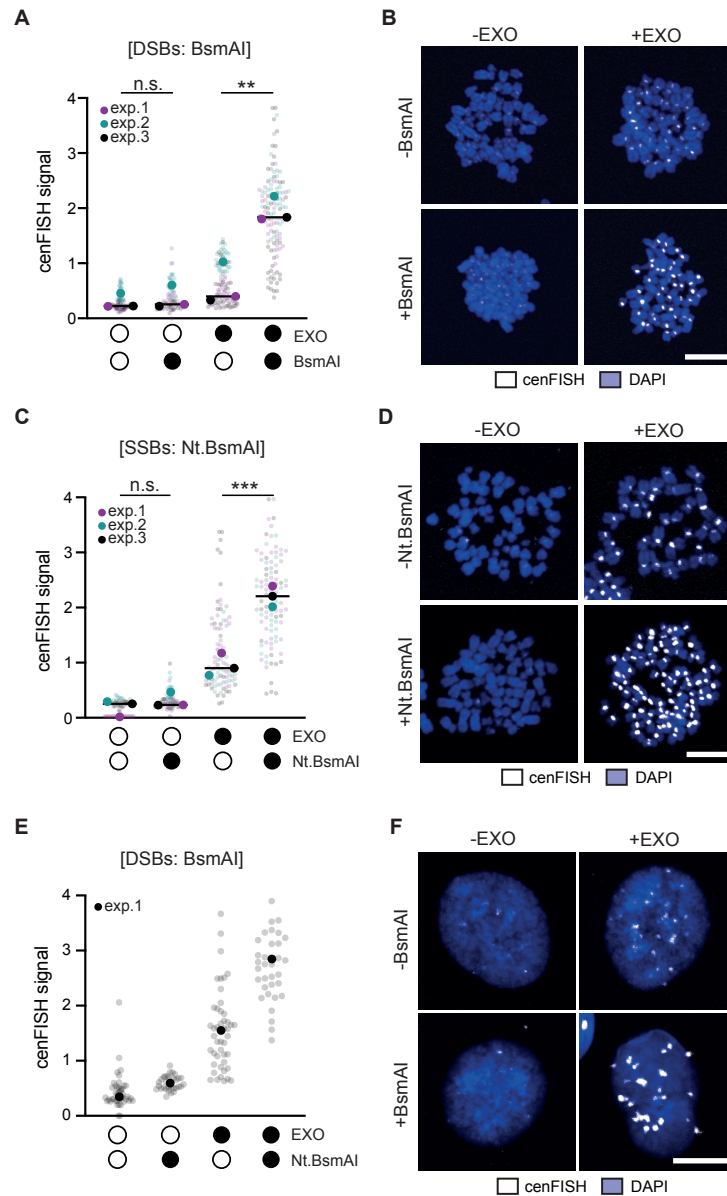


Figure 3.7: Validation of exo-FISH. (A-D) Quantification and representative images of DNA break detection by exo-FISH. hTERT-RPE1 cells were arrested in mitosis by S-trityl-L-cysteine (STLC), harvested and fixed. Double-strand breaks (DSBs) or single-strand breaks (SSBs) were induced by restriction enzyme digestion using BsmAI or Nt.BsmAI, respectively. Experimental controls not including Exonuclease III (EXO) digestion or restriction enzyme digestion are included. Quantification was performed as in **Figure 3.6B**. (E, F) As in A-D, using hTERT-RPE1 asynchronous cells with or without BsmAI treatment and EXO treatment (200 mU/μl). Quantification was performed as in **Figure 3.6B**. Scale bar represents 10 μm. FISH signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median value of a cell, with the median value of each biological repeat indicated in darker hues. At least 20 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$).

While optimizing exo-FISH, I observed a consistent increase in the centromeric FISH (cenFISH) signal upon ExoIII treatment, suggesting the presence of spontaneous DNA breaks at human centromeres even

in unperturbed conditions (**Figure 3.6C**, **Figure 3.7A, C, E**). This was always the case with the exception of **Figure 3.7A**, where centromeres appeared to be less responsive to ExoIII treatment in the absence of BsmAI. However, the normal responsiveness to ExoIII treatment may be masked in this experiment series, as microscopy settings were amended to capture the FISH signal upon BsmAI-induced DSBs without saturating the signal. As a result, smaller differences between the ExoIII- and ExoIII+ samples may no longer be readily apparent. To ensure that the ExoIII-dependence of centromere FISH signals was not an artefact of the protocol itself, I included a FISH probe complementary to telomeric sequences (telFISH) as an internal control. Indeed, there was no change in telFISH signal intensity upon ExoIII treatment, suggesting that these spontaneous breaks are indeed specific to centromeres (**Figure 3.8A, B**). This trend was equally prevalent in interphase (i.e., asynchronously dividing, non-mitotic) cells (**Figure 3.8C, D**), further suggesting that centromeric DNA breaks are not restricted to mitosis or overtly dependent on cell-cycle stage.

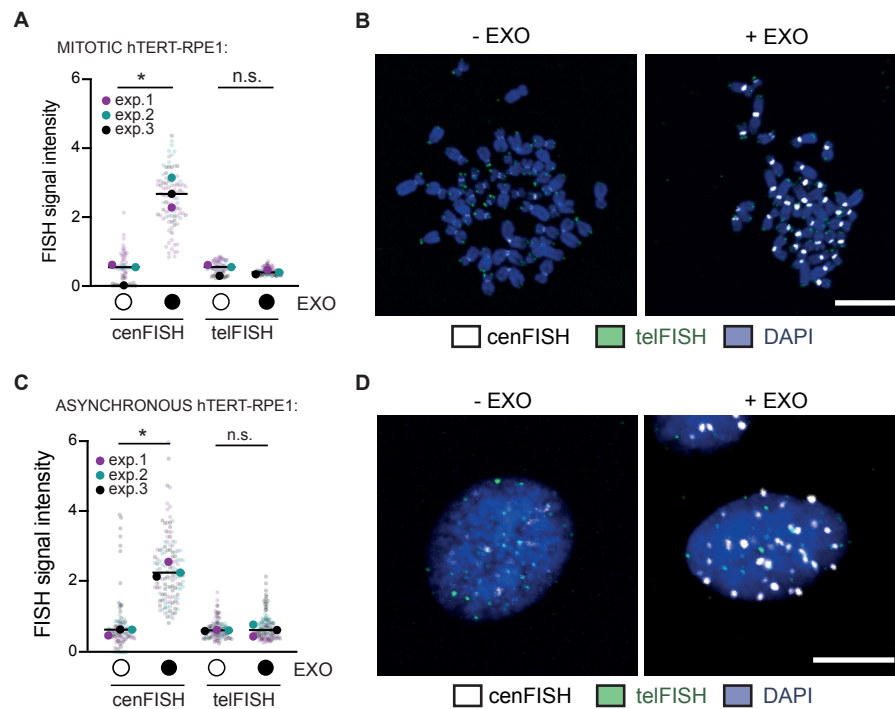


Figure 3.8: Characterizing spontaneous DNA breaks at centromeres. (A, B) Quantification and representative images of exo-FISH applied to mitotic hTERT-RPE1 cells with and without Exonuclease III (EXO) treatment. FISH signal intensity was calculated as described in **Figure 3.6B** for both centromeric (cenFISH) or telomeric (telFISH) FISH foci. The same threshold was used to define cenFISH and telFISH foci. cenFISH, telFISH and DAPI signals are shown in white, green, and blue, respectively. (C, D) Same as in (A, B) but in hTERT-RPE1 asynchronous cells. Quantification performed as described in **Figure 3.6B**. Scale bar represents 10 μ m. FISH signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median value of a cell, with the median value of each biological repeat indicated in darker hues. At least 20 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$).

3.2.1.4 Centromere-specific chromatin in spontaneous centromeric DNA breakage

As described above, centromeres are unusually complex regions of the genome on both a genetic and epigenetic level. Genetically, centromeres are composed of highly ordered tandem repeats that are prone to secondary structure formation (Kasinathan and Henikoff, 2018). Epigenetically, centromeres contain a highly specialized histone variant CENP-A that adopts unusual conformations within DNA (Takizawa *et al.*, 2020). In addition, CENP-A also precipitates the recruitment of hundreds of proteins as part of the kinetochore plate, potentially subjecting centromeric chromatin to mechanical stress not experienced by the rest of the genome (Barnhart *et al.*, 2011; Hori *et al.*, 2013; Mendiburo *et al.*, 2011). Therefore, to distinguish whether these spontaneous breaks are related to the DNA structure or chromatin structure of centromeres, I asked whether the localization of CENP-A at centromeres affects the levels of spontaneous centromeric breaks.

To verify that CENP-A could be efficiently depleted in hTERT-RPE1 cells, asynchronous cells were fixed and stained for both CENP-A and CENP-B using immunofluorescence (IF). CENP-B, as a sequence-specific centromere binding protein, serves as a marker of satellite repeats to define centromere location. To quantify CENP-A intensity within CENP-B-defined centromeres, an approach is taken similar to the exo-FISH quantification (**Figure 3.6B**), except quantifying CENP-A intensities based on CENP-B foci positioning (**Figure 3.9A**). CENP-B, as a sequence-specific centromere binding protein, functions as an important internal control to ensure that the efficiency of immunofluorescence staining is consistent between experimental repeats. First, CENP-B foci are identified in an automated fashion based on a threshold arbitrarily determined by eye. A 10x10-12x12 pixel centromere ‘box’ is then drawn around each CENP-B focus. Second, the signal intensities of both CENP-B and CENP-A are measured within each CENP-B-defined box. The superpositioning of all foci within a given experimental condition then generates a ‘median focus’ for each channel, which can be plotted as a representative depiction of signal intensity within centromeres. Finally, to quantify the signal intensity with minimal interference from background staining, a background subtraction is performed based on the median perimeter reading. The sum of the background subtracted IF signal within each cell is plotted to visualize biological variation within each experimental condition (as in **Figure 3.9B**).

Using this pipeline, I verified that 96 hours of RNAi-mediated depletion of CENP-A was sufficient to reduce CENP-A levels almost entirely within CENP-B-defined centromeres (**Figure 3.9B, C, D**). The same long-term depletion strategy was then applied to assess the levels of centromeric breaks using exo-FISH. Depletion of CENP-A did not affect the levels of spontaneous centromeric DNA breaks (**Figure 3.9E, F**), suggesting that these breaks are largely independent of the unique chromatin structure of CENP-A-containing nucleosomes.

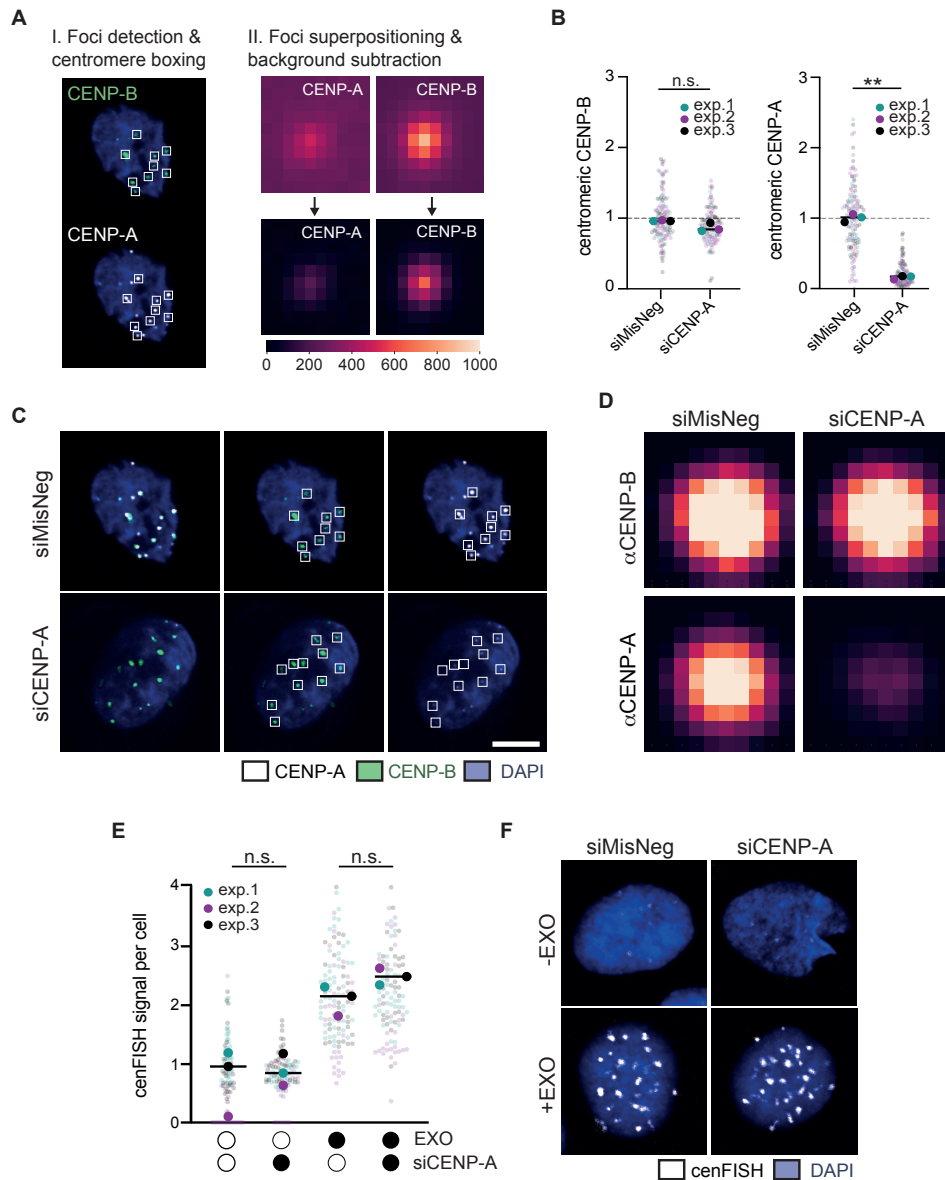


Figure 3.9: Investigating the contribution of CENP-A to spontaneous centromeric SSBs. (A) For the quantification of centromere-specific immunofluorescence signals, CENP-B foci were automatically detected after a threshold was arbitrarily assigned by eye, and 10x10-12x12 pixel boxes were drawn around CENP-B foci. The sum of signal intensities within the CENP-B-defined box were then quantified following a background subtraction based on perimeter readings. Representative ‘median foci’ are depicted here, displaying the median signal of all centromeres across all cells within a single representative biological repeat. (B, C, D) Quantification, representative images and representative median foci of immunofluorescence staining against CENP-A and CENP-B using the method described in **Figure 3.9A**. hTERT-RPE1 cells were fixed 96 hours after siRNA-mediated depletion of CENP-A. (E, F) Quantification and representative images of exo-FISH following the same experimental protocol as in (B, C, D). Quantification was performed as described in **Figure 3.6B**. Scale bar represents 10 μ m. FISH signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median value of a cell, with the median value of each biological repeat indicated in darker hues. At least 20 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$).

3.2.1.5 DNA replication in spontaneous centromeric DNA breakage

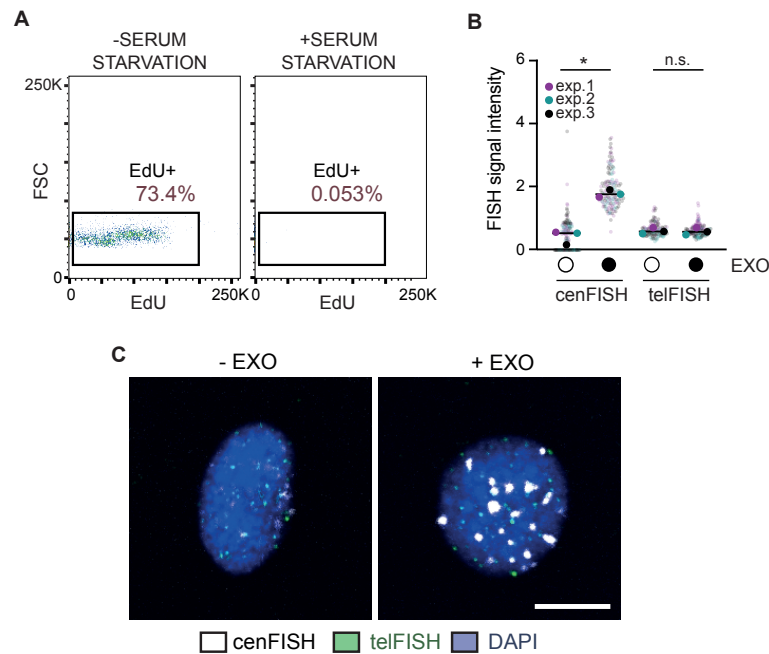


Figure 3.10: DNA replication as a potential source for spontaneous centromeric SSBs. (A) Validation of serum starvation halting cell proliferation. hTERT-RPE1 were incubated with serum-starved media (0.5% FBS) for 24 hours and then incubated with EdU for another 24 hours, as proliferating cells will incorporate EdU into DNA during replication. Cells were then fixed and analyzed by flow cytometry. Gates were generated using -serum starvation to define the EdU+ population. Forward scatter (FSC) is an indicator of cell size. (B, C) Quantification and representative images as in **Figure 3.8** but in hTERT-RPE1 serum-starved (i.e., non-dividing G1 cells). cenFISH, telFISH and DAPI signals are shown in white, green, and blue, respectively. Scale bar represents 10 μ m. FISH signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median value of a cell, with the median value of each biological repeat indicated in darker hues. At least 20 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$).

Given that the spontaneous centromeric DNA breaks appear to be independent of chromatin structure, I next turned to the underlying DNA at centromeres. Human centromeres are repetitive and tend to form secondary structures, suggesting they may be inherently difficult to replicate (Kasinathan and Henikoff, 2018), as supported by the observation that centromeres replicate late in the cell cycle (Erliandri *et al.*, 2014; Ten Hagen *et al.*, 1990). Therefore, I hypothesized that replication difficulties may underlie the inherent propensity for DNA breakage at centromeres. To address this hypothesis, I induced hTERT-RPE1 cells into a state of quiescence using serum-starved media to prevent any proliferation and, thereby, DNA replication. As little as 24 hours in serum-starved media was sufficient to entirely halt cell proliferation as detected by EdU incorporation (**Figure 3.10A**). Under these conditions, the levels of spontaneous centromeric and telomeric breaks were then assessed using exo-FISH (**Figure 3.10B, C**). The exo-FISH signals of cenFISH, but not telFISH, showed the same response to ExoIII treatment as normally divid-

ing cells (refer to **Figure 3.8**), suggesting that preventing DNA replication does not affect the levels of spontaneous DNA breaks at centromeres.

These investigations overall suggest that spontaneous centromeric breaks are not overtly dependent of DNA replication. While the source of spontaneous DNA nicks at centromeres remains to be identified, these data ultimately imply the presence of an alternative source of DNA breakage that is particularly enriched at human alpha satellite HORs and independent of CENP-A-defined chromatin environments.

3.2.2 Characterizing DNA damage signaling at human centromeres

3.2.2.1 DNA damage signaling in response to spontaneous centromeric DNA breaks

The spontaneous centromeric DNA breaks observed in non-transformed human cells was somewhat unexpected, since a baseline level of DNA fragility in unstressed cells had not previously been reported at human centromeres. In addition to direct break detection methods analyzed above (e.g., GLOE-seq, BLESS), DNA breaks have also been commonly detected using indirect methods, such as the locus-specific activation of the DNA damage response (DDR) by microscopy or chromatin immunoprecipitation. The current gold standard for DDR activation is S139-phosphorylated H2A.X (γ H2A.X), as H2A.X is robustly phosphorylated in response to both double-strand breaks and extended ssDNA gaps (Mah *et al.*, 2010; Sedelnikova *et al.*, 2002; Ward and Chen, 2001; Ward *et al.*, 2004).

To test whether spontaneous centromeric breaks can initiate canonical DNA damage signaling, I looked for an enrichment of γ H2A.X phosphorylation at centromeres. Analyzing γ H2A.X ChIP-seq performed in unperturbed cells, there did not appear to be any baseline enrichment of γ H2A.X across centromeres, as would be expected given the spontaneous enrichment of DNA breaks (**Figure 3.11A**). On a mechanistic level, this observation raises two distinct possibilities. The first possibility is that the single-strand breaks accumulated at centromeres are insufficient to induce DNA damage signaling. The second possibility is that centromeres are refractory to canonical DDR responses. To distinguish between these possibilities, I sought conditions under which DNA double-strand breaks at centromeres were induced such that the corresponding DDR could be analyzed.

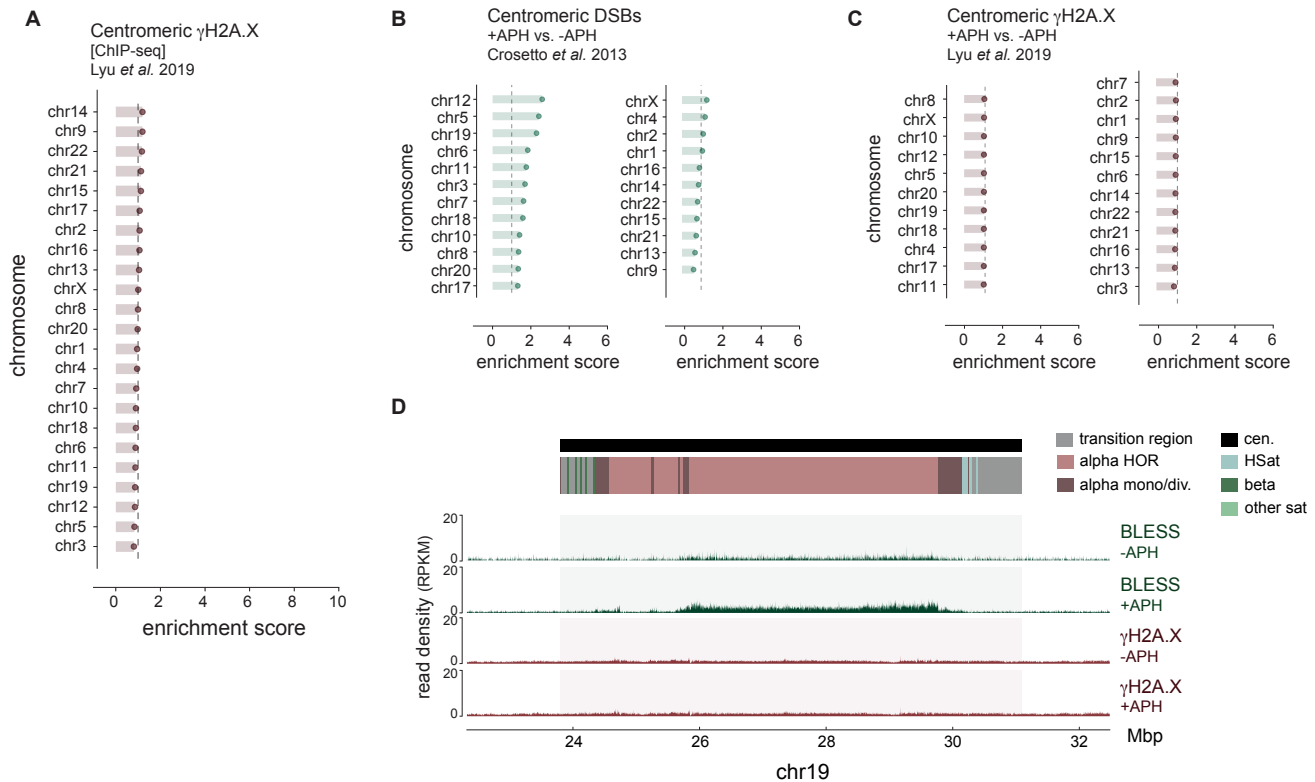


Figure 3.11: **Investigating DNA damage signaling at human centromeres.** (A) Enrichment scores of γ H2A.X ChIP-seq (Lyu *et al.*, 2019) across human autosome centromeres as calculated by the sum of mapped sequencing reads within centromeres relative to that of the corresponding input control, after normalizing to read depth, as in **Figure 3.3B**. Chromosomes are ranked by descending enrichment score. (B) Enrichment scores of DSB mapping (BLESS) (Crosetto *et al.*, 2013) across human autosome centromeres as calculated by the sum of mapped sequencing reads within centromeres following replication stress treatment (0.4 μ M aphidicolin; APH) relative to the non-treated control. Quantification as in **Figure 3.3B**. (C) Enrichment scores of γ H2A.X ChIP-seq (Lyu *et al.*, 2019) across human autosome centromeres as calculated by the sum of mapped reads within centromeres following replication stress treatment (0.3 μ M aphidicolin; APH) relative to the non-treated control. (D) Alignment of short-read sequencing data detecting DNA double-strand breaks (BLESS) (Crosetto *et al.*, 2013) with and without APH as well as γ H2A.X ChIP-seq (Lyu *et al.*, 2019) with and without APH across chromosome 19 of T2T-CHM13. Centromeric regions (black bar) are composed of alpha satellites within higher order repeats (HORs), monomeric / diverged alpha satellites, human satellites (HSat), beta satellites, transition regions and other satellites. Read density is calculated as the reads per kilobase per million mapped reads (RPKM), smoothed using a sliding window for visualisation.

3.2.2.2 DNA damage signaling capacity of centromeres

To understand the DDR capacity of centromeres, I wanted to determine whether γ H2A.X at centromeres could be detected under conditions whereby DNA breaks had been induced specifically at centromeres. In Crosetto *et al.* 2013, replication stress was induced with the drug aphidicolin (APH) and DNA double-strand breaks were detected genome-wide using BLESS. From this, the researchers identified aphidicolin-sensitive regions (ASRs), which were enriched for alpha satellite repeats. This suggested that DNA DSBs could be induced at centromeres under APH-induced replication stress. A study from another lab in 2019 similarly used a low-dose aphidicolin treatment, albeit in a different cell line, while mapping γ H2A.X

occupancy genome-wide using ChIP-seq. Analyzing both datasets, enrichment scores were calculated in treated (+APH) relative to untreated (-APH) conditions to determine how replication stress impacted both centromeric DSBs and γ H2A.X signaling. Indeed, this analysis confirmed that DNA DSBs were mildly induced upon APH treatment, whereas γ H2A.X levels at centromeres did not change (**Figure 3.11B, C, D**). Importantly, caution should be exercised since these were two independent studies from two independent labs, using different cell lines. However, these results served as an indicator that perhaps centromeres are in fact refractory to canonical DDR mechanisms.

To further investigate this hypothesis experimentally, I sought to induce conditions under which DNA DSBs could be induced at human centromeres to map both DNA breaks (using exo-FISH) and DDR activation. Although CRISPR-mediated breaks targeting alphoid repeats is currently in use by other groups and could precisely and efficiently induce DNA breaks, this approach would hypothetically generate tens of thousands of breaks within each centromere, hence exceeding a physiological representation of canonical DDR signaling thresholds. Therefore, conditions were chosen that could induce DNA breaks at centromeres at more physiologically representative levels. To this end, I took advantage of the natural enrichment of the human Topoisomerase II α (TOPII α), which acts at mitotic centromeres and generates transient DNA breaks while resolving topological stress (Rattner *et al.*, 1996). I first verified that TOPII α is enriched at centromeres in mitosis in hTERT-RPE1 cells by co-staining TOPII α and the centromere marker CENP-B. While TOPII α localization at centromeres was more diffuse than CENP-B foci, a clear enrichment at centromeres was observed (**Figure 3.12A, B**).

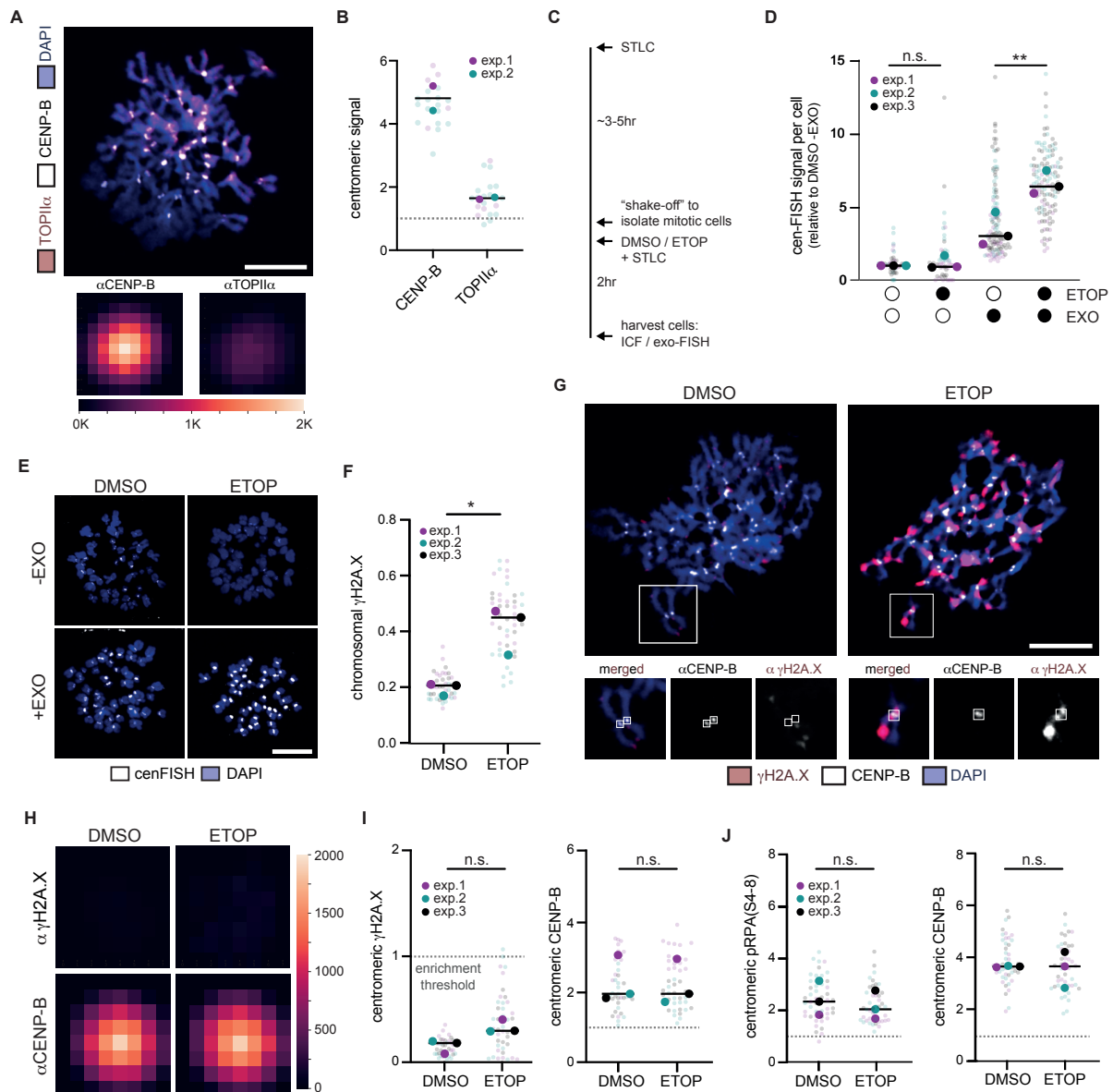


Figure 3.12: DNA damage signaling capacity at human centromeres. (A, B) Representative image and quantification of native immuno-cytofluorescence (ICF), detecting CENP-B and TOP1α at mitotic hTERT-RPE1 cells. Quantification performed as in **Figure 3.9A**. (C) Experimental schematic. hTERT-RPE1 cells were enriched in mitosis with S-trityl-L-cysteine (STLC) for ~3-5 hours. Mitotic cells were collected and treated with drugs (DMSO or Etoposide (ETOP)) for 2 hours in the presence of STLTC and harvested for ICF or exo-FISH. (D, E) Quantification and representative images of exo-FISH following 2 hours DMSO or ETOP treatment. Quantification as described in **Figure 3.6B**, and signal intensities normalized to the DMSO-EXO condition. cenFISH and DAPI signals are shown in white and blue, respectively. (F, G) Quantification and representative images of γH2A.X and CENP-B detection with ICF following 2 hours DMSO or ETOP treatment. Quantification of chromosomal γH2A.X was performed by calculating the sum of the total γH2A.X fluorescence intensity relative to the sum of the DAPI signal within each spread. (H) Median foci of γH2A.X and CENP-B within CENP-B defined centromeres. (I) Quantification of centromeric γH2A.X and CENP-B within CENP-B defined centromeres, quantified as in **Figure 3.9A**. (J) Quantification of centromeric pRPA(S4-8) and CENP-B within CENP-B defined centromeres, quantified as in **Figure 3.9A**. Arbitrary enrichment thresholds set at 10,000 A.U. Scale bar represents 10 μm. Each data point represents a median value of a cell, with the median of each biological replicate indicated in darker hues. IF and FISH signal intensities are X10,000 A.U., except for in (D) and (F). At least 10 (MSO-ICF) or 30 (cenFISH) cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (*p < 0.05, **p < 0.01).

Poisoning TOPII α with the drug Etoposide (ETOP) converts transient DNA breaks into permanent breaks to rapidly induce DSBs at centromeres in hTERT-RPE1 mitotic cells, resulting in conditions in which DDR activation can be assessed (**Figure 3.12C**). I first confirmed, using exo-FISH, that Etoposide treatment for 2 hours in mitotically arrested cells resulted in an increase in centromeric DNA breaks (**Figure 3.12D, E**). Under the same conditions, I then assessed DDR marker activation by quantifying the levels of γ H2A.X in response to Etoposide treatment. There was a notable induction of γ H2A.X signaling on chromosome arms in Etoposide-treated cells, as quantified by the sum of the signal intensity on DAPI-stained DNA, normalized to the total DAPI intensity of a given cell (**Figure 3.12F, G**). When quantifying the background-subtracted signal intensity within CENP-B-defined centromeres, there was no baseline enrichment of γ H2A.X at centromeres. Upon Etoposide treatment, there was a small but non-significant ($p=0.14$) increase in γ H2A.X staining (**Figure 3.12H, I**). An alternative marker for DNA breaks - phosphorylated RPA - was also tested to determine if the DDR suppression was specific to γ H2A.X. A spontaneous enrichment of phosphorylated RPA (pRPA) at mitotic centromeres was detected, as previously reported (Kabeche *et al.*, 2018). However, there was no change in the levels of centromeric pRPA upon DNA damage induction with Etoposide treatment, suggesting that other DDR markers are also not responsive to the DNA damage induced at centromeres (**Figure 3.12J**).

Overall, these results do not unequivocally address whether the spontaneous DNA nicks observed at centromeres can activate the DDR. However, these results do support the notion that DDR activation is at least partially suppressed at human centromeres. This may explain how the spontaneous DNA breaks at centromeres are able to accumulate over time without interrupting the cell cycle, and potentially why this centromeric fragility has not been previously detected.

3.3 Discussion

3.3.1 Investigating innate fragility at human centromeres

Structural abnormalities originating at centromeres have been of increasing interest in recent years (Barra and Fachinetti, 2018; Black and Giunta, 2018). Chromosome rearrangements and breaks originating at or near centromeric regions have been identified in numerous tumour samples (Adeyinka *et al.*, 2007; Alsop *et al.*, 2006; Backsch *et al.*, 2011; Balachandar *et al.*, 2007; Bièche *et al.*, 1997; Cooke *et al.*, 2008; Fernandez-Banet *et al.*, 2014; Ganly *et al.*, 2004; Hermsen *et al.*, 2005; Hermsen *et al.*, 1996; Lacle *et al.*, 2015; Le Baccon *et al.*, 2001; Martínez *et al.*, 2012; Martins *et al.*, 1999; Smedley *et al.*, 2000; Thompson *et al.*, 1997). Chromosome rearrangements originating at centromeres or pericentromeres can result in chromosome arm copy-number alterations, which are one of the most common form of copy-number alterations observed in a pan-cancer analysis of 3,131 cancer specimens (Beroukhi *et al.*,

2010). Furthermore, the data presented here suggest that arm-CNAs are not only frequent outcomes of highly mutagenic cancer environments but in fact act as potent drivers of early carcinogenesis (**Figure 3.3**). Centromeres therefore appear to act as drivers of genome instability. In this chapter, I set out to investigate if and why centromeres are inherently fragile regions of the genome.

3.3.1.1 Potential sources of centromere-derived structural abnormalities

Arm-CNAs can be an outcome of unrepaired DSBs, NHEJ-mediated translocations or HR-mediated rearrangements, all of which are initiated by either DNA DSBs or ssDNA gaps or nicks (collectively SSBs). I therefore set out to determine if, and under which conditions, centromeres exhibit DNA breakage, either in the form of DSBs or SSBs. Analyses of publicly available datasets revealed a notable accumulation of spontaneous DNA single-strand breaks, with a very mild accumulation of DSBs, across human centromere HORs as well as human satellites 1 (HSat1) and 2 (HSat 2).

To verify this observation and determine whether it was similarly the case in non-cancerous cell lines, I developed a microscopy-based approach (exo-FISH), which can detect both DSBs and SSBs at repetitive elements such as centromeres and telomeres. With exo-FISH, hTERT-RPE1 centromeres, but not telomeres, displayed a high degree of DNA breaks in unperturbed conditions, confirming the notion that DNA breaks accumulate spontaneously at human centromeres. Notably, it is perhaps unexpected that telFISH probes are unresponsive to EXOIII treatment because telomeres appear to accumulate some DNA breaks in the GLOE-seq analysis (**Figure 3.5**). However, one possibility is that the DNA breaks detected by GLOE-seq reflect only the ends of telomeres, and that telomere cap protection complexes prevent Exonuclease III digestion. Another possibility is that the total number of breaks at centromeres is much higher than breaks found at telomeres, and that exo-FISH is not sensitive to detect lower densities of DNA breaks, as might be at telomeres. Regardless, the consistent increase in cenFISH signals upon Exonuclease III treatment confirms that DNA breaks frequently occur at human centromeres. Centromeres in lower eukaryotes also accumulate DNA breaks at centromeres as detected by GLOE-seq (Sriramachandran *et al.*, 2020), but to my knowledge this is the first evidence that this spontaneous accumulation of centromeric DNA breaks also occurs in higher eukaryotes.

Given the predominance of SSBs over DSBs across human centromeres, it seems likely that SSBs are primarily present at centromeres, of which a variable proportion are converted to DSBs. Further, given that SSBs, which reflect both ssDNA gaps formed during active DNA replication and DNA nicks, can be repaired by HR, the predominance of centromeric SSBs further imply that HR is more active at centromeres in baseline conditions as compared to NHEJ, which predominantly acts on DSBs. It is tempting to speculate that this accumulation of DNA SSBs at centromeres may underlie the propensity for centromeric breaks and/or rearrangements that initiate the chromosome arm copy-number alterations

observed during early carcinogenesis.

3.3.1.2 A potential role for DNA SSBs in driving chromosomal rearrangements

It was initially surprising to observe the incidence of endogenous SSBs, rather than DSBs, at human centromeres. Centromeric sister-chromatid exchanges, one of the two major outcomes of HR, are detectable to some extent in almost all mitotic cells, suggesting that HR occurs frequently at centromeres even in unperturbed conditions (Giunta and Funabiki, 2017; Jaco *et al.*, 2008). Because HR is typically considered to a DSB repair pathway, the initial expectation was therefore that spontaneous DSBs at centromeres would be at least somewhat detectable. However, the data revealed very little enrichment of spontaneous DSBs at centromeres (**Figure 3.4**), and only a mild induction upon treatment of drug-induced replication stress (**Figure 3.11**). The disconnect between frequently observed SCEs and a paucity of centromeric DSBs implies that centromeric recombination could, in fact, be driven independently of DSB repair. Indeed, the high incidence of spontaneous SSBs at centromeres provides an attractive alternative explanation.

As discussed fully in **section 1.1**, HR has not always been considered a DSB repair pathway. Earlier models of HR posited single-strand nicks as initiating substrates for HR (**Figure 1.1**). Attention increasingly shifted towards DSB as initiating substrates following the Szostak model in 1983. However, more recent comparative studies using CRISPR-induced DNA breaks reveal that the originally proposed SSB-induced model is also viable (**section 1.1.1.4**). It is therefore not inconceivable that the endogenous SSBs at centromeres could drive HR and sister-chromatid exchanges seen at high baseline levels at human centromeres. In turn, if this is improperly regulated, recombination between inverted repeats or non-allelic chromosomes could in fact directly drive arm-CNAs.

3.3.1.3 The relative contribution of centromeric and pericentromeric breaks

One of the outstanding questions related to arm-CNA formation is whether the propensity towards rearrangements and/or DNA breaks near centromeres results from DNA fragility at centromeres or at pericentromeres. Indeed, there has been some ambiguity on this matter in recent decades, especially given the clear examples of structural abnormalities originating from pericentromeres. For example, Robertsonian translocations are a form of rearrangement involving acrocentric chromosomes (chr13, 14, 15, 21 and 22) that have been shown to break at pericentromeres regions (Gravholt *et al.*, 1992). In addition, the most common form of isochromosome (isochromosome X) involves pericentromeric, rather than centromeric, breakage (Wolff *et al.*, 1996). Neoplasia-associated isochromosome i17q also appears to originate from pericentromeric rearrangements (Barbouti *et al.*, 2004). And finally, whole arm translocation breakpoints in head and neck squamous cell carcinoma appear to originate from pericentromeres rather

than centromeric cores (Martínez *et al.*, 2012). However, most studies assessing (peri)centromere-derived rearrangements have lacked the spatial resolution to conclusively differentiate between centromeres and pericentromeres, and so their relative contribution remain unclear.

The analysis of DNA SSBs through GLOE-seq revealed an enrichment of DNA SSBs at both inner HOR alpha satellites and, to a slightly lesser extent, human satellite repeats 1 and 2, which are found at pericentromeric regions (Lee *et al.*, 1997). If it stands true that SSBs can drive rearrangement events through NAHR, these data imply that both centromeres and pericentromeres can contribute to rearrangements, potentially underlying the historical ambiguity about their relative contributions towards structural rearrangements. Moving forward, it will be imperative to determine whether pericentromeres and centromeres accumulate SSBs through similar mechanisms, and how downstream repair dynamics diverge between the two regions.

3.3.1.4 Investigating sources of spontaneous centromeric DNA breaks

Identifying the sources of spontaneous centromeric SSBs would be useful for two principal reasons. First, to better understand the inherent fragility of centromeres. Second, and perhaps more importantly, to experimentally interrogate the relationship between centromeric SSBs and downstream consequences including centromeric recombination and chromosomal rearrangements. While this work was not able to conclusively identify the source of spontaneous centromeric SSBs, significant progress has been made in understanding its origins.

Endogenous SSBs constitute up to 75% of the estimated 70,000 DNA lesions experienced per cell per day and can also result from a wide variety of cellular processes (Lindahl and Barnes, 2000). First, DNA replication produces small DNA nicks not only at the junctions but also larger ssDNA gaps at stalled or collapsed replication forks (Smith *et al.*, 2009; Sriramachandran *et al.*, 2020; Van *et al.*, 2010). Second, SSBs have also been associated with active transcription in both *S. cerevisiae* promoters and human transcription termination sites (Hegedüs *et al.*, 2018; Sriramachandran *et al.*, 2020). Third, single-strand cutting endonucleases such as Topoisomerase I and Topoisomerase II and various repair-associated endonucleases (e.g., XPF, XPG) generate SSBs. Topoisomerase I, for example, induces transient SSBs to relax both positive and negative DNA supercoiling ahead and behind elongating Polymerase II as well as initiate transcription (Pommier *et al.*, 2016). SSBs are also generated as repair intermediates by endonucleases during repair pathways such as the mismatch-repair pathway (MMR), nucleotide excision repair (NER) and ribonucleotide excision repair (RER) (Sassa *et al.*, 2019; Tubbs and Nussenzweig, 2017). XPF and XPG can also target persistent RNA-DNA hybrids to generate SSBs (Cristini *et al.*, 2019). Fourth, endogenous SSBs can be directly induced by replicative DNA transposases such as DDE transposases and HUH nucleases-like transposases (Y1-transposases and Y2-transposases), as well as domesticated

transposases (Arinkin *et al.*, 2019). Given the many potential sources of SSBs at centromeres, identifying the source of spontaneous SSBs was a challenging task.

First, I set out to determine whether spontaneous centromeric SSBs were related to the DNA structure or the CENP-A-defined chromatin structure of centromeres, both of which have atypical features. To interfere with centromeric chromatin structure, I depleted CENP-A until levels of CENP-A were hardly detectable at centromeres and found no notable change in the levels of spontaneous centromeric breaks (**Figure 3.9**). These data suggest that centromeric chromatin structure imbued by CENP-A chromatin occupancy probably does not underlie the susceptibility of centromeres to spontaneous breaks. An important caveat to this analysis is that I did not verify that kinetochore recruitment was fully abrogated under these conditions, so it remains inconclusive as to whether spontaneous centromeric breaks are also independent of kinetochore formation. These data further imply that the underlying DNA structure, rather than chromatin structure, may render centromeres susceptible to DNA breaks.

DNA structure (i.e., non-B-form DNA conformations and RNA-DNA hybrids) at centromeres can interfere with DNA replication as well as act as a target for various structure-specific endonucleases (fully described in **section 1.3.5**). Therefore, I next addressed the role of DNA replication in the formation of centromeric SSBs. An accumulation of secondary structures at centromeres can interfere with replication progression and result in mitotic errors (Li *et al.*, 2018), suggesting that centromeric SSBs may reflect ssDNA gaps exposed by stalled or collapsed replication forks. To determine whether the SSBs observed at centromeres are associated with ongoing DNA replication, I compared the levels of spontaneous centromeric breaks in serum-starved quiescent cells with that of actively replicating cells, using exo-FISH (**Figure 3.8**, **Figure 3.10**). Although no direct comparison was made, it was evident that there was no significant change in the responsiveness of the cenFISH signal to exonuclease treatment in the different experimental settings. This suggests that the breaks observed at centromeres are not strongly associated with replication-associated processes. An alternative possibility is that active DNA replication does in fact underlie centromeric SSBs, but that the repair kinetics are too slow to see an effect at the point of analysis. However, it should also be noted that SSBs at centromeres have also been observed in budding yeast (Sriramachandran *et al.*, 2020), where centromeres are replicated early in S-phase and lack complex repetitive structures (Furuyama and Biggins, 2007; McCarroll and Fangman, 1988). The observation of replication independence of human centromeric SSBs is therefore in line with the trends observed in *S. cerevisiae*.

An important next step will be to address the role of DNA transcription in the generation of spontaneous centromeric DNA breaks. In addition to hard-to-replicate regions, a recent review of NGS-based break detection datasets revealed that spontaneous DNA breaks are frequently enriched at actively transcribed

regions (Saayman and Esashi, 2021). Transcription-induced fragility could conceivably relate to both the action of transcription itself, as well as RNA-DNA hybrids that are associated with ongoing transcription. Indeed, centromeres have been shown to be both actively transcribed and spontaneously accumulate RNA-DNA hybrids, particularly during mitosis (Chan *et al.*, 2012; Liu *et al.*, 2015; McNulty *et al.* 2017; Kabeche *et al.* 2018). In future work, this possibility could be addressed using RNA polymerase degron systems to transiently and efficiently reduce centromeric transcription with minimal off-target effects, and subsequently assess the extent of centromeric DNA breakage. Alternatively, the levels of centromeric transcription may be specifically up- or down- regulated by tethering transcription activators or repressors to centromeres, as has been previously performed using human artificial chromosomes (Nakano *et al.* 2008), to determine directly how centromeric transcription impacts centromere DNA fragility.

Centromeric DNA structure can also act as a target for structure-specific nucleases. Yeast Rad51-Mre11 complexes have been shown to have endonuclease activity towards DNA hairpins and MRE11/RAD50 have been shown to be spontaneously enriched on centromeric DNA (Aze *et al.*, 2016; Trujillo and Sung, 2001). In addition, DNA2 cleaves centromere secondary structures (Li *et al.*, 2018) and Topoisomerase II α has been shown to directly target alpha satellite hairpins, providing further possibilities for the source of centromeric SSBs described here (Jonstrup *et al.*, 2008). Given these studies and the apparent independence of spontaneous SSBs from centromeric chromatin structure and DNA replication, it seems plausible that the source of centromeric SSBs may indeed be related to a structure-specific endonuclease activity targeting secondary structures (e.g., hairpins) at human centromeres. However, this hypothesis remains to be tested and there are several additional potential sources of centromeric SSBs including intermediates of DNA repair mechanisms, topoisomerase activity and RNA-DNA hybrids.

One additional intriguing yet untested source of DNA breaks involves CENP-B, which has a nickase domain homologous to transposases and which may still retain its nicking activity (Tudor *et al.*, 1992). Pogo-like transposases appear to have been domesticated independently at least three times in metazoans (Mateo and González, 2014), arguing that there may be some evolutionary advantage to DNA nicking activity at CENP-B-bound regions. Given the unexpected enrichment of DNA nicks on CENP-B-bound alpha satellites, it is tempting to speculate that these nicks are directly generated by CENP-B. Whether CENP-B nickase activity is truly still retained, and its relationship to spontaneous DNA breaks at centromeres, remains to be addressed.

3.3.2 Investigating DNA damage responses (DDR) at human centromeres

3.3.2.1 Evidence for DDR suppression at human centromeres

One of the most surprising outcomes of this study was the observation that γ H2A.X signalling may be suppressed at human centromeres. Indeed, prior to the development of exo-FISH, I and others spent considerable time and effort exploring conditions with which to detect DDR markers such as γ H2A.X as a read-out for centromeric DNA damage. Despite testing a variety of different techniques (ChIP-qPCR, immunofluorescence, PLA) in different experimental settings, I was unable to detect any convincing induction of the DDR at centromeres.

There are three lines of evidence supporting the notion of DDR suppression at centromeres (**Figure 3.11**, **Figure 3.12**). Firstly, while analysing publicly available datasets, I observed no spontaneous enrichment of γ H2A.X that would reflect the enrichment of spontaneous DNA breaks detected at centromeres. Secondly, under low-dose replication stress conditions in which DSBs at alpha satellite regions have previously been reported (Crosetto *et al.*, 2013), there was no induction of γ H2A.X at centromeres (Lyu *et al.*, 2019). And finally, inducing permanent centromeric DSBs by poisoning Topoisomerase II α , which is naturally enriched at mitotic centromeres, did not result in any notable increase in centromeric γ H2A.X signalling, despite a clear induction of γ H2A.X signalling on chromosome arms.

There are several possible caveats to be considered with these data. For example, it is conceivable that the absence of spontaneous centromeric γ H2A.X signalling might be because single-stranded DNA breaks, as detected by GLOE-seq and exo-FISH, are not sufficient to activate γ H2A.X signalling, as previously suggested (Vriend and Krawczyk, 2017). In addition, the datasets under comparison in both the first and second lines of evidence were acquired by different lab groups, in different cell lines and several years apart. Therefore, technical variation cannot be excluded as a reason for the discrepancy between DSBs and γ H2A.X induction. The experimental validation through TopoII α poisoning in mitotically arrested hTERT-RPE1 cells provides more direct evidence for the suppression of the DDR at human centromeres. Although there was a small increase in γ H2A.X staining within CENP-B-defined centromeres upon Etoposide treatment, it should be noted that this could be due to peripheral staining and appears under-enriched relative to DNA damage induction detected by exo-FISH and the chromosomal changes in γ H2A.X staining. However, it is worth consideration that exoFISH only assesses breaks at CENP-B-containing satellite repeats. Therefore, one possibility is that centromeric TopoII α is somehow resistant to Etoposide poisoning, and that the breaks generated at centromeres are relatively infrequent compared to surrounding regions, which is reflected in the pattern of γ H2A.X staining. While this possibility seems unlikely, it is nonetheless worth consideration.

These data conflict with several reports of γ H2A.X at centromeres using ChIP-based or immunofluorescence methods (Giunta *et al.*, 2021; Guerrero *et al.*, 2010; Lee *et al.*, 2014; Mendez-Bermudez *et al.*, 2018; Mendez-Bermudez *et al.*, 2021; Yilmaz *et al.*, 2021). For example, a recent paper detailing the role of CENP-A in preventing replication stress-induced DNA damage at centromeres reports an increase in centromeric γ H2A.X signal upon CENP-A depletion (Giunta *et al.*, 2021). It is plausible that the discrepancy between the data reported here, and these studies, is because these reports are primarily quantifying relative amounts, rather than definitive amounts, of γ H2A.X between experimental conditions. For example, while some detectable level of γ H2A.X can be found at centromeres in certain conditions, this may be largely under-enriched relative to the degree of actual physical breaks or DNA damage in the region. Indeed, Yilmaz *et al.*, 2021 found a spontaneous colocalization of γ H2A.X at centromeres, but γ H2A.X foci appeared to colocalize adjacent to, rather than directly overlapping, centromeres cores. More convincing microscopy-based reports of γ H2A.X at centromeres use CRISPR-induced breaks targeting satellite repeats (Tsouroula *et al.*, 2016; Yilmaz *et al.*, 2021), which would hypothetically generate tens of thousands of breaks within each centromere, hence exceeding a physiological representation of canonical DDR signalling thresholds.

In support of these observations, studies in *Arabidopsis* quantifying the activation of γ H2A.X at pericentric heterochromatin relative to the genome-wide distribution of H2A.X also report an under-enrichment of γ H2A.X at peri-centromeres following DNA damage, with which the authors conclude that pericentromeric heterochromatin may be refractory to DDR (Lorkovic *et al.*, 2017). In addition, earlier studies using γ H2A.X also report a remarkably lower detection of γ H2A.X at satellite 2 (pericentromeres) and alpha satellites (centromeres) in response to DNA damage as compared to euchromatin, although the authors of this study conclude this to mean that fewer DSBs were generated in these regions upon IR treatment (Karagiannis *et al.*, 2007). These data ultimately suggest that caution should be exercised in the use of γ H2A.X as a reliable read-out of DNA damage at centromeres, and that greater emphasis should be placed on more direct read-outs for DNA breaks. While the evidence reported here points to a suppression of the DDR (or at least γ H2A.X and phosphorylated RPA) at centromeres, further research explicitly exploring this possibility is warranted.

3.3.2.2 Potential mechanisms and implications of centromeric DDR suppression

There are several potential mechanisms underlying the suppression of γ H2A.X. The first possibility is that the lack of γ H2A.X induction at centromeres is due to inherently low levels of the histone variant H2A.X at centromeres. Indeed, previous reports have found a depletion of H2A.X near centromeres, as little as 25% of the genome average H2A.X signal in areas directly surrounding centromeres (Iacovoni *et al.*, 2010). Although it is unlikely that H2A.X is entirely depleted within centromeres, this could

be one potential contributing factor towards the diminished levels of γ H2A.X induction at centromeres. The second possibility is based on reports that γ H2A.X spreading is inhibited by the presence of cohesin (Caron *et al.*, 2012), thereby restricting its localization between topologically-associated domains (Arnould *et al.*, 2020). Cohesin loading is enriched at simple centromeres of *S. cerevisiae* (Blat and Kleckner, 1999; Glynn *et al.*, 2004), yet cohesin at longer region centromeres such as those found in *S. pombe* seems to depend on heterochromatin formation, suggesting cohesin is primarily recruited to adjacent pericentromeres (Bernard *et al.*, 2001; Nonaka *et al.*, 2002). Therefore, the high density of cohesin at (peri)centromeres may block the amplification or spreading of γ H2A.X signalling in response to DNA damage. These two possibilities could explain why γ H2A.X specifically is suppressed at centromeres. However, the observation that another frequently used DNA damage marker (i.e., phosphorylated RPA) is also unresponsive to DSBs at mitotic centromeres suggests that the suppression may in fact be a more universal suppression of the DDR, beyond just γ H2A.X. Attempts were made to assess other markers of DNA damage including PARylation and 53BP1, but the fluorescent signals of these markers could not be validated. Therefore, the third possibility is that the upstream kinases of γ H2A.X signalling (ATM, ATR and DNA-PKcs) or even the damage recognition machinery (e.g., MRN for DSBs, RPA for SSBs) are somehow suppressed in their canonical functions at centromeres.

Despite the underlying mechanisms of centromeric DDR suppression, such a phenomenon would undoubtedly have significant impacts for genome stability. First, this observation could explain how DNA nicks at centromeres are able to accumulate over time without significantly perturbing cell cycle progression and why the inherent fragility at centromeres has not been previously detected. Second, an inefficient DNA response and repair system may explain why arm-CNAs occur so frequently, potentially increasing the incidence of failed break repair as well as translocations or non-allelic recombination events. To my knowledge, this work is one of the first to highlight potential pitfalls in relying on signalling cascades as markers for DNA damage, highlighting that DNA damage signalling thresholds or mechanisms may not be as spatially homogenous as previously anticipated.

Altogether, these findings are especially interesting considering the recent attention to genome instability originating at centromeres (Barra and Fachinetti, 2018; Black and Giunta, 2018). One intriguing possibility is that the high incidence of spontaneous DNA SSBs in combination with a suppressed DNA damage response may in fact underlie this propensity towards breakage, genome rearrangements and the onset of cancer. We are now beginning to understand the mechanism underlining centromere integrity control, which undoubtedly deserves further investigation.

Chapter 4

Investigating the role of homologous recombination (HR) in the repair of centromere breaks

4.1 Introduction

Human centromeres are prone to DNA breakage and rearrangements, contributing to human disease (Barra and Fachinetti, 2018; Black and Giunta, 2018). In **Chapter 3**, evidence is provided that human centromeres spontaneously accumulate DNA strand breaks, primarily in the form of single-stranded DNA breaks, even in non-cancerous cells and without exogenous perturbation. This surprising observation raises important questions about how centromeric breaks are repaired to ensure genome stability. As comprehensively discussed in **section 1.3.6**, there are several lines of evidence suggesting that the repair pathway homologous recombination (HR) acts at centromeres. Therefore, in this chapter, I set out to investigate the role of HR in the protection of human centromeric DNA from spontaneous DNA damage.

4.1.1 Homologous recombination: mechanisms and regulation

HR is catalysed by the highly conserved RecA/RAD51 family of DNA recombinases, using homologous sequences as repair templates (Baumann *et al.*, 1996; Li and Heyer, 2008; Sung, 1994). The substrates initiating HR as well as the proposed mechanism of crossover formation vary according to the HR model (Holliday, 1964; Meselson and Radding, 1975; Radding, 1982; Szostak *et al.*, 1983) (**Figure 1.2**). Although the current consensus model is the DSB-initiated Szostak model (**Figure 1.2D**), HR can be initiated by DNA DSBs and SSBs, the latter including both ssDNA nicks and ssDNA gaps (**section 1.1.1.4**).

As repair templates, homologous sequences can be alleles on sister chromatid following DNA replication,

known as allelic recombination. Alternatively, using DNA sequences with high sequence homology but on distinct parts of the chromosome is known as non-allelic homologous recombination (NAHR) (Stankiewicz and Lupski, 2002). NAHR can occur both within the same chromosome (intra-chromosomal NAHR) or between chromosomes (inter-chromosomal NAHR) and is particularly prevalent within repetitive regions given the abundance of non-allelic templates available for repair (**Figure 4.1**).

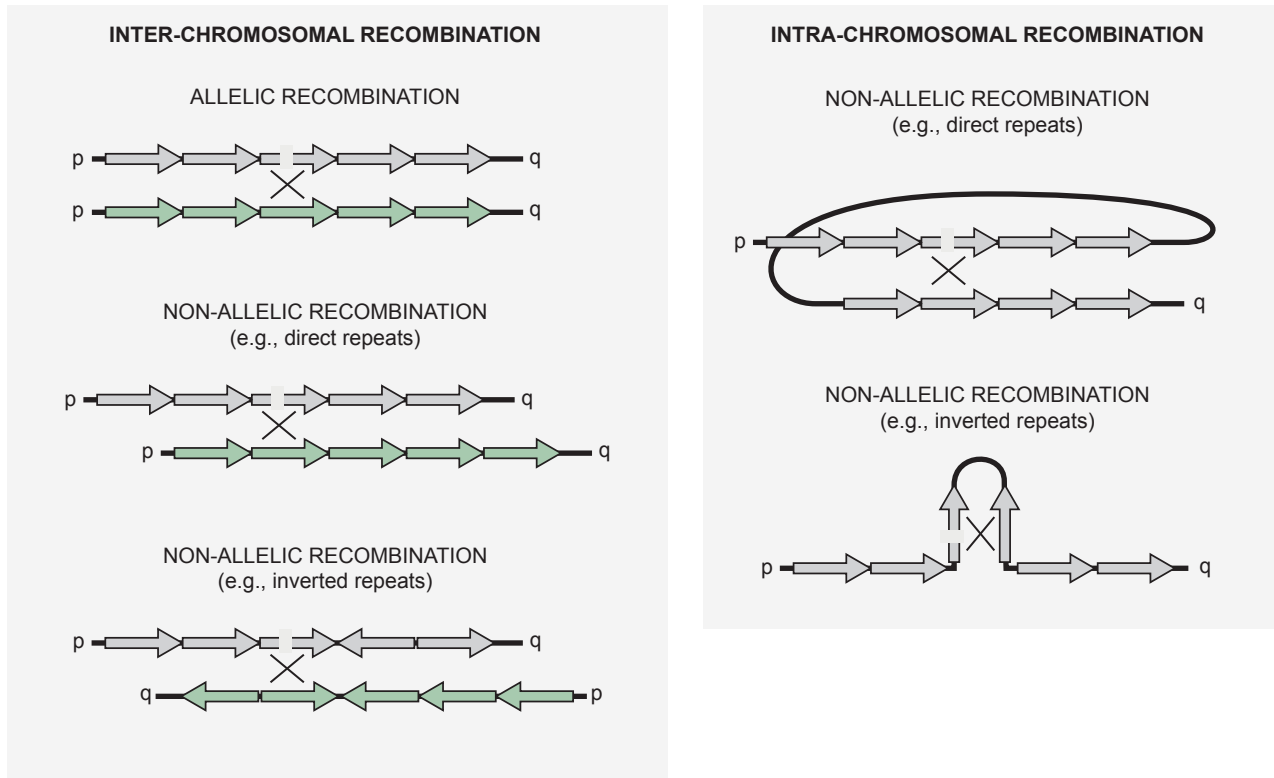


Figure 4.1: **Models of homologous recombination at repetitive regions.** Depictions of allelic and non-allelic homologous recombination between chromosomes (inter-chromosomal) and within chromosomes (intra-chromosomal). p and q arms of chromosomes denote chromosome orientation.

As discussed in the **Chapter 1**, RAD51 is canonically recruited by the HR mediator complex composed of breast cancer 1 (BRCA1) and breast cancer 2 (BRCA2) as well as the partner and localiser of BRCA2 (PALB2) proteins (Jensen *et al.*, 2010; Sy *et al.*, 2009; Yuan *et al.*, 1999). However, numerous alternative RAD51 loaders including the MMSL-TONSL complex, NBS1, and RAD52 have been identified, highlighting the complexity of HR regulation in genome maintenance (Chastain *et al.*, 2016; O'Donnell *et al.*, 2010; Park *et al.*, 2019; Teng *et al.*, 2018; Yata *et al.*, 2014; Yata *et al.*, 2012).

4.1.2 Investigating homologous recombination at human centromeres

The evidence for HR actively repairing centromeres come from two principal observations, as fully discussed in **section 1.3.6.1**. First, reciprocal exchanges (i.e., SCEs) are observed between mammalian centromeres at much higher frequencies than other regions of the genome, and in the absence of any kind

of exogenous perturbation (Giunta and Funabiki, 2017; Jaco *et al.*, 2008). Second, alpha satellite arrays appear to homogenize from within functionally active centromere cores, providing evidence for frequent unidirectional genetic exchanges at centromeres (Alexandrov *et al.*, 2001; Lee *et al.*, 2005; Miga, 2019; Talbert *et al.*, 2018; Warburton and Willard, 1995; Willard, 1985). This is backed by experimental data finding high rates of gene conversion between centromeric repeats in maize (Shi *et al.*, 2010). However, no direct assessment of constitutive HR functionality at human centromeres has been performed. In this chapter, I set out to determine whether HR constitutively acts at human centromeres by assessing the role of the recombinase RAD51 at centromeres.

Up until now, centromeric recombination has been primarily evaluated using CO-FISH, which detects centromere SCEs (Giunta and Funabiki, 2017; Jaco *et al.*, 2008). While SCEs can be used as a proxy for centromeric HR (Bailey *et al.*, 2004), this approach is limited for three principal reasons. First, SCEs are relatively rare in non-cancerous cells such as hTERT-RPE1 cells, limiting the dynamic range to detect changes upon candidate protein depletion. Second, SCEs only reflect a fraction of actual recombination events, as HR can equally act between intra-chromosomal repeats and/or in a non-crossover mode. Indeed, crossovers may even be actively suppressed at centromeres independently of the rate of recombination (**section 1.3.6.2**), complicating the relationship between SCE and recombination rates (Onaka *et al.*, 2016; Zafar *et al.*, 2017). Third, CO-FISH relies on actively dividing cells that reach mitosis, systematically preventing CO-FISH analysis when depleting candidate factors that severely impact cell cycle progression.

Assessing a direct role for the canonical HR machinery at centromeres has therefore been technically challenging until now. With the development of exo-FISH (as described and validated in **Chapter 3**), I was in the position to experimentally investigate whether the HR recombinase RAD51 facilitates homologous recombination repair at centromeric DNA. An inability to facilitate DNA repair by HR would result in an accumulation of DNA breaks, which can be detected using exo-FISH independently of any downstream crossover suppression mechanisms. I also set out to assess the contribution of BRCA2, as a key regulator of RAD51, to centromeric recombination.

4.1.3 Human artificial chromosomes (HACs) as centromere models

qPCR analyses of repetitive elements can be challenging due to their repetitive nature and propensity towards DNA secondary structures. Therefore, performing ChIP-qPCR for the assessment of protein localization to centromeres poses several technical limitations. To circumvent this issue, this study makes use of small extrachromosomal elements called human artificial chromosomes (HACs). HACs have recently been widely adopted for the study of centromere assembly and function (Molina *et al.*, 2017). HACs replicate autonomously and form functional centromeres, thereby segregating independently

(Kouprina *et al.*, 2012). They are therefore valuable centromere models that can be easily manipulated and studied with little impact to overall cell survival.

Several iterations of HACs have been developed. In this study, I use a HeLa cell line containing a HAC composed of alphoid^{tetO} repeats and constitutively expressing tetR-eYFP, as was previously generated and validated (Ohzeki *et al.*, 2012) (**Figure 4.2**). These alphoid repeats were synthetically generated by rolling-circle amplification and transformation-associated recombination. In this particular iteration, the synthetic alphoid^{tetO} HACs were generated based on a native dimer from chromosome 21 alpha satellite DNA, with one of the monomers containing a tetO sequence instead of a CENP-B box (Ohzeki *et al.*, 2012). The tetO repeat is bound by tetR-eYFP, which can be used for selective targeting of fusion proteins, but also functions as a read-out for protein-HAC association without any complications from DNA secondary structures.

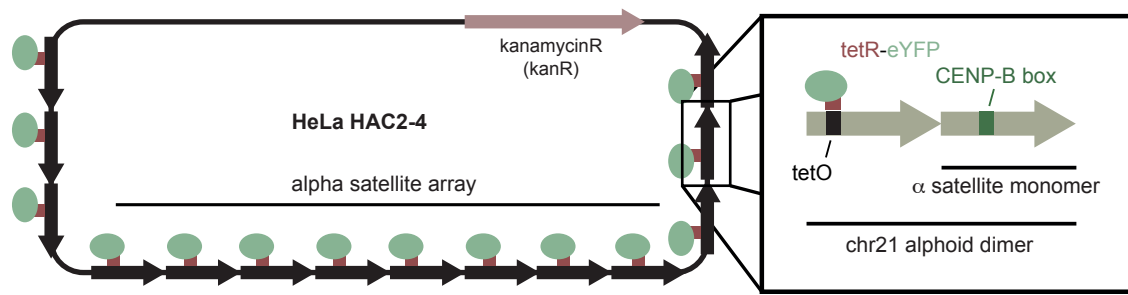


Figure 4.2: **Human Artificial Chromosomes (HACs) as centromere models.** Schematic of the HeLa HAC2-4 cell line containing a human artificial chromosome (HAC) composed of native alpha satellite dimers from chromosome 21. Within each dimer, one alpha satellite monomer contains the canonical 17-bp CENP-B box, whereas the other contains a tetO sequence. Constitutive expression of tetR-eYFP results in binding of tetR-eYFP to tetO repeats. A kanamycin resistance cassette enables HAC maintenance selection.

4.2 Results

4.2.1 HR factor localization to centromeres

To establish a potential role for RAD51 in centromere DNA repair, I first sought to investigate whether RAD51 localizes to centromeres. To this end, I used the HeLa cell line containing a HAC composed of alphoid^{tetO} repeats and constitutively expressing tetR-eYFP (**Figure 4.2**). ChIP-qPCR was performed to detect RAD51 association with tetO sequences as a proxy for centromere binding. First, I optimized the chromatin shearing sonication conditions to generate ~300-500bp fragments, as recommended for the detection of protein-DNA interactions (**Figure 4.3A**). I also ensured that RAD51 was being immunoprecipitated appropriately by Western blotting (**Figure 4.3B**). Using agarose gel analysis, I assessed the final ChIP qPCR products to ensure single products were obtained using the chosen probes against

both HAC centromeres (tetO) and endogenous centromeres. The HAC resistance cassette kanamycinR (kanR) was also assessed to determine any non-centromeric HAC binding (**Figure 4.3C**). Both the kanR and endogenous centromere probes produced products that were slightly larger than is optimal for qPCR analysis, and so longer extensions times were used to account for their size. In addition, the probe for endogenous centromeres produced multiple bands, an important caveat that highlights the value of using tetO as a proxy for centromere binding using qPCR.

Using these conditions, I assessed the localization of RAD51 to tetO repeats at the HAC as a proxy for centromere binding, using tetR-eYFP association with tetO as a positive control (**Figure 4.3D**). qPCR signals were normalized to both IgG and GAPDH to estimate fold enrichment. From these results, an enrichment of tetO sequences was consistently detected when pulling down RAD51. To ensure the specificity of the RAD51-HAC signal, I treated HAC2-4 tetR-eYFP asynchronous cells for 1 hour with the RAD51 small molecule inhibitor B02 (Huang *et al.*, 2011) (**Figure 4.3E**). Once again normalizing to both IgG and GAPDH, there was a reduction of RAD51 association with both tetO and kanR upon B02 treatment, suggesting that the detected RAD51-HAC association is indeed specific.

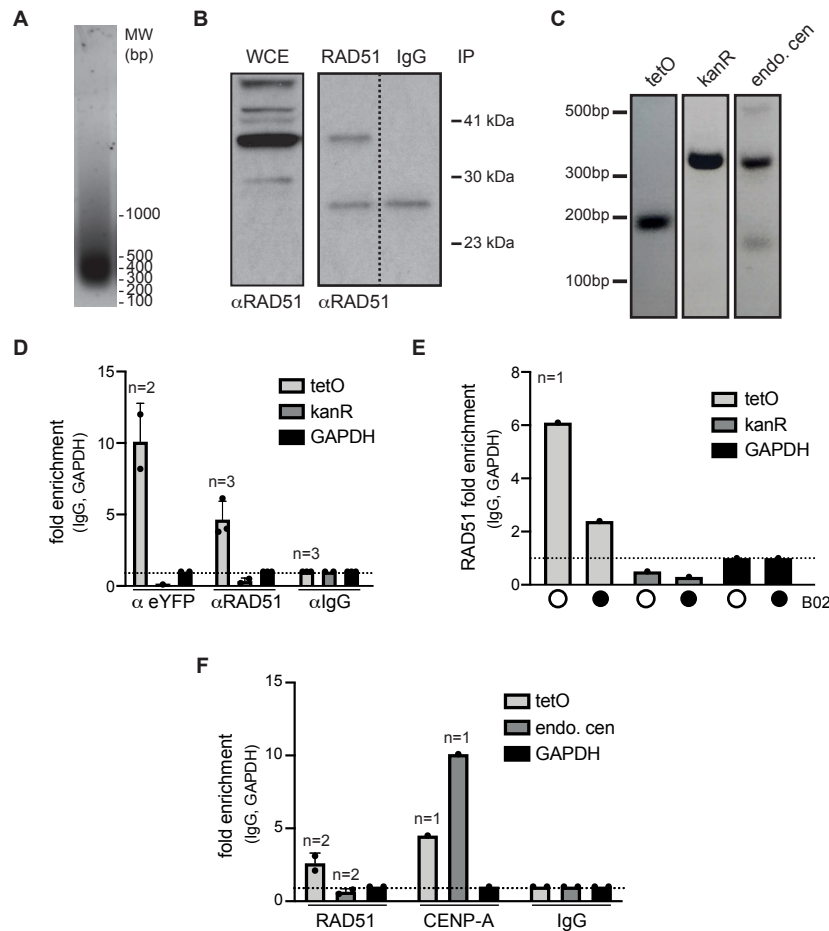


Figure 4.3: Investigating HR factor localization to HAC centromeres. HeLa cell line containing human artificial chromosome (HAC) with tetO-containing alpha satellite repeats as a proxy for human centromeres and expressing tetR-eYFP. (A) HeLa HAC2-4 tetR-eYFP cell line harvested and fixed. Chromatin shearing was optimized to produce ~300-500bp fragments, as recommended for the detection of protein-DNA interactions. Agarose gel depicted here visualizing the fragment size distribution. (B) Verification of immunoprecipitation of RAD51. Western blot showing whole-cell extract (WCE) illustrates RAD51 antibody staining patterns. Following ChIP pull-down with RAD51, but not the negative control IgG, the RAD51 band were detected at the predicted molecular weight (36 kDa). For visualisation, the Western blot was cropped as indicated by the dashed line. (C) Agarose gel analysis of qPCR products for the detection of the tetO sequence, the HAC resistance marker kanR and a pan-alpha satellite probe for endogenous centromeres (endo. cen). (D) ChIP-qPCR analysis in HeLa HAC2-4 tetR-eYFP cell line, pulling down either eYFP, RAD51 or the negative control IgG. qPCR was used to detect enrichment of the tetO sequence as a proxy for centromeres, kanR as a control for general non-centromeric HAC binding, and GAPDH as a negative genomic control. Fold enrichment was calculated as $2^{-\Delta\Delta C_t}$, using IgG and GAPDH for normalization. (E) ChIP-qPCR repeated following a 1-hour in vivo treatment with the RAD51 small molecular inhibitor B02 (Huang *et al.*, 2011), which prevents RAD51-DNA association. Fold enrichment was calculated as $2^{-\Delta\Delta C_t}$, using IgG and GAPDH for normalization. (F) ChIP-qPCR repeated but assessing the enrichment of RAD51 at HAC centromeres (tetO) as well as endogenous centromeres (endo. cen). CENP-A was used as a positive control for endogenous centromere binding. Fold enrichment was calculated as $2^{-\Delta\Delta C_t}$, using IgG and GAPDH for normalization.

To determine whether RAD51 binding could be detected at endogenous centromeres as well as HAC centromeres, I repeated the ChIP-qPCR in the HAC2-4 tetR-eYFP cell line assessing both tetO enrich-

ment as well as a probe for alpha satellite sequences found at endogenous centromeres (**Figure 4.3F**). Although this analysis is more complicated due to unusual DNA structure of alpha satellites, CENP-A was used as a positive control to ensure endogenous centromere binding could be detected. Curiously, RAD51 appeared to only bind the tetO-containing HAC sequences but not endogenous centromeres, suggesting that either RAD51 binding is exclusive to HAC centromeres, RAD51 binding is specific to chromosome 21 alphoid DNA found in the HAC, or that RAD51 binding at endogenous centromeres cannot be adequately detected using the chosen alpha satellite probe. Therefore, while it appears that there is some RAD51 association with HAC centromeres, it remains uncertain whether RAD51 interacts with endogenous centromeres at detectable levels.

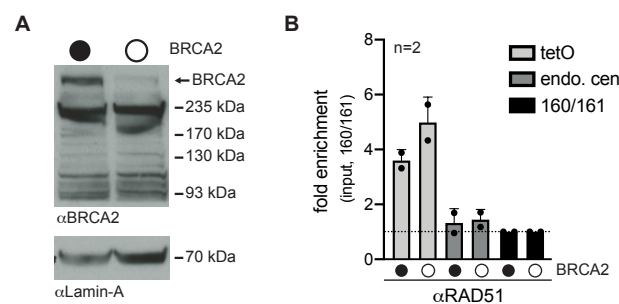


Figure 4.4: BRCA2 in the recruitment of RAD51 to centromeres. (A) Verification of efficient siRNA-mediated knockdown of BRCA2 48 hours after siRNA treatment. Western blot of whole-cell extract (WCE) staining against BRCA2, with the BRCA2 protein band indicated with an arrow. Lamin-A was used as a loading control. (B) ChIP-qPCR repeated as in **Figure 4.3**, assessing RAD51 localization to both HAC centromeres (tetO) as well as endogenous centromeres (endo. cen). Fold enrichment was calculated as $2^{-\Delta \Delta C_t}$, using the percentage input and a negative control genome sequence 160/161 for normalization in place of GAPDH.

I next sought to determine whether RAD51-HAC association is dependent on the canonical RAD51 regulator BRCA2. BRCA2 was depleted in HAC2-4 tetR-eFYP cells, as verified by Western blotting (**Figure 4.4A**). ChIP-qPCR was performed, but in this series using the percentage input and a new negative control (160/161) for normalization, as it was subsequently observed that RAD51 displays a small degree of binding to GAPDH. Under these conditions, RAD51-HAC interaction did not decrease, and perhaps even slightly increased (**Figure 4.4B**). Therefore, it appears that the RAD51-HAC association is not positively promoted by its canonical regulator BRCA2.

4.2.2 RAD51-mediated protection of human centromeres

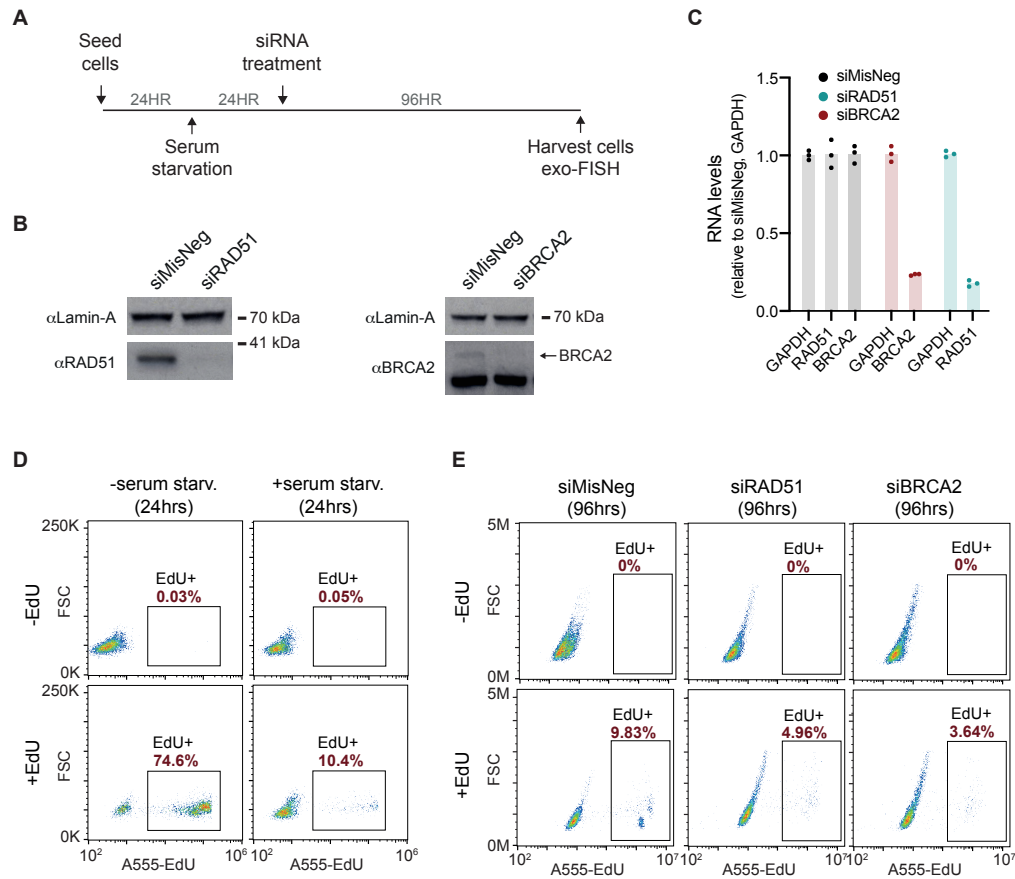


Figure 4.5: Optimization for the investigation of HR at centromeres. (A) Experimental schematic for the detection of centromeric breaks using exo-FISH upon RAD51 or BRCA2 siRNA-mediated depletion in hTERT-RPE1 serum-starved cells. (B) Western blot confirming the depletion of RAD51 and BRCA2 protein levels following 72 hours of siRNA-mediated depletion in hTERT-RPE1 serum-starved cells. Lamin-A levels were used as a loading control. (C) RT-qPCR of RNA extracted from serum-starved hTERT-RPE1 cells following 96 hours depletion of RAD51, BRCA2 or a negative control (siMisNeg). RNA subsequently probed for GAPDH, RAD51 and BRCA2 mRNA to determine expression levels. (D) Validation of serum starvation preventing cell proliferation. hTERT-RPE1 were incubated with serum-starved media (0.5% FBS) for 24 hours and then incubated with EdU for another 24 hours, as proliferating cells will incorporate EdU into DNA during replication. Cells were then fixed and analyzed by flow cytometry. Gates were chosen using the -EdU negative control to define the EdU+ population. Forward scatter (FSC) is an indicator of cell size. (E) Cells were incubated with serum-starved media (0.5% FBS) for 24 hours and then treated with siRNA targeting either RAD51, BRCA2 or a universal negative control. Cells were then incubated for 96 hours in the presence of EdU in serum-starved media, fixed and analyzed by flow cytometry to determine what proportion of the population had incorporated EdU during proliferation. Gates were chosen using the -EdU negative control to define the EdU+ population. Forward scatter (FSC) is an indicator of cell size.

To investigate HR repair at centromeres, I next sought to investigate the roles of RAD51 and BRCA2 in repairing centromeric DNA breaks using exo-FISH. Given that both RAD51 and BRCA2 play several roles in DNA metabolism including DNA replication and repair, I decided to test whether RAD51 or BRCA2 protect centromeres independently of their roles in DNA replication. To this end, hTERT-

RPE1 cells were serum starved for 24 hours to induce quiescence and subsequently treated with siRNA targeting RAD51 or BRCA2. Following 96 hours of further incubation, centromere breaks were assessed using exo-FISH (**Figure 4.5A**). RAD51 and BRCA2 depletion was verified with both western blotting (**Figure 4.5B**) and RT-qPCR (**Figure 4.5C**). The induction of quiescence was verified by quantifying the proportion of cells having incorporated the thymidine analog EdU over 24 hours with and without serum starvation, as EdU will only be incorporated in actively replicating cells (**Figure 4.5D**). Finally, I verified that the RNAi treatment over 96 hours did not result in the resumption of cellular proliferation, as relatively few cells had incorporated any EdU over the full 96 hours of siRNA treatment (**Figure 4.5E**).

After having verified that cellular proliferation was indeed halted and that BRCA2 and RAD51 could be adequately depleted in the hTERT-RPE1 cells, exo-FISH was used to determine the extent of centromere breaks (**Figure 4.6**). As described in **Chapter 3**, FISH foci were automatically detected based on an arbitrarily chosen threshold and the ‘median focus’ analysis, including background subtraction, was used to estimate a median FISH focus signal intensity per cell. From this analysis, I observed that siRNA-mediated depletion of RAD51 for 96 hours resulted in an increase in centromeric, but not telomeric, FISH signals in response to ExoIII treatment (**Figure 4.6A, B**). Studies assessing the role of BRCA2 in centromere protection revealed that BRCA2 depletion did not have any major impact on cenFISH signals upon ExoIII treatment, although these results were more variable between biological replicates (**Figure 4.6C, D, E**). These data support a role for RAD51, but not BRCA2, in suppressing the accumulation of spontaneous DNA breaks observed specifically at centromeres. The analyses of publicly available datasets revealed that RAD51 depletion in MCF7 cells did not induce DSBs at centromeres as detected by sBLISS (**Figure 4.6F**) (Hazan *et al.*, 2019), suggesting that the breaks detected by exo-FISH most likely correspond to an accumulation of SSBs rather than *de novo* DSBs.

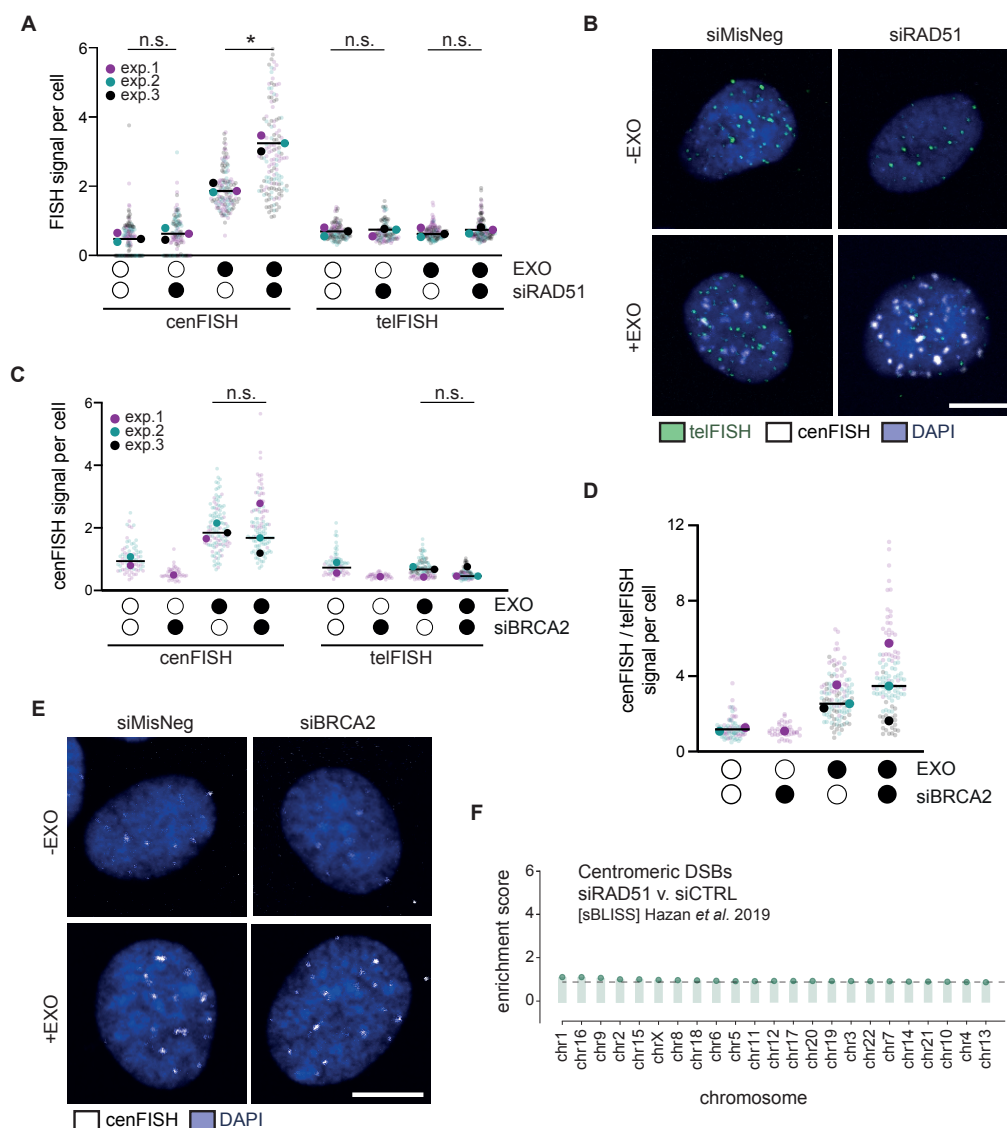


Figure 4.6: RAD51 prevents the accumulation of centromeric DNA breaks. (A, B) Quantification and representative images of exo-FISH probing for both telomeric (telFISH) and centromeric (cenFISH) sequences with and without Exonuclease III (EXO) treatment following RAD51 siRNA-mediated depletion. cenFISH, telFISH and DAPI signals are shown in white, green and blue, respectively. (C, D, E) Quantification and representative images of exo-FISH with and without Exonuclease III (EXO) treatment following BRCA2 siRNA-mediated depletion. cenFISH and DAPI signals are shown in white and blue, respectively. Quantification was performed as described in **Chapter 3**. Scale bar represents 10 μ m. FISH signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median value of a cell, with the average of each biological replicate indicated in darker hues. The averages of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$, ** $p < 0.01$). (F) Enrichment scores of DNA double-strand break (DSB) mapping (sBLISS) (Hazan *et al.*, 2019) following siRNA-mediated RAD51 depletion across human autosome centromeres. Enrichment scores were calculated as described in **Figure 3.3**, defining the centromeric DSB read density in siRAD51 relative to siCTRL. Chromosomes are ranked by descending enrichment score.

To determine whether any DDR activation (i.e., γ H2A.X) could be detected upon RAD51 depletion, I co-stained γ H2A.X and CENP-B under the same experimental conditions employed in **Figure 4.5** and **Figure 4.6** and quantified the centromeric γ H2A.X signal by median focus analysis. Supporting previous

observations of DDR suppression at centromeres (**Chapter 3**), RAD51 depletion did not result in any induction of γ H2A.X at centromeres in quiescent cells (**Figure 4.7A, B, C**).

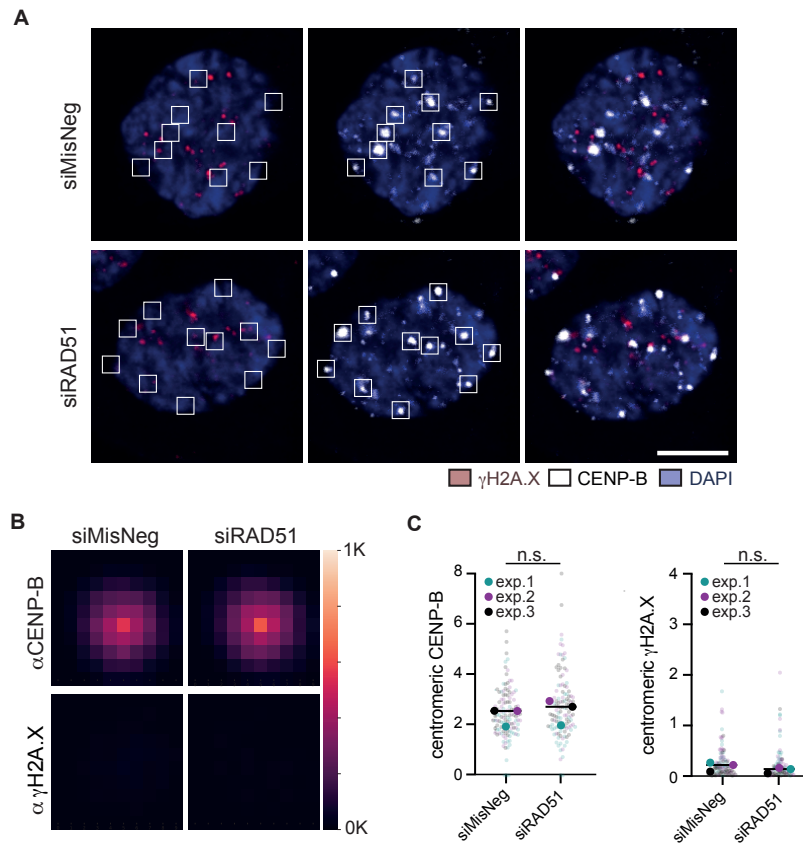


Figure 4.7: Centromeric DDR activation in the absence of RAD51. (A-C) Representative images, representative median foci and quantification of immunofluorescence staining against γ H2A.X and CENP-B. hTERT-RPE1 cells were fixed 96 hours after siRNA-mediated depletion of RAD51. γ H2A.X and CENP-B levels were subsequently quantified at CENP-B-defined centromeres. Quantification was performed as described in **Figure 3.9**. Scale bar represents 10 μ m. IF signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median value of a cell, with the median of each biological replicate indicated in darker hues. At least 30 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$, ** $p < 0.01$).

4.2.3 RAD51 mechanism of protection

After observing an increase in centromeric DNA breaks in the absence of RAD51, I next sought to determine whether this was an accumulation of DNA breaks as a result of defective HR repair or rather due to an incidence of *de novo* DNA breaks in the absence of RAD51. Aside from directly repairing DNA breaks via HR, RAD51 also has protective roles through its ability to bind exposed ssDNA. For example, RAD51 binds ssDNA (e.g., at stalled replication forks) to prevent excessive nuclease degradation independently of its enzymatic activity (Mason *et al.*, 2019; Schlacher *et al.*, 2011; Schlacher *et al.*, 2012). Therefore, the observed increase in centromeric DNA breaks in the absence of RAD51 could conceivably be a result of either defective RAD51 strand-exchange (i.e., recombinational repair) or protective ssDNA

binding.

To investigate the mode of RAD51 function in suppressing centromeric DNA breaks, I exploited a series of RAD51 separation-of-function variants (**Figure 4.8A**). During HR, human RAD51 binds ssDNA to form presynaptic filaments prior to initiating homology search and DNA strand exchange in an ATP-dependent manner (Benson *et al.*, 1994) (**Figure 1.3**). RAD51 also exhibits DNA-dependent ATPase activity, and ATP hydrolysis by RAD51 is required for the disassembly of filaments from dsDNA (Chi *et al.*, 2006; Morrison *et al.*, 1999; van Mameren *et al.*, 2009). Therefore, the first RAD51 variant is RAD51-K133R, which binds ATP and is defective in ATP hydrolysis, hence retaining its DNA binding and strand exchange activities (Chi *et al.*, 2006; Morrison *et al.*, 1999). Second, RAD51-II3A contains three amino acid substitutions (R130A, R303A, K313A) and can bind DNA but is unable to catalyze strand exchange (Cloud *et al.*, 2012; Mason *et al.*, 2019). Finally, RAD51-T131P is unable to bind DNA on its own in any capacity (Wang *et al.*, 2015). Using these tools, I set out to determine whether RAD51 DNA binding and/or strand exchange activity is important in its ability to protect centromeres from DNA break accumulation.

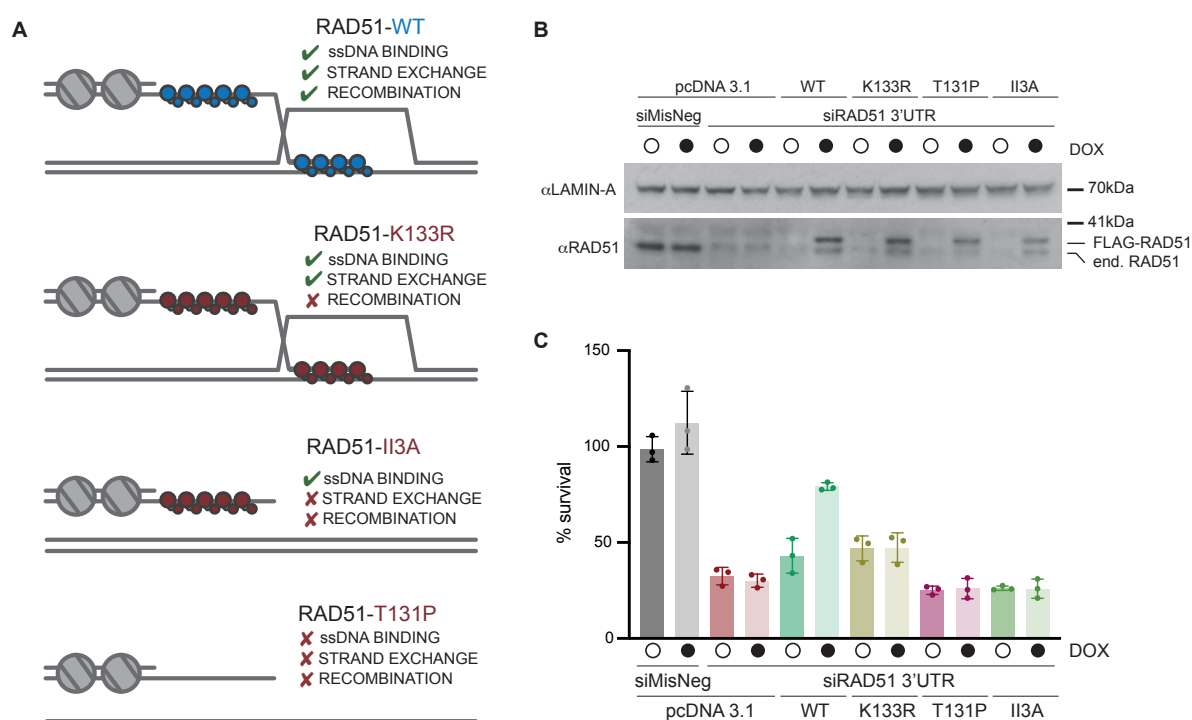


Figure 4.8: Generation and validation of RAD51 separation-of-function (SOF) variants. (A) Schematic of RAD51 separation-of-function variant functionalities, as generated using the hTERT-RPE1 Flp-In T-Rex cell line. (B) Western blot confirming the depletion of endogenous RAD51 (end. RAD51) protein levels upon siRNA treatment for 48 hours, as well as the conditional expression of separation-of-function variants upon doxycycline (DOX) treatment. Lamin-A levels were used as a loading control. (C) WST-1 assay to measure the cell survival after 120 hours endogenous RAD51 siRNA-mediated depletion and following expression of the RAD51 separation-of-function mutants.

The RAD51 separation-of-function variants were generated using the Flp-In T-REx integration system, which produces isogenic stable cell line using an hTERT-RPE1 Flp-In T-REx parental cell line. Unlike random integration, this form of integration allows for precise insertion of the desired sequences at a targeted locus and limits variation in construct copy number and expression. Flp-In constructs expressing Flag-tagged wildtype RAD51 (RAD51-WT), RAD51-K133R and RAD51-T131P were previously generated in the lab (Wassing *et al.*, 2020). A construct expressing RAD51-II3A was generated through site-directed mutagenesis of RAD51-WT at R130A, R303A and K131A. The existing resistance cassette of all constructs (hygromycin resistance) was switched with a neomycin resistance cassette, as the hTERT-RPE1 Flp-In T-REx cell line was already resistant to hygromycin. Constructs were analyzed by sequencing to verify both individual base mutations as well as resistance marker switches.

Genes encoding FLAG-fusions of these RAD51 variants were then stably integrated in the hTERT-RPE1 Flp-In T-REx cell line, such that their expression could be conditionally induced by doxycycline. I depleted endogenous RAD51 by siRNA targeting the 3'UTR and asked whether expression of the RAD51 variants would be able to rescue the centromeric DNA damage phenotype. First, I verified that endogenous RAD51 could be depleted and that the RAD51 variants were conditionally expressed (**Figure 4.8B**). An additional band was detected at the endogenous RAD51 predicted size upon variant expression, which might reflect some stabilization of endogenous RAD51 protein due to physically interaction between variant RAD51 and endogenous RAD51 monomers, or a cleavage product of the RAD51 variants. Therefore, I sought to verify that no endogenous RAD51 functionality was maintained, which could conceivably interfere with downstream analyses. Assessing cell survival using the cell proliferation and viability marker WST-1 after 120 hours of endogenous RAD51 depletion, it was evident that there was no rescue of cell viability even in conditions with an apparent protein band corresponding to endogenous RAD51 (**Figure 4.8C**).

I then used exo-FISH to determine the level of centromeric breaks in these cells in the absence of endogenous RAD51 (**Figure 4.9A, B, C**). Although depletion of endogenous RAD51 was less apparent in the hTERT-RPE1 Flp-In T-REx cells than what was previously observed in the hTERT-RPE1 cells, there was still a notable accumulation of centromeric DNA breaks in the absence of RAD51. Expression of either RAD51-WT or RAD51-K133R was able to rescue this phenotype to a similar degree. However, expression of RAD51-II3A or RAD51-T131P did not reduce the proportion of cells with high levels of centromeric breaks. These data suggest that RAD51 strand-exchange activity, which is selectively present in RAD51-WT and RAD51-K133R, is critical for preventing the accumulation of DNA breaks at human centromeres. Conversely, the ability of RAD51 to bind single-stranded DNA in a protective capacity, which is present in the RAD51-II3A variant, appears inconsequential to the overall levels of DNA breaks at centromeres. These data support a model whereby both the homology-pairing and strand-exchange

activities of RAD51 (i.e., recombinational repair activity) are critical for limiting the accumulation of centromeric breaks.

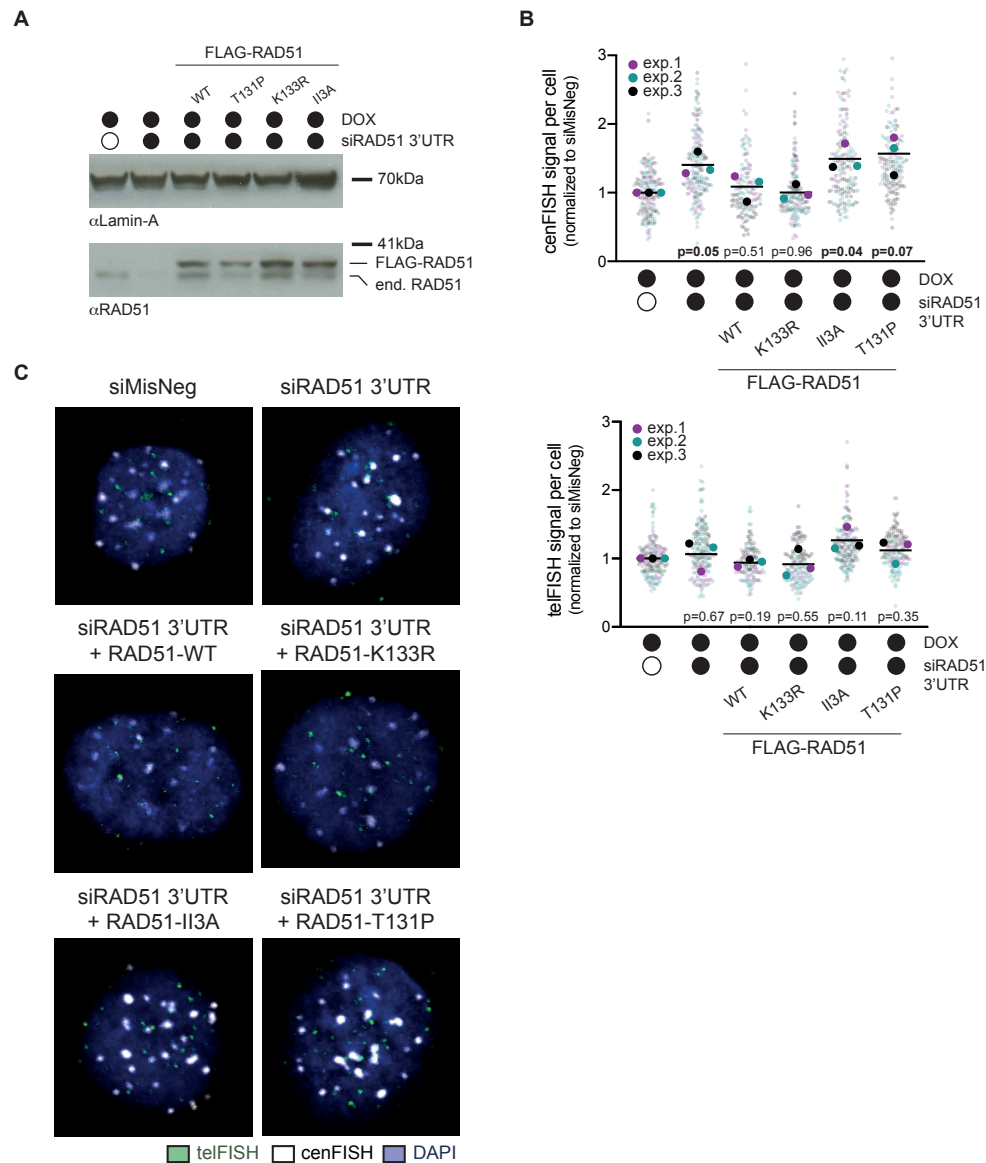


Figure 4.9: Investigating RAD51 mechanism of centromeric protection. (A) Western blot confirming the depletion of endogenous RAD51 (end. RAD51) protein levels upon siRNA treatment for 48 hours, as well as the conditional expression of separation-of-function variants upon doxycycline (DOX) treatment. Lamin-A levels were used as a loading control. (B, C) Quantification and representative images of exo-FISH upon endogenous RAD51 depletion and RAD51 variant re-expression. Quantification of cenFISH and telFISH signals was done as described in **Chapter 3**. The p-values corresponding to each treatment condition, as calculated relative to the siMisNeg control, were attained using a two-sided paired t-test. At least 30 cells were imaged per experimental condition. cenFISH, telFISH and DAPI signals are shown in white, green, and blue, respectively. Scale bar represents 10 μ m.

4.3 Discussion

In this chapter, I set out to investigate whether centromeric DNA breaks are repaired by homologous recombination. As described above, there are several lines of evidence suggesting that HR actively repairs centromeric DNA. First, high rates of sister-chromatid exchanges (SCEs) have been observed at both murine and human centromeres (Giunta and Funabiki, 2017; Jaco *et al.*, 2008). Second, evolutionary analyses of centromeric alpha satellites suggests that arrays undergo frequent array rearrangements and gene conversions (Shi *et al.*, 2010). However, both observations are based on indirect read-outs of HR, and rates of centromeric SCEs are very low in non-cancerous cells (i.e., hTERT-RPE1). It therefore remained unclear to what extent recombination is active at human centromeres in baseline conditions (i.e., non-cancerous cells, unperturbed conditions), when spontaneous DNA breaks are observed. Given that HR is typically catalyzed by the evolutionarily conserved RecA/RAD51 family of DNA recombinases (Baumann *et al.*, 1996; Sung, 1994), I addressed this question by investigating the role of the recombinase RAD51 in the repair of centromeric DNA breaks.

4.3.1 HR factor recruitment to human centromeres

To determine whether HR recombinase RAD51 functions at centromeres, I first sought to investigate their localization to centromeres by ChIP-qPCR (**Figure 4.3**). Using the synthetic HAC alphoid^{tetO} system, RAD51 appeared to consistently interact with tetO sequences found within alpha satellite dimers.

It was peculiar to note that although RAD51 reproducibly interacted with HAC centromeres, no enrichment could be detected at endogenous centromeres when probing for the consensus alpha satellite sequence. This observation can be explained by several possibilities. First, it is plausible that the probe used to assess endogenous centromere enrichment is not sensitive enough to detect a transient RAD51-centromere enrichment. Indeed, CENP-A itself is enriched only ~10-fold at endogenous centromeres, which is perhaps lower than might be expected given that CENP-A is almost exclusively found at centromeres. Second, there may be mechanisms in place to actively prevent the stable association of RAD51 to endogenous centromeres, perhaps to prevent excessive recombination between non-allelic centromeric repeats. If these mechanisms do not function at HAC centromeres, this could explain the disparity in enrichment patterns. Third, the HAC system specifically uses alphoid arrays from chromosome 21, and the association of RAD51 may reflect some chromosome-specific trends. For example, chromosome 21 is the only chromosome to exhibit CENP-A localization and DNA breaks outside of the annotated HOR (**Figure 3.5A**), a phenomenon which has not been experimentally investigated to date. Finally, there may be some intrinsic feature of the HAC that renders HAC centromeres more susceptible to DNA damage and/or replication stress which could result in RAD51 association. One possibility is that the frequent

tetR-eYFP binding of the cell line results in excessive replication stress, although to my knowledge this possibility has not been experimentally addressed.

Overall, while there is some indication that RAD51 is enriched at centromeric DNA, the inability to detect RAD51 at endogenous centromeres remains an important caveat, and the conflicting evidence warrants further investigation into the localization of RAD51 at centromeres. However, a recent report shows RAD51 recruitment to centromeres upon Cas9-induced DSBs, suggesting that RAD51 can in fact be recruited to centromeres, but it is possible that SSB-induced RAD51 recruitment is less readily detectable (Yilmaz *et al.*, 2021).

An important aspect of this study was also to identify potential regulatory factors in centromeric recombination. Specifically, I set out to determine whether BRCA2 influences RAD51 recruitment to and function at centromeres. BRCA2, as part of the HR mediator complex, generally promotes the recruitment of RAD51 to sites of DNA damage in S and G2 phases of the cell cycle (Jensen *et al.*, 2010; Yuan *et al.*, 1999). However, BRCA2 can also have a negative regulatory role in certain contexts, such as when CDK2-mediated phosphorylation of BRCA2 prevents the C-terminal RAD51 interaction (Esashi *et al.*, 2005), which can de-stabilize RAD51 filaments. Therefore, it was conceivable that BRCA2 has both positive and negative regulatory effects on RAD51 centromere interaction. The depletion of BRCA2 had little to no effect on RAD51 association with the HAC, perhaps even slightly increasing RAD51-HAC interaction, and no effect on RAD51 enrichment at endogenous centromeres (**Figure 4.4**). This observation implies that RAD51 recruitment to centromeres is largely independent of its canonical regulator BRCA2, or perhaps even negatively regulated. Several BRCA2-independent RAD51 loaders have been identified including MMS22L-TONSL, NBS1, RAD52, etc. (Moudry *et al.*, 2016; O'Donnell *et al.*, 2010; Teng *et al.*, 2018; Yata *et al.*, 2014; Yata *et al.*, 2012), providing potential avenues for future research.

4.3.2 HR as a repair mechanism for centromere DNA breaks

To address the role of HR in the repair of centromeric breaks in physiological settings, I sought to determine whether centromeric DNA breaks would accumulate in the absence of the mammalian recombinase RAD51. To avoid any major cell cycle or cell death impacts of RAD51 depletion, I applied exo-FISH in non-dividing hTERT-RPE1 cells. These analyses revealed that the depletion of RAD51 reproducibly increased the cenFISH, but not telFISH, signal, implying a specific accumulation of centromeric DNA breaks in the absence of RAD51 (**Figure 4.6**). In support of the previous observation that BRCA2 does not impact RAD51 recruitment to centromeres, no change in the level of centromeric breaks were observed upon BRCA2 depletion (**Figure 4.6**). These results collectively imply that RAD51 function in centromeric DNA break repair is largely independent of its canonical regulator BRCA2.

An accumulation of centromeric DNA breaks in the absence of RAD51 implies that RAD51 functions at centromeres, conceivably either by repairing existing DNA breaks or by preventing the incidence of new DNA breaks. While the analysis of publicly available datasets suggests that RAD51 depletion does not result in *de novo* DSBs at centromeres (**Figure 4.6F**), it is difficult to entirely exclude the latter possibility using exo-FISH alone. Therefore, to distinguish between the possibilities, I sought to determine whether RAD51 recombinase/catalytic activity is required for preventing the accumulation of centromeric DNA breaks, as this would imply that RAD51 was actively repairing existing centromeric breaks.

Exploiting a series of RAD51 variants, I determined that exogenously expressing RAD51 variants retaining both ssDNA binding and DNA strand-exchange activities (RAD51-WT, RAD51-K133R) were able to rescue the level of centromeric breaks upon endogenous RAD51 depletion (**Figure 4.9**). Conversely, expression of RAD51 variants lacking strand-exchange activity (RAD51-T131P, RAD51-II3A) was unable to rescue the level of centromeric breaks. This observation suggests that RAD51 strand-exchange activity is imperative for preventing the accumulation of centromeric breaks, further implying that RAD51 is acting in a repair capacity at centromeres.

Collectively, these data support the notion that HR does in fact play a role in the repair of centromeric DNA breaks, even in non-cancerous cell lines and without exogenous perturbation. This finding highlights a previously unappreciated role for HR in cell physiology, perhaps even relating to the poorly understood embryonic lethality observed upon RAD51 depletion in mammalian cells (Tsuzuki *et al.*, 1996). This finding also provides experimental evidence supporting conclusions derived from both evolutionary studies and CO-FISH observations, improving our understanding of DNA metabolism at human centromeres.

4.3.3 Additional repair pathways in the repair of centromeric DNA breaks

One critical limitation of this study is that only one repair pathway (i.e., HR) was addressed as a potential repair mechanism for centromeric DNA breaks. Therefore, while the data suggest that HR is at least partially required for the repair of centromeric breaks, there is no indication as to the relative importance of HR as compared to other repair pathways. NHEJ, often cited as a parallel pathway choice to HR, is an unlikely competing pathway in this context since NHEJ uses DSBs as an initiating substrate, and centromeric DNA breaks appear to be predominantly single-stranded. However, several other SSB repair pathways have been well established.

A 'global' SSB repair mechanism, as proposed by Keith Caldecott (Caldecott, 2008), is comprised of four principal stages: detection, end processing, gap filling and ligation. Several pathways involving SSB repair (i.e., base excision repair, sugar damage repair, abortive TOP1-SSB repair) have distinct enzyme complexes facilitating each step of the response. It will be of interest to investigate a potential

role of these different enzyme complexes in the repair spontaneous centromeric SSBs. It is, however, intriguing to note that global SSB repair pathways are thought to occur rapidly (i.e., within minutes) (Frankenberg-Schwager, 1989). The observation that SSBs accumulate at centromeres therefore suggests that either their incidence is extremely frequent, or that they are repaired with unusually slow kinetics. To my knowledge, no work has addressed any spatial heterogeneity in SSB repair pathways, so it is unclear whether centromeres specifically could be refractory to the canonical and rapid SSB repair pathways. SSB repair pathways, like other DNA damage repair pathways, rely on detection and signalling mechanisms. It will therefore be crucial to address whether these signalling pathways are suppressed, analogous to the γ H2A.X suppression described in **Chapter 3**. A spatial disparity in SSB repair signalling could perhaps explain why HR might play a more predominant role in the repair of SSBs at centromeres compared to the rest of the genome, and why SSBs are able to accumulate at centromeres despite the presence of highly efficient and global SSB repair pathway.

In addition, the predominant SSBs observed at human centromeres may alternatively represent intermediates of a DNA repair pathway itself, suggesting the activity of alternative DNA repair pathways at human centromeres. Repair pathways such as mismatch repair (MMR) and ribonucleotide excision repair (RER) both involve transient SSB intermediates. In humans, base mismatch recognition is facilitated by the MSH2-6 complex, which initiates downstream strand nicking and repair synthesis (Jiricny, 2006). Indeed, mass spectrometry analysis of replicating human centromeric alpha-satellite bacterial artificial chromosomes (BACs) using *Xenopus laevis* egg extract revealed an enrichment of the MMR complex MSH2-6 as well as MRE11-RAD50, XRCC1, XRCC5 and PARP1 (Aze *et al.*, 2016). Therefore, MMR may be active on centromeric DNA, and may initiate the nicks described here. However, MSH2-6 in *S. cerevisiae* has also been shown to interact with Holliday Junctions (Marsischky *et al.*, 1999), so the exact mode of MSH2-6 recruitment to and function at alpha satellite DNA remains unclear. Alternatively, ribonucleotide excision repair (RER) is initiated by RNaseH for the removal of mis-incorporated ribonucleotides (Eder *et al.*, 1993; Rydberg and Game, 2002). Curiously, GLOE-seq analysis of SSBs genome-wide revealed a slight bias towards leading strand synthesis, which experiences elevated levels of mis-incorporated ribonucleotides (Nick McElhinny *et al.*, 2010). The authors subsequently proposed that ribonucleotide incorporation significantly contributes to spontaneous SSBs genome-wide (Sriramachandran *et al.*, 2020), opening the possibility that centromeric SSBs may in fact reflect RER intermediates. However, as with MMR, RER is finalized by nick ligation with DNA ligase I (Sparks *et al.*, 2012). Therefore, if centromeric SSBs do indeed reflect repair intermediates, it stands to reason that either the frequency of repair is extremely high or that the downstream ligation is inefficient.

Although this study does not comprehensively address the role of all DNA repair pathways in the repair of centromeric DNA breaks, the establishment of exo-FISH provides a useful tool with which questions

such as these can be addressed in future studies. HR appears to play at least some role in the repair of centromeric breaks, but it will be important to determine how centromere integrity is maintained with respect to other potential repair mechanisms.

Chapter 5

Investigating RAD51 function in centromere specification

5.1 Introduction

5.1.1 CENP-A in the specification of centromeres

CENP-A positioning within chromatin spatially defines centromere functionality in terms of kinetochore assembly and microtubule attachment (Heun *et al.*, 2006; Ho *et al.*, 2014; Logsdon *et al.*, 2019; Mendiburo *et al.*, 2011; Van Hooser *et al.*, 2001). As fully described in **section 1.3.3** and **section 1.3.4**, CENP-A has a well-characterized self-propagating deposition pathway mediated by its chaperone HJURP and the MIS18 complex (Carroll *et al.*, 2010; Dunleavy *et al.*, 2009; Foltz *et al.*, 2009; Moree *et al.*, 2011; Shuaib *et al.*, 2010). In brief, the presence of pre-existing CENP-A directs the deposition of new CENP-A within centromeric chromatin in early G1, following each round of cell division (**Figure 1.7**) (Jansen *et al.*, 2007; Mellone *et al.*, 2011).

5.1.2 Alternative pathways for CENP-A deposition

The presence of CENP-A nucleosomes within non-centromeric chromatin as well as the formation of neocentromeres implies that alternative chromatin landscapes can support CENP-A deposition (Bodor *et al.*, 2014). However, our current understanding of alternative mechanisms promoting CENP-A deposition remains limited. In recent years, roles for centromeric transcription and, to a lesser extent, DNA damage have been increasingly studied as potential alternative mechanisms.

5.1.2.1 Transcription in the epigenetic definition of centromere identity

A role for transcription has been recently investigated as a potential contributor to centromere definition. Centromeres are transcribed by RNA Polymerase II to produce non-coding RNA (ncRNA) that is re-

quired for the correct localization of CENP-A, HJURP and other centromere-associated factors (Blower, 2016; Chan *et al.*, 2012; Quenet and Dalal, 2014; Rosic *et al.*, 2014). RNA polymerase II and other transcription-associated factors localize to human centromeres in mitosis and disrupting transcription results in CENP-A deposition defects (Chan *et al.*, 2012; Quenet and Dalal, 2014). In support of this notion, histone modifications indicative of active transcription (e.g., H3K4me2 and H3K9 acetylation) as well as chromatin remodelling factors associated with active transcription (e.g., FACT, RbAp48) have been shown to promote CENP-A deposition (Bergmann *et al.*, 2011; Chen *et al.*, 2015; Furuyama *et al.*, 2006; Ohzeki *et al.*, 2012; Okada *et al.*, 2009). Curiously, excessive centromeric transcription can also interfere with CENP-A chromatin occupancy, suggesting that while transcription is mechanistically important, it's levels must be finely regulated to ensure optimal centromere function (Bergmann *et al.*, 2011).

An ongoing debate exists on precisely how centromeric transcription promotes CENP-A deposition, and why such additional specification should exist. Some evidence supports a model in which chromosome-specific centromeric ncRNAs remain associated at their origin centromeres, complex with CENP-A and CENP-C, and directly recruit CENP-A to centromeres to facilitate its deposition (McNulty *et al.*, 2017; Rosic *et al.*, 2014; Wong *et al.*, 2007). However, other reports claim that centromeric RNAs do not remain associated with centromeres, suggesting that the physiological importance of centromeric transcription is through establishing a permissive chromatin environment for CENP-A deposition rather than a direct role of the RNA itself (Bury *et al.*, 2020). Centromeric transcription has alternatively been proposed to convert CENP-A from a transient chromatin association to stable incorporation (Bobkov *et al.*, 2018). However, the temporal discrepancy between centromeric transcription and CENP-A deposition argues against this hypothesis. Indeed, centromeric transcription is thought to peak in G2 and mitosis, although centromeric transcripts have been detected throughout the cell cycle by other groups (Bobkov *et al.*, 2018; Bury *et al.*, 2020; Chan *et al.*, 2012; Ferri *et al.*, 2009; Liu *et al.*, 2015; McNulty *et al.*, 2017).

5.1.2.2 DNA repair in the epigenetic definition of centromere identity

As described above, the recruitment of CENP-A to ectopic sites and formation of neocentromeres implies the presence of CENP-A assembly mechanisms independent of the feed-forward centromere-defined deposition pathway (Bodor *et al.*, 2014). The precise mechanisms of neocentromere formation in humans are poorly understood, although one study revealed a close association between neocentromere formation sites and rearrangement boundaries following chromosome breakage (Ventura *et al.*, 2003). An association between neocentromere formation and recombination sites has also been described in chromosome 13-derived neocentromeres, and it has been subsequently proposed that genome rearrangements may initiate neocentromere formation (Marshall *et al.*, 2008).

On the basis of these observations, an alternative centromere specification pathway could be related to DNA damage and repair. As described above (**section 1.3.4**), HJURP is entirely sufficient for CENP-A deposition. Curiously, HJURP was originally identified as a DNA damage repair protein involved in HR repair and named after its ability to directly bind Holliday junctions *in vitro* (Kato *et al.*, 2007). In addition, HJURP directly binds both members of the MRN complex as well as structural intermediates of the HR repair pathway (Barnhart *et al.*, 2011; Kato *et al.*, 2007). It is thereby feasible that DNA damage or repair intermediates may directly promote or reinforce HJURP recruitment and, in turn, CENP-A deposition. Although very few studies have investigated a potential connection between DNA damage, repair and centromere specification, some preliminary evidence may support its feasibility.

First, HJURP forms nuclear foci in response to DNA damage and CENP-A itself can also be recruited to sites of DNA DSBs (Kato *et al.*, 2007; Yilmaz *et al.*, 2021; Zeitlin *et al.*, 2009), providing evidence that DNA damage leads to the recruitment of centromere-associated proteins. Curiously, CENP-A recruitment to DNA breaks seems to depend on active DNA repair mechanisms (in this report, NHEJ repair factors), suggesting that it is the process of DNA repair, rather than the DNA damage itself, underlying CENP-A recruitment (Zeitlin *et al.*, 2009). Second, two studies have identified potential roles for HR repair factors in promoting CENP-A occupancy at centromeres. In *C. albicans*, depletion of HR proteins Rad51 and Rad52, but not NHEJ repair factors, resulted in a reduction of CENP-A (Cse4) at centromeres (Mitra *et al.*, 2014). A recent study also reported the reduction of centromeric CENP-A upon depletion of HR mediator complex factor BRCA1 (Racca *et al.*, 2021). Therefore, it remains conceivable that active DNA repair somehow functions to promote CENP-A chromatin occupancy at centromeres. Importantly, these observations would collectively imply that a DNA repair-induced CENP-A recruitment pathway is integrated with canonical centromere specification mechanisms at endogenous centromeres, and not specific to *de novo* neocentromere formation. However, few additional studies have addressed this hypothesis and no consensus has been reached as to the mechanisms underlying these observations.

5.2 Results

5.2.1 Reduction of CENP-A in the absence of HR recombinase RAD51

While assessing the localization of RAD51 to endogenous centromeres in HeLa HAC2-4 cells (**Figure 4.4**), I observed that CENP-A association with endogenous centromeres was reduced to ~50% upon BRCA2 depletion (data not shown). Given the previously reported association between HR factors and CENP-A occupancy at centromeres (Mitra *et al.*, 2014; Racca *et al.*, 2021), I sought to further investigate this association through an independent assay and in the non-cancerous human hTERT-RPE1 cell line.

First, I asked whether CENP-A was lost at centromeres following long-term depletion of the central mammalian recombinase RAD51. Depleting RAD51 for 96 hours in dividing hTERT-RPE1 cells, I determined the level of CENP-A at centromeres by immunofluorescence. I verified that RAD51 and CENP-A, as a positive control, could be efficiently depleted 72 hours after siRNA treatment (as shown later in **Figure 5.3A, C**). The depletion of CENP-A was not very apparent by Western blot even after 72 hours of depletion, presumably due to its previously described stability on chromatin (Bodor *et al.*, 2013; Falk *et al.*, 2015; Jansen *et al.*, 2007).

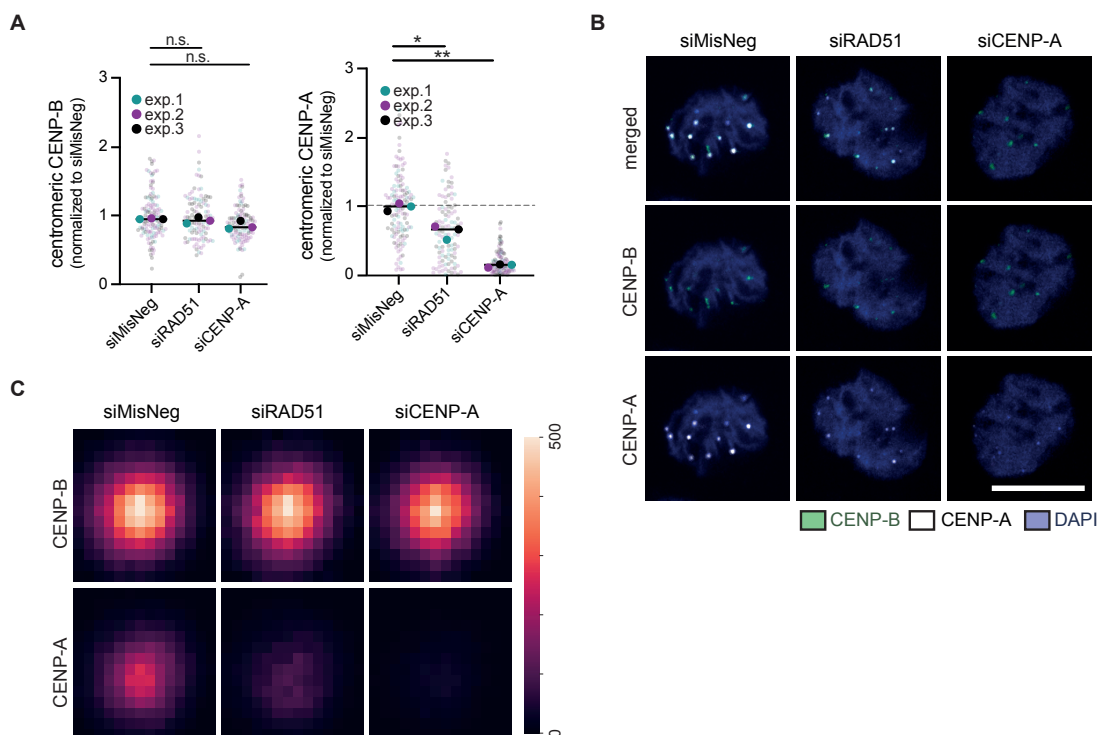


Figure 5.1: Reduction of CENP-A in the absence of HR recombinase RAD51. (A, B, C) Quantification, representative images, and representative median focus analysis of immunofluorescence staining against CENP-A and CENP-B. hTERT-RPE1 cells were fixed 96 hours after siRNA-mediated depletion of RAD51, or CENP-A. CENP-A and CENP-B levels were subsequently quantified at CENP-B-defined centromeres, as described in **Figure 3.9**. Scale bar represents 10 μ m. IF signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median value of a cell, with the median of each biological replicate indicated in darker hues. At least 30 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (*p < 0.05, **p < 0.01).

Co-staining for both CENP-A and CENP-B, I used the previously described centromere box analysis (**Figure 3.9A**) to quantify the levels of centromeric CENP-B and centromeric CENP-A at CENP-B-defined centromeres. Upon depletion of RAD51 or CENP-A itself, I observed a consistent reduction of centromeric CENP-A without any changes in CENP-B levels (**Figure 5.1A, B, C**). Indeed, even though the depletion of CENP-A was only minimally apparent by Western blotting (**Figure 5.3C**), there was

an almost complete reduction of CENP-A at CENP-B-defined centromeres upon CENP-A depletion. This observation highlighted that the analysis of CENP-A protein levels by Western blotting is a poor indicator of centromeric occupancy. These data also supported a role for the HR recombinase RAD51 in promoting CENP-A occupancy at centromeres. Although the decrease in CENP-A upon RAD51 depletion is relatively mild, it should be noted that RAD51-depleted cells typically exhibit striking cell cycle arrests, thereby preventing cellular proliferation and limiting CENP-A turnover. Therefore, the changes in CENP-A centromere occupancy upon RAD51 depletion are likely an under-estimation of the role of RAD51 in maintaining centromeric CENP-A levels.

5.2.2 CENP-A reduction upon RAD51 depletion is independent of active DNA replication

As with the centromeric DNA breaks described in **Chapter 3**, I sought to determine whether the reduction of CENP-A upon RAD51 depletion was dependent on active DNA replication, to determine whether cell cycle effects or replication stress upon RAD51 depletion could interfere with CENP-A centromeric occupancy. Using the same experimental protocol for serum starvation (as described and verified in **Figure 4.5**), I quantified CENP-A centromere occupancy upon RAD51 and CENP-A depletion. Depletion of RAD51 or CENP-A itself reduced CENP-A occupancy at centromeres even in the absence of active DNA replication (**Figure 5.2A, B, C**), suggesting that defective centromeric replication does not underlie the reduction of CENP-A upon RAD51 depletion.

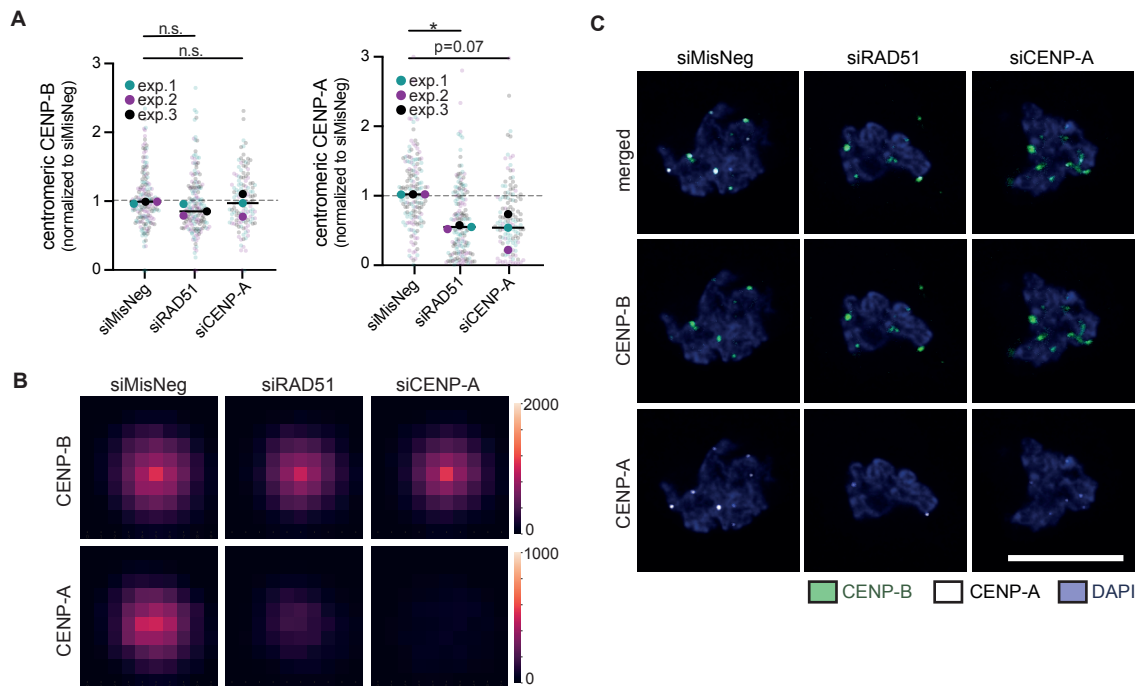


Figure 5.2: Reduction of CENP-A upon RAD51 depletion is independent of active DNA replication. (A, B, C) Quantification, representative images, and representative median focus analysis of immunofluorescence staining against CENP-A and CENP-B. hTERT-RPE1 cells were serum-starved for 24 hours before undergoing 96 hours siRNA-mediated depletion of RAD51 or CENP-A. CENP-A and CENP-B levels were subsequently quantified at CENP-B-defined centromeres, as described in **Figure 3.9**, 96 hours after depletion. Scale bar represents 10 μ m. IF signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median value of a cell, with the median of each biological replicate indicated in darker hues. At least 30 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$, ** $p < 0.01$).

5.2.3 CENP-A protein and mRNA levels in the absence of RAD51

The initial observation that HR mediator RAD51 promotes CENP-A levels at centromeres was interesting in light of previously reported links between centromere specification and DNA repair (as described in **section 5.1.2**). However, it was important to determine whether the reduction of CENP-A was in fact due to defective RAD51 function, rather than an indirect outcome of total CENP-A protein level disruptions. Therefore, I assessed total CENP-A protein levels by Western blotting upon RAD51 depletion in hTERT-RPE1 dividing cells (**Figure 5.3A**). Intriguingly, depletion of RAD51 resulted in a reduction of CENP-A protein levels, suggesting a potential indirect mechanism through which RAD51 promotes CENP-A occupancy at centromeres. In hTERT-RPE1 quiescent (i.e., serum-starved) cells, a similar phenomenon was observed whereby RAD51 depletion reduced CENP-A protein levels (**Figure 5.3B**).

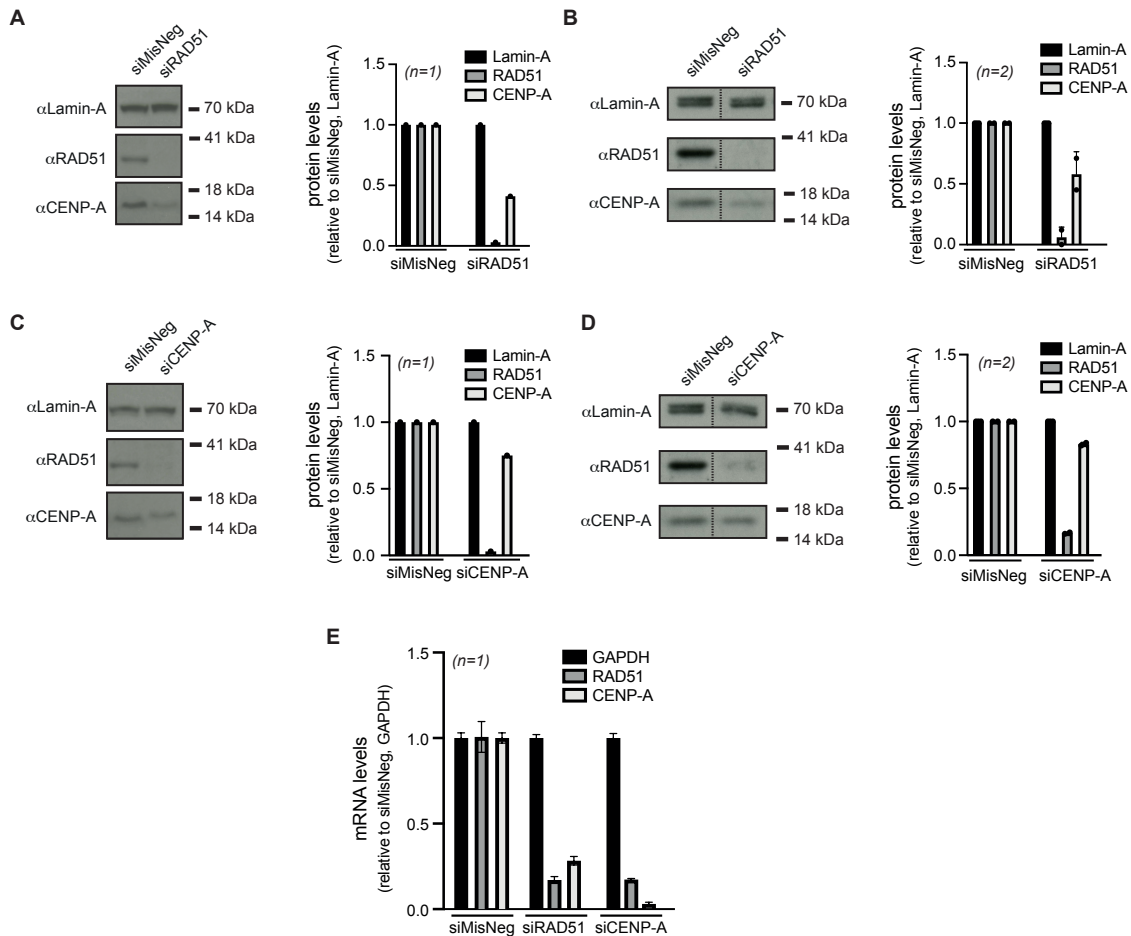


Figure 5.3: CENP-A protein and mRNA in the absence of RAD51. (A) Western blot depiction and quantification of Lamin-A, RAD51 and CENP-A protein levels following 72 hours RAD51 siRNA-mediated depletion in asynchronous hTERT-RPE1 cells. Protein levels were quantified as the total background-subtracted signal intensity of protein bands, normalizing to both the universal negative control siMisNeg and Lamin-A. (B) Western blot depiction and quantification, as in (A), for hTERT-RPE1 quiescent cells serum-starved 24 hours prior to siRNA treatment. (C) Western blot depiction and quantification of Lamin-A, RAD51 and CENP-A protein levels following 72 hours CENP-A siRNA-mediated depletion in asynchronous hTERT-RPE1 cells. Quantification as in (A). (D) Western blot depiction and quantification, as in (A), for hTERT-RPE1 quiescent cells serum-starved 24 hours prior to siRNA treatment. (E) GAPDH, RAD51 and CENP-A mRNA levels assessed by RT-qPCR following 96 hours depletion of RAD51, CENP-A or the universal negative control siMisNeg. mRNA levels were calculated as $2^{-\Delta\Delta C_t}$, using siMisNeg and GAPDH for normalization.

Curiously, a reciprocal relationship was noted whereby the depletion of CENP-A in both dividing and non-dividing hTERT-RPE1 cells resulted in a striking reduction in RAD51 protein levels (**Figure 5.3C, D**). Initially, this suggested that these proteins may be part of a mutual protein complex, whereby the depletion of one factor destabilizes other complex constituents. However, analysis of mRNA levels in hTERT-RPE1 quiescent cells subsequently refuted this hypothesis (**Figure 5.3E**). Depletion RAD51 resulted in a notable reduction of CENP-A mRNA levels. Equally, the depletion of CENP-A reduced

RAD51 mRNA levels. Therefore, the observed changes in protein levels most likely reflect disruptions in upstream transcription levels rather than protein stability.

Several potential explanations could account for the disruption in CENP-A transcription upon RAD51 depletion. First, this effect could be due to off-target siRNA effects. This explanation is unlikely since the siRNA concentrations used were towards the lower end of the manufacturers recommendations, and a dual-targeting siRNA system was used to target RAD51, thereby minimizing the chances of off-target effects. Second, accumulated DNA damage in the absence of HR-mediated repair could interfere with the transcription of CENP-A. For both the first and second possibilities, however, the reciprocity of mRNA disruption between RAD51 and CENP-A argues against this explanation. Finally, it is possible that depletion of RAD51 or CENP-A induce a cellular state whereby transcription networks are altered, perhaps through activation of the p53 response. Indeed, depletion of CENP-A has been shown to induce premature p53-dependent cellular senescence, which can directly impact RAD51 protein expression (Arias-Lopez *et al.*, 2006; Maehara *et al.*, 2010).

In support of the third possibility, I observed different trends using a HeLa cervical cancer cell line in which the p53 response is innately absent. Analysis of proteins levels upon RAD51 and CENP-A depletion in a SNAP-tagged CENP-A HeLa cell line revealed that depletion of RAD51 had little impact on CENP-A levels, and depletion of CENP-A similarly did not impact RAD51 levels to the same extent (**Figure 5.4A**). Therefore, it seems plausible that disruptions in CENP-A transcript and protein levels upon HR factor depletion in hTERT-RPE1 cells (**Figure 5.3**) are at least partially due to p53-induced transcription changes.

Importantly, in the HeLa CENP-A-SNAP system, without any evident changes in CENP-A protein levels upon HR factor depletion, RAD51 depletion continued to reduce centromeric CENP-A occupancy equivalent to that of CENP-A depletion itself (**Figure 5.4B, C**). Therefore, these data support a functional link between RAD51 activity and CENP-A centromere occupancy.

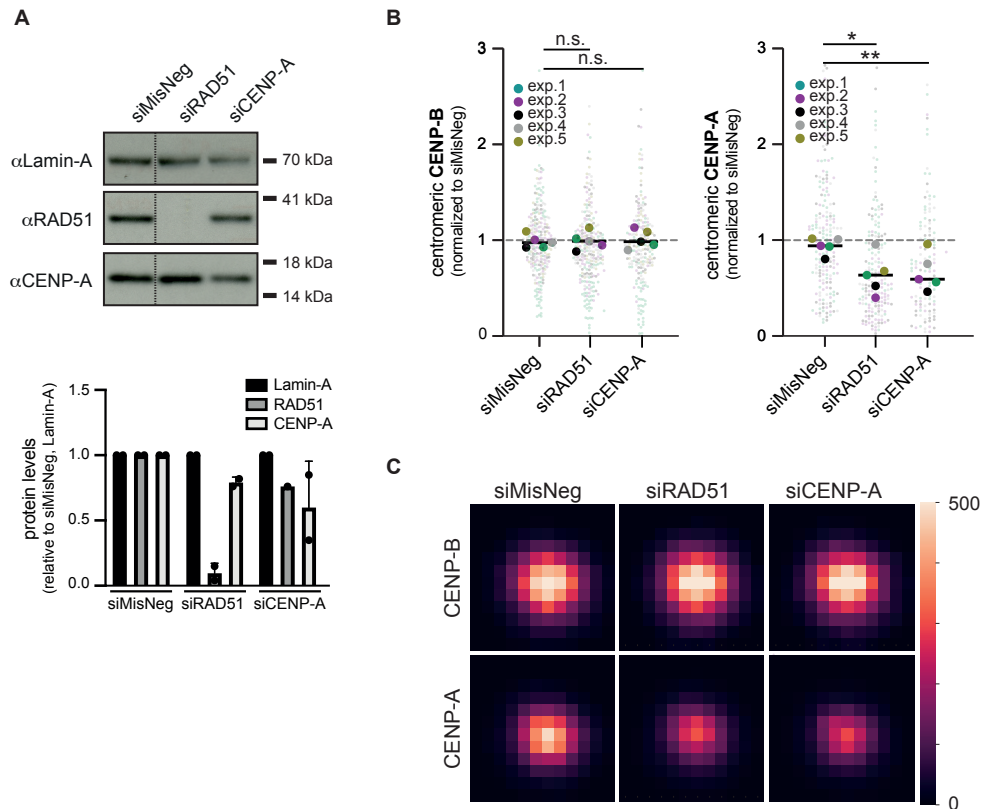


Figure 5.4: **CENP-A centromeric occupancy in HeLa CENP-A-SNAP.** (A) Western blot and quantification verifying the depletion of RAD51 and CENP-A in HeLa CENP-A-SNAP cells. Lamin-A is indicated as loading control. Protein levels were quantified as the total background-subtracted signal intensity of protein bands, normalizing to both the universal negative control siMisNeg and Lamin-A. (B, C) Quantification and representative median focus analysis of immunofluorescence staining against CENP-A and CENP-B. HeLa CENP-A-SNAP cells were fixed 72 hours after siRNA-mediated depletion of RAD51, or CENP-A. CENP-A and CENP-B levels were subsequently quantified at CENP-B-defined centromeres, as described in **Figure 3.9**. IF signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median value of a cell, with the median of each biological replicate indicated in darker hues. At least 30 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$, ** $p < 0.01$).

5.2.4 HeLa CENP-A-SNAP to distinguish CENP-A deposition and stability

Previously, DNA replication was excluded as a potential mechanism by which RAD51 influences CENP-A centromere occupancy (**Figure 5.2**). However, there remained many potential mechanisms by which RAD51 activity could impact CENP-A centromeric occupancy. These mechanisms ranged from defective DNA repair impacting local nucleosome stability to non-repair functions disrupting the centromeric chromatin environment to interfere with the deposition of new CENP-A. I sought to narrow down the possibilities by determining whether the reduction of CENP-A at centromeres was due to reduced stability within centromeric chromatin, or due to an inability to deposit new CENP-A upon cellular division, or some combination of the two. To address this question, I made use of a HeLa cell line expressing CENP-A tagged to a SNAP moiety. This system was designed to distinguish between CENP-A deposition and

stability (Bodor *et al.*, 2012; Jansen *et al.*, 2007). As previously described, the HeLa CENP-A-SNAP cell line had been validated for the depletion of RAD51 as well as the disruption to CENP-A centromeric occupancy (**Figure 5.4**).

5.2.4.1 CENP-A-SNAP deposition in the absence of RAD51

In the HeLa CENP-A-SNAP system, CENP-A is tagged with a SNAP moiety, which can be either be rapidly fluorescently labelled ('pulsed') or blocked ('quenched') *in vivo* (**Figure 5.5A**). The deposition of new CENP-A can therefore be assessed in cells that are initially quenched to block all available SNAP moieties of existing CENP-A and subsequently pulsed to fluorescently label any newly synthesized CENP-A (**Figure 5.5B**). The extent of newly deposited CENP-A is then determined by quantifying SNAP intensity at CENP-B-defined centromeres. Typically, there is a 24-hour chase window between quenching and pulsing to allow at most one round of CENP-A-SNAP incorporation into centromeric chromatin. However, a chase time course over 24-72 hours was performed including siRNA treatment against the negative control (**Figure 5.5C, D**) and it was evident that in these conditions, SNAP deposition (i.e., SNAP foci) could only be observed 48 hours after the initial quenching. This is conceivably because the RNAi treatment itself increases the division time of cells, therefore requiring up to 48 hours for a single round of CENP-A-SNAP incorporation.

With these conditions optimized, I tested the impact of RAD51 depletion on the deposition of new CENP-A-SNAP, with CENP-A depletion included as a positive control. One important caveat to this analysis is that if there are any significant disruptions in cell-cycle progression upon depletion of a candidate protein, fewer cells may have incorporated any CENP-A-SNAP (i.e., having not reached early G1), which may produce false positives. Therefore, to ensure that cell cycle perturbations were not biasing the quantification, only cells that were classified as 'SNAP-positive' were included in the quantification. More specifically, if cells had >10 SNAP foci, as detected by an automated algorithm, they are classed as SNAP-positive and included in the quantification. From this analysis, RAD51 depletion reduced the deposition of new CENP-A-SNAP at CENP-B-defined centromeres to a similar extent as CENP-A depletion itself (**Figure 5.5E, F**). The recombinase RAD51 therefore promotes CENP-A deposition at centromeres.

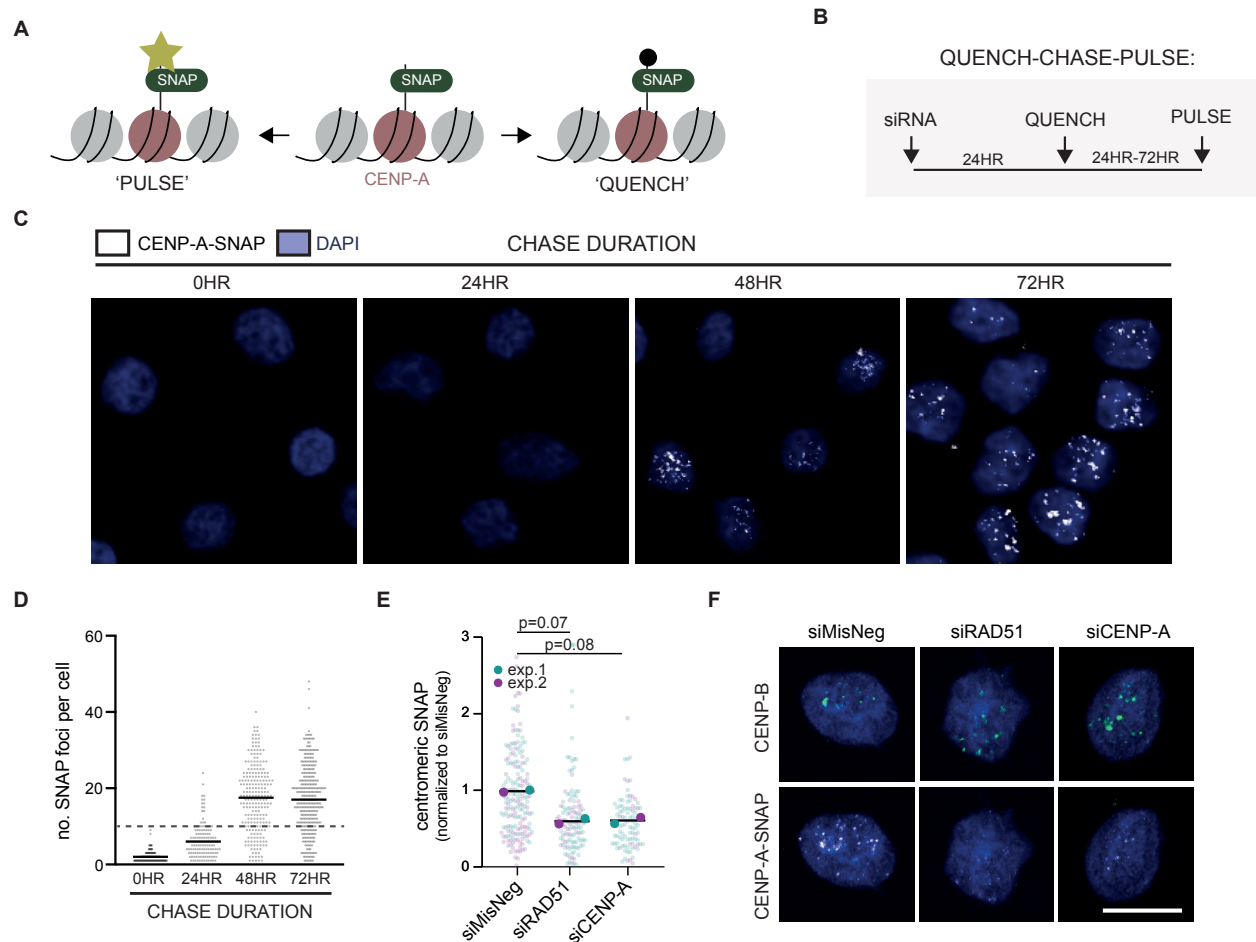


Figure 5.5: CENP-A deposition in the absence of RAD51. (A) Schematic of CENP-A-SNAP functionality in HeLa system. CENP-A molecules are tagged with a SNAP moiety, which can either be covalently labelled with a fluorescent molecule ('pulsed') or blocked with a non-fluorescent chemical modification ('quenched'). (B) Experimental schematic of CENP-A deposition assay ('quench-chase-pulse') using HeLa CENP-A-SNAP cells. After 24 hours siRNA-mediated depletion of candidate factors, cells are quenched to block all existing CENP-A-SNAP molecules. After the indicated time (24-72 hours), any newly incorporated CENP-A is pulsed to quantify newly deposited CENP-A molecules. (C, D) Representative images and quantification of chase optimization of quench-chase-pulse assay to determine pulse timing for the detection of newly loaded CENP-A-SNAP molecules. (E, F) Quantification and representative images of quench-chase-pulse assay following the siRNA treatment against RAD51, CENP-A or the universal negative control (siMisNeg). Quantification of SNAP within CENP-B-defined centromeric loci was performed as in **Figure 3.9**. Importantly, only cells classed as SNAP-positive (i.e., > 10 SNAP foci) were included in the quantification. Each data point represents a median value of a cell, with the median of each biological replicate indicated in darker hues. At least 20 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$, ** $p < 0.01$).

5.2.4.2 CENP-A-SNAP stability in the absence of RAD51

I next wanted to determine whether RAD51 depletion would have any impact on the stability of existing CENP-A within centromeric chromatin. Indeed, one intriguing possibility is that the failure to repair centromeric DNA breaks, for example, might result in nucleosome eviction or remodelling. To test this possibility, I used a complementary experiment measuring CENP-A stability over time upon a single

pulse (**Figure 5.6A, B**). The levels of fluorescently labelled CENP-A-SNAP decreased by roughly two-fold every 24 hours, consistent with cellular division halving the existing CENP-A chromatin association. However, the depletion of RAD51 did not appear to have any significant impact of the levels of CENP-A-SNAP after 24 or 48 hours pulsing (**Figure 5.6C, D**). Together, these data suggest that RAD51 promotes CENP-A centromere occupancy by enabling the deposition of new CENP-A, rather than the retention of existing CENP-A.

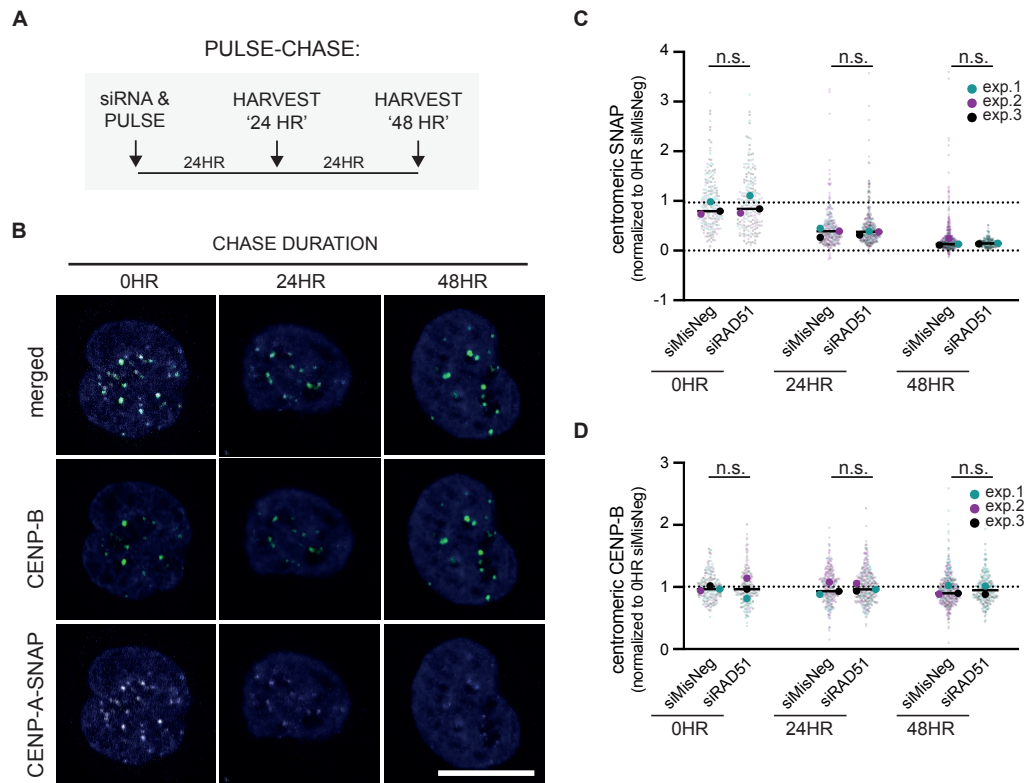


Figure 5.6: CENP-A stability in the absence of RAD51. (A) Experimental schematic of CENP-A stability assay ('pulse-chase') using HeLa CENP-A-SNAP cells. siRNA-mediated depletion of candidate factors is performed directly after initial pulse. Centromeric SNAP intensity is measured every 24 hours to determine chromatin stability. (B) Representative images of SNAP foci intensity 0, 24 and 48 hours after initial pulse in untreated cells. Scale bar represents 10 μ m. (C, D) Quantification of SNAP and CENP-B signal intensities at CENP-B-defined centromeres 0, 24 and 48 hours following SNAP pulsing and siRNA mediated depletion of RAD51 or the universal negative control (siMisNeg). CENP-B is used to define centromeres and the SNAP intensity is quantified as in **Figure 3.9**. Each data point represents a median value of a cell, with the median of each biological replicate indicated in darker hues. At least 20 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$, ** $p < 0.01$).

5.2.5 Centromeric transcription in the absence of RAD51

Based on the observation of reduced CENP-A deposition, but not stability, upon RAD51 depletion, the potential mechanisms by which RAD51 promotes CENP-A centromere occupancy were significantly narrowed down. The deposition of *de novo* CENP-A at centromeres is a well-characterized mechanism

that occurs in late mitosis / early G1 of each cell cycle (Jansen *et al.*, 2007; Mellone *et al.*, 2011). As described fully in **section 1.3.4**, HJURP and the MIS18 complex function to direct CENP-A deposition to pre-existing CENP-A-containing chromatin. In addition, a potential role for centromeric transcription has emerged in promoting CENP-A deposition (**section 5.1.2.2**). Based on this, I reasoned that defects in CENP-A deposition upon RAD51 depletion could be related to centromeric transcription or the recruitment of CENP-A chaperones such as HJURP.

I set out to determine whether HR could impact non-coding transcription at centromeres. Centromeric transcripts were detected using a RT-qPCR approach, whereby total cellular RNA is reverse-transcribed into cDNA followed by quantification by qPCR. I sought to quantify the RNA levels of centromeric transcripts as well as another RNA Polymerase II-transcribed, non-coding RNA (e.g., telomeric TERRA) and the RNA Polymerase I-transcribed 5S rDNA, as negative controls. First, I optimized the cDNA input to be within a linear range of amplification to allow the detection of changes in cDNA levels (**Figure 5.7A**). I tested 5S rDNA and two telomeric TERRA probes, two centromeric probes, as well as two non-coding but non-repetitive RNA controls. From this, it was clear that at higher cDNA inputs, RNA from repetitive regions resulted in a saturation of the qPCR signal. Therefore, 10 ng cDNA input (or an input corresponding to the equivalent Ct value) was used to prevent be within the optimized linear amplification range. In addition, the analysis of RNA from repetitive regions can be challenging due to interference from the high abundance of contaminating genomic DNA (gDNA). Despite two successive rounds of DNA removal using DNaseI prior to reverse transcription, there was still a notable qPCR signal from the non-reverse-transcribed control samples. Therefore, I verified that the qPCR signal from the centromeric probes using 10 ng cDNA was still sufficient to detect RNA levels (RT+) above that of the gDNA contamination (RT-) (**Figure 5.7B**). For alphoid transcripts, this was 66-fold enrichment of RNA over contaminating gDNA using 10 ng cDNA. However, the D17Z-1 centromeric probe did not show sufficient enrichment of the RT+ signal over RT-, and so this probe was discontinued.

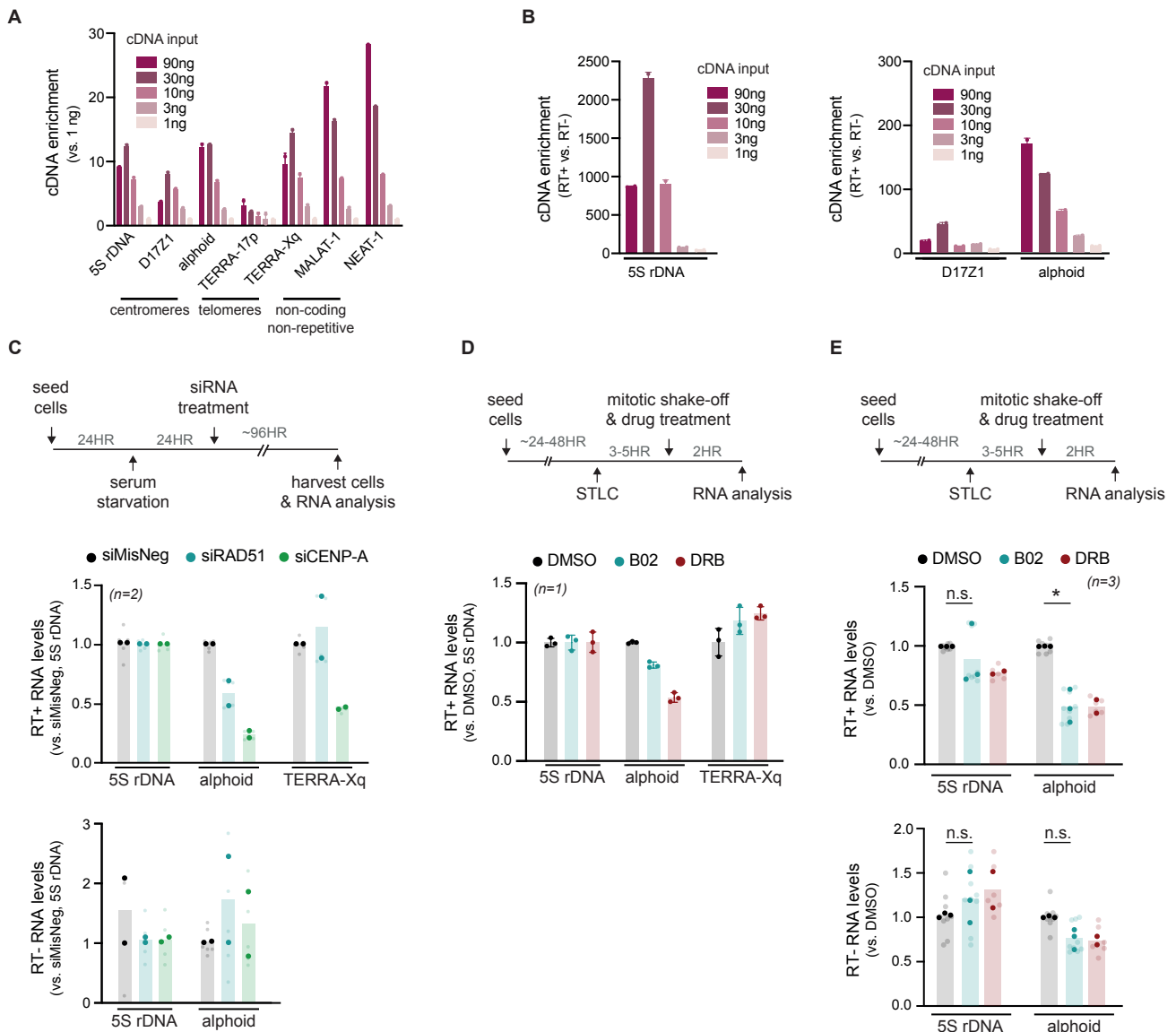


Figure 5.7: Centromeric transcription in the absence of RAD51. (A) Optimization of RT-qPCR conditions for the detection of RNA produced at repetitive regions. Probes for 5S rDNA, centromeres (D17Z-1, alphoid), telomeres (TERRA-17P, TERRA-Xq) and non-coding non-repetitive regions (MALAT-1, NEAT-1) were tested. After reverse-transcribing whole cell RNA, increasing inputs of cDNA were quantified by qPCR, after which the cDNA enrichment was calculated by the fold enrichment compared to the 1ng cDNA input. (B) Verification of RNA-specific signal (RT+) enrichment over contaminating genomic DNA signal (RT-) for 5S rDNA, D17Z1 and alphoid qPCR probes over various cDNA inputs. (C) Assessment of whether long-term depletion of RAD51 impacts centromeric transcript levels in serum-starved hTERT-RPE1 cells. After 96 hours siRNA treatment, RNA was extracted and transcripts were quantified for 5S rDNA, alpha satellite transcripts, and non-coding telomeric transcripts TERRA-Xq by RT-qPCR. (D) Assessment of whether short-term inhibition of RAD51 with the small molecule inhibitor B02 impacts centromeric transcript levels in mitotic hTERT-RPE1 cells. Mitotic cells were enriched with 3-5 hours treatment using S-trityl-L-cysteine (STLC), collected by mitotic shake-off and treated with DMSO, B02 or DRB for 2 hours in the presence of STLC prior to RNA extraction. As in (C), 5S rDNA, alpha satellite, and telomeric ncRNA (TERRA-Xq) were analyzed and normalized to DMSO and 5S rDNA. (E) As in (D) but for HeLa CENP-A-SNAP mitotic cells instead of mitotic hTERT-RPE1 cells. For all qPCR reactions, the equivalent of 10-30ng cDNA was used for input (corresponding to optimized Ct values) and RT- controls were included when possible. Where indicated, the averages of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$).

Having optimised the conditions for RT-qPCR detection of centromeric transcripts, I tested the long-term depletion of RAD51 in quiescent hTERT-RPE1 cells and observed a consistent reduction in alpha-satellite transcripts, but not telomeric transcripts (**Figure 5.7C**). Additional analysis of the RT- sample was included, where possible, to verify that the change in qPCR signal upon RAD51 depletion was not an indirect reflection of the levels of contaminating gDNA. Curiously, the depletion of CENP-A seemed to markedly reduce both alpha-satellite as well as telomeric transcripts. Therefore, although there appeared to be a disruption in centromeric transcription upon HR factor depletion, this interpretation is complicated by the observation that CENP-A depletion itself interferes with centromeric transcription. In other words, it is challenging to discern whether RAD51 first reduces centromeric transcription, thereby impacting CENP-A deposition, or whether a decrease in CENP-A chromatin occupancy is sufficient to produce the observed disruptions in centromeric transcription.

To discern between these two scenarios, centromeric transcripts were once again quantified following short-term RAD51 inhibition using the small-molecule inhibitor B02. After arresting hTERT-RPE1 cells in mitosis with STLC, mitotic cells were collected and treated with B02, the negative control DMSO, or a positive control for transcriptional inhibition, DRB, for 2 hours (**Figure 5.7D**). Short-term inhibition of RAD51 was sufficient to reduce centromeric transcripts, about 50% the effect size of DRB itself, and a similar trend was observed in HeLa CENP-A-SNAP cells (**Figure 5.7E**). Once again, RT- controls were included when possible. Curiously, the RT- samples similarly showed a reduction in alpha satellite levels upon B02 and DRB treatment, but to a lesser extent than the RT+ samples. Overall, under these short-term inhibition conditions, when CENP-A levels could not yet be affected by RAD51 inhibition, centromeric transcription appears to already be disrupted. Therefore, the perturbation of transcription upon RAD51 depletion is likely upstream of CENP-A deposition defects, providing a potential mechanism through which RAD51 can promote CENP-A deposition at centromeres.

5.2.6 Involvement of the canonical RAD51 regulator BRCA2 in centromere specification

During this investigation, I once again sought to determine whether the function of RAD51 in centromere specification was shared by its canonical regulator BRCA2. To that end, several of the experiments described above were performed in tandem with BRCA2 depletion. Remarkably, and unlike the DNA break results described in **Chapter 4**, BRCA2 depletion appeared to largely phenocopy RAD51 depletion in terms of centromere specification. First, long-term BRCA2 depletion in hTERT-RPE1 cells reduced CENP-A occupancy at CENP-B-defined centromeres (**Figure 5.8A**). This was also the case, to a lesser extent, in quiescent hTERT-RPE1 cells (**Figure 5.8B**). As with RAD51 depletion, BRCA2 depletion in the HeLa CENP-A-SNAP cell line also reduced CENP-A centromeric occupancy (**Figure 5.8C**)

and reduced the deposition of new CENP-A-SNAP to centromeres (**Figure 5.8D**). Finally, I measured whether centromeric transcript levels were also dependent on BRCA2 (**Figure 5.8E**). However, BRCA2 depletion in quiescent hTERT-RPE1 cells produced variable results between biological replicates, and so it remains unclear whether BRCA2 function maintains centromeric transcript levels.

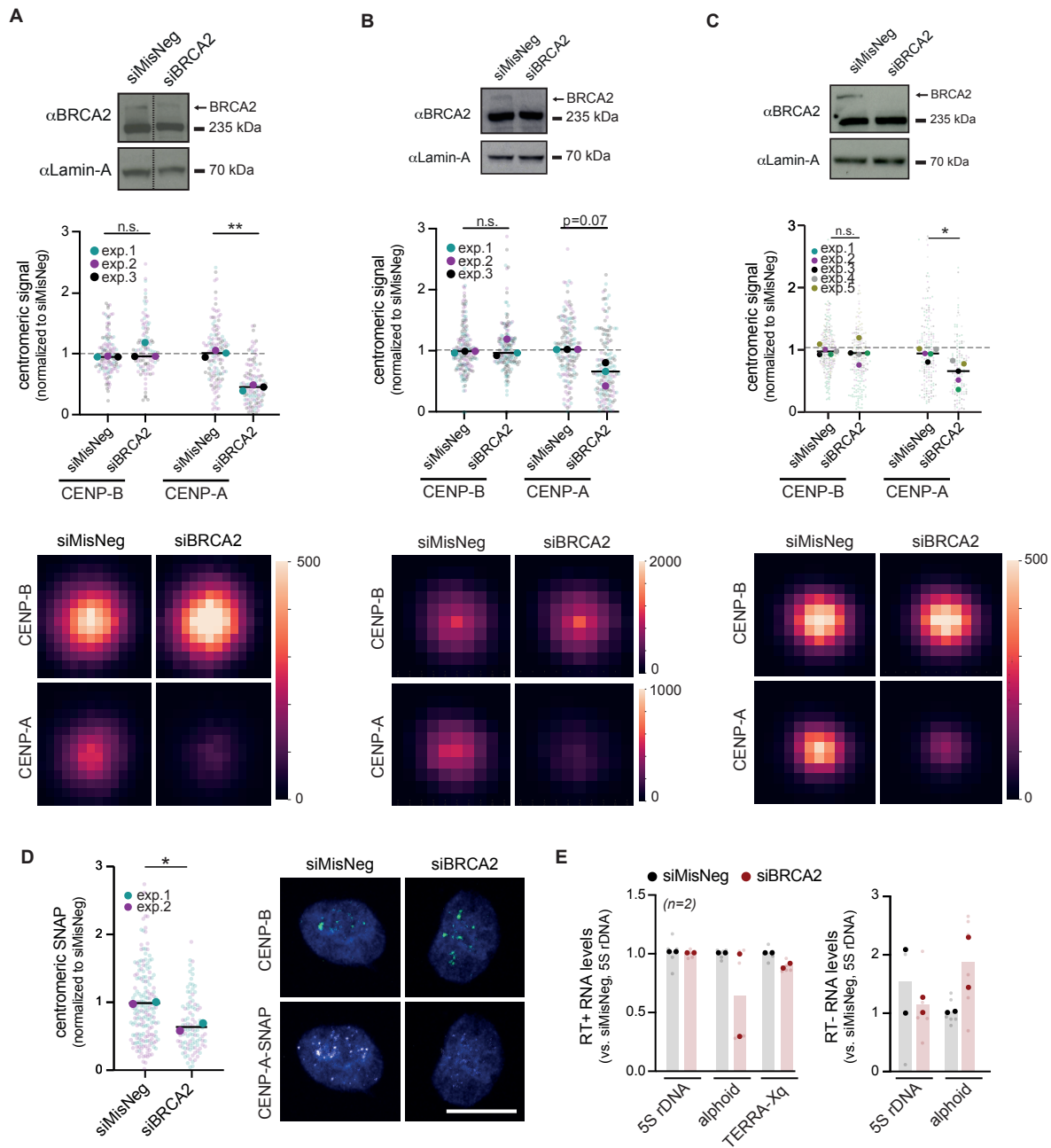


Figure 5.8: Involvement of the canonical RAD51 regulator BRCA2 in centromere specification. (A) As described in **Figure 5.2** but for BRCA2 depletion in asynchronously dividing hTERT-RPE1 cells. Western blot confirming BRCA2 depletion 72 hours after siRNA treatment, quantification of CENP-B and CENP-A within CENP-B-defined centromeres and representative median foci. Immunofluorescence was performed 96 hours after siRNA-mediated depletion of BRCA2. (B) As in (A) but for BRCA2 depletion in serum-starved (i.e., non-dividing) hTERT-RPE1 cells, 96 hours after siRNA-mediated depletion. (C) As in (A) but for BRCA2 depletion in asynchronously dividing HeLa CENP-A-SNAP cells, 72 hours after siRNA-mediated depletion. (D) Quantification and representative images of SNAP quench-chase-pulse assay, as described in **Figure 5.6**. (E) As in **Figure 5.7**, RT-qPCR analysis for the quantification of transcript levels, 96 hours after siRNA-mediated depletion. Transcript levels of 5S rDNA, alpha satellite transcripts, and non-coding telomeric transcripts TERRA-Xq were quantified RT-qPCR. Optimization of probes in **Figure 5.7**, and RT- samples included where possible. All microscopy quantification as described in **Figure 3.9**. Scale bar represents 10 μ m. IF signal intensity is X10,000 arbitrary units (A.U.). Each data point represents a median value of a cell, with the median of each biological replicate indicated in darker hues. At least 30 cells were imaged per experimental condition. The medians of each experimental condition were used to perform a two-sided paired t-test (* $p < 0.05$, ** $p < 0.01$).

Overall, BRCA2 appears to play a comparable role in CENP-A deposition to RAD51, suggesting that this previously unrecognized role of RAD51 in centromere specification may be shared by the canonical HR mediator complex.

5.3 Discussion

5.3.1 Evidence for the role of HR recombinase RAD51 in promoting CENP-A deposition

In this chapter, I investigate a potential role for HR in promoting centromere specification through CENP-A occupancy. In the absence of the central mammalian recombinase RAD51, I observed a reduction of CENP-A at CENP-B-defined centromeres in both dividing and non-dividing hTERT-RPE1 cells (**Figure 5.1, 5.2**). This phenotype was also apparent upon depletion of the canonical RAD51 regulator BRCA2 (**Figure 5.8**). Although initially there were some potentially confounding disruptions in CENP-A transcription and protein levels (**Figure 5.3**), these disruptions appeared to correspond at least partially to cellular p53 status, as p53-negative cell lines did not exhibit similar disruptions (**Figure 5.4A**). Importantly, I confirmed that CENP-A centromere occupancy was reduced upon RAD51 depletion, under conditions where CENP-A protein levels were largely unchanged (**Figure 5.4**). From these results, it is evident that both RAD51 and BRCA2 function to ensure centromere specification. This finding coincides with previous reports of CENP-A (or Cse4 in *S. cerevisiae*) reduction upon HR factor depletion, contributing to a growing body of literature detailing a potential link between DNA repair proteins and centromere specification (Mitra *et al.*, 2014; Racca *et al.*, 2021).

Exploiting the HeLa CENP-A-SNAP cell line to distinguish between CENP-A deposition and stability, it further became evident that depletion of RAD51 or BRCA2 reduced the deposition of CENP-A at centromeres (**Figure 5.5, 5.8**). Although no clear defect in CENP-A stability was observed (**Figure 5.6**), an important caveat to this interpretation is that the experiment is missing a positive control for CENP-A stability defects. For example, including a known CENP-A maintenance factor such as CENP-C or Spt6 could have ensured that the experimental conditions suffice to detect changes in CENP-A stability (Bobkov *et al.*, 2020; Falk *et al.*, 2015).

5.3.2 Specificity of RAD51 at centromeres in CENP-A deposition

An important question stands regarding the specificity of RAD51 function for CENP-A deposition at centromeres. First, it is unclear whether this phenomenon is restricted to CENP-A-containing nucleosomes or applicable to all nucleosomes. Second, it is unclear whether this phenomenon is restricted to centromeres or a genome-wide phenomenon. Indeed, it is entirely plausible that RAD51 supports

nucleosome occupancy on a genome-wide level.

Increased chromatin mobility, for example, occurs in response to DNA breaks both locally at the site of the break as well as genome-wide (Dion and Gasser, 2013; Dion *et al.*, 2012; Herbert *et al.*, 2017; Mine-Hattab and Rothstein, 2012). Although the mechanisms of chromatin mobility changes upon DNA breaks are still being investigated, chromatin decompaction (i.e., removal or remodelling of nucleosomes) is thought to drive changes in chromatin mobility (Amitai *et al.*, 2017; Hauer *et al.*, 2017; Hinde *et al.*, 2014). Therefore, it is feasible that a global decompaction of nucleosomes, including but not specific to CENP-A-containing nucleosomes, occurs upon depletion of the repair factor RAD51, explaining the reduction of centromeric CENP-A. However, it should be noted that other models for damage-induced chromatin mobility changes propose a stiffening of chromatin, rather than chromatin decompaction, as well as deposition, rather than eviction, of H3.3 at sites of DNA damage (Herbert *et al.*, 2017; Luijsterburg *et al.*, 2016). The notion of global chromatin decompaction upon DNA damage is therefore still a matter of debate. In addition, a chromatin decompaction hypothesis would imply that BRCA2-RAD51 depletion impact CENP-A chromatin retention, rather than CENP-A deposition. The observation of defective CENP-A deposition upon BRCA2-RAD51 depletion thereby supports an alternative mechanism. Unfortunately, the current data cannot assess whether this phenomenon is restricted to CENP-A or even centromeres, warranting further investigation into the specificity of this previously unrecognised function for HR factors BRCA2 and RAD51.

5.3.3 Potential mechanisms by which RAD51 promotes CENP-A deposition

Given the many cellular functions mediated by RAD51, there are several potential mechanisms by which RAD51 promote CENP-A deposition. One immediate question is whether the role of RAD51 in CENP-A deposition is related their repair functionality during homologous recombination (**Figure 5.9**). If HR is involved, it is feasible that the direct repair of SSBs, or the formation of recombination intermediates bound by HJURP, serve as links between RAD51 and CENP-A deposition. Alternatively, HR-independent pathways could include DNA replication (i.e., related to the replication fork stabilization roles of BRCA2 and RAD51), DNA transcription, RNA-DNA hybrids, or even a physical association between HR factors and CENP-A. Given the involvement of both BRCA2 and RAD51 and the similarity of their phenotypic expression, it would seem likely that the mechanism in question relates to BRCA2-RAD51 complex functionality, and likely HR. However, the current data do not suffice to directly address this possibility and both HR-dependent and HR-independent potential mechanism are therefore considered.

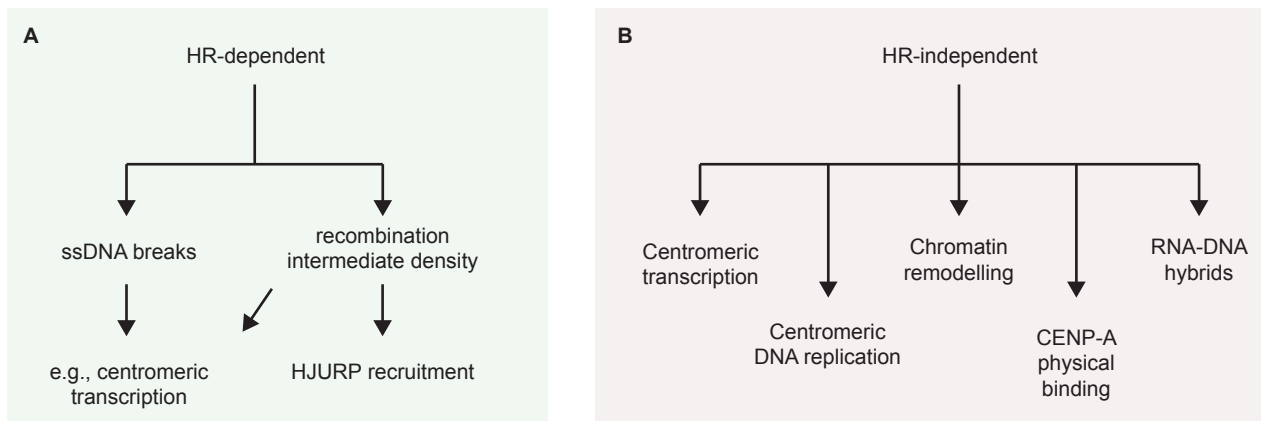


Figure 5.9: **Potential mechanisms by which RAD51 promotes CENP-A deposition.** CENP-A deposition defects in the absence of RAD51 can be explained by several different possibilities, loosely categorized into repair-related or non-repair-related pathways. Homologous recombination (HR)-dependent pathways include single-strand DNA break (SSB) repair or the generation of centromeric recombination intermediates. Pathways that are HR-independent including centromeric transcription and DNA replication, as well as physical binding to CENP-A and accumulated RNA-DNA hybrids.

5.3.3.1 HR-dependent pathways linking RAD51 and CENP-A deposition

In previous chapters, I provide evidence that RAD51 facilitates DNA repair at centromeres, raising a potential link between centromeric DNA repair and CENP-A centromeric occupancy. There are at least two potential mechanisms by which HR as a repair mechanism could promote CENP-A deposition (**Figure 5.9A**).

First, defects in HR may lead to an accumulation of unrepaired DNA breaks at centromeres, indirectly interfering with CENP-A deposition. While this possibility accounts for RAD51-related phenotypes, a notable counterargument is that long-term depletion of BRCA2 does not appear to induce centromeric DNA breaks, as detected by exo-FISH (**Figure 4.5C, D**). This is in direct contrast with the observation that BRCA2 depletion consistently reduces CENP-A centromeric occupancy, as well as CENP-A deposition, to a similar degree as RAD51 (**Figure 5.8**). However, given the inconsistent exo-FISH results upon BRCA2 depletion, this possibility cannot be entirely excluded. Alternatively, it is equally plausible that BRCA2 prevents ssDNA gap accumulation independent of HR-mediated SSB repair. Supporting a role for BRCA2 in preventing ssDNA gap accumulation, several recent reports have argued that the basis of chemosensitivity in BRCA2-deficient cell lines is ssDNA gap accumulation behind re-primed replication forks, rather than the previously proposed DSB accumulation (Cong *et al.*, 2021; Panzarino *et al.*, 2019; Tagliatela *et al.*, 2017; Tirman *et al.*, 2021). While these reports conclude that RAD51 prevents ssDNA gap accumulation by directly catalysing recombination repair, BRCA2 appears to prevent ssDNA gap accumulation by regulating MRE11 nuclease resection (Piberger *et al.*, 2020; Tirman *et al.*, 2021). Therefore, it is possible that BRCA2 and RAD51 both prevent ssDNA accumulation through different

mechanisms. In this model, BRCA2 depletion results in extended ssDNA gaps through MRE11 resection, thereby interfering with CENP-A deposition, without affecting the number of ssDNA gaps, as detected by *exo*-FISH.

It is challenging to envision all the indirect ways through which CENP-A deposition could be perturbed upon ssDNA gap accumulation, some of which may not have been identified yet, and I discuss only three possibilities here. First, a strong contender based on our current knowledge about CENP-A deposition mechanisms is centromeric transcription. Indeed, unrepaired SSBs can block RNA polymerase progression (Kathe *et al.*, 2004; Zhou and Doetsch, 1993, 1994), and this relationship could explain how DNA repair defects selectively interfere with CENP-A deposition rather than CENP-A chromatin stability. This possibility is supported by observations that the depletion of RAD51 as well as BRCA2, albeit less consistently, reduce centromeric transcript levels (**Figure 5.7**).

Second and third, the inability to perform HR at centromeres could impact the local density of recombination intermediates at centromeres, which could subsequently influence either centromeric transcription or HJURP recruitment. For the former, given that centromeres appear to lack canonical transcription initiation sites, it remains unclear how transcription is initiated in these regions. Recent evidence has emerged that secondary DNA structures such as RNA-DNA hybrids can initiate transcription of long non-coding RNAs (Tan-Wong *et al.*, 2019), opening the possibility that Holliday junctions could similarly expose DNA and initiate transcription. Once again, the reduced centromeric transcript levels observed upon RAD51 and BRCA2 depletion support this possibility (**Figure 5.7**). For the latter, changes to recombination intermediate densities could directly influence HJURP recruitment. This model posits that part of the spatial specification for CENP-A deposition is through HJURP directly binding recombination intermediates (Kato *et al.*, 2007). Indeed, centromeres likely have an elevated frequency of recombination intermediates based on the high frequency of SCEs observed at centromeres compared to the rest of the genome (Jaco *et al.*, 2008). In addition, depletion of the MIS18 complex members Mis18 α and Mis18 β only reduce HJURP centromere recruitment by ~60-70%, implying that there may be alternative mechanisms recruiting HJURP to centromeres that have yet to be identified (Wang *et al.*, 2014). It is therefore feasible that recombination intermediates at centromeres act as a redundant mechanism to direct CENP-A towards regions of the genome with high recombination activity.

5.3.3.2 HR-independent pathways linking RAD51 and CENP-A deposition

Independent of HR activity, there are several potential mechanisms by which RAD51 and/or BRCA2 could promote the deposition of CENP-A at centromeres (**Figure 5.9B**).

First, BRCA2 could theoretically influence centromeric transcription independent of its DNA repair role.

BRCA2 has been known to have transcriptional activation functions through its transactivation domain for several decades (Milner *et al.*, 1997). Although both the mechanism and targets of transactivation are still being investigated, BRCA2 has been shown to interact with transcriptional co-activators such as P/CAF (Fuks *et al.*, 1998). More recently, researchers have shown that BRCA2 directly interacts with RNA Polymerase II to allow transcription elongation by preventing the accumulation of RNA-DNA hybrids at transcriptionally active loci (Shivji *et al.*, 2018). Therefore, one possibility is that BRCA2 promotes centromeric transcription through this functionality, in turn promoting CENP-A deposition. However, one flaw with the latter possibility is that the proposed mechanism involves BRCA2 facilitating the release of RNA Polymerase II from promoter-proximal pausing (PPP) sites. PPP sites are positioned 20-50 nucleotides downstream of transcription start sites and found in most RNA Pol II-transcribed genes (Krumm *et al.*, 1995; Rasmussen and Lis, 1993; Rougvie and Lis, 1988). As centromeres do not contain active promoters and the mechanism of centromeric transcription initiation remains unclear, it is unlikely that this explains the observed disruptions in centromeric transcription or CENP-A deposition upon BRCA2 depletion. Notably, this would also not explain the disruptions in CENP-A centromeric occupancy observed upon RAD51 depletion.

Second, CENP-A deposition defects could be related to incomplete centromeric DNA replication. As discussed more extensively in **Chapter 3**, centromeres are frequently described as ‘difficult-to-replicate’ due to their propensity to form DNA secondary structures (Erliandri *et al.*, 2014; Kasinathan and Henikoff, 2018; Ten Hagen *et al.*, 1990). RAD51, independently of BRCA2, can protect stalled replication forks from degradation and even facilitate the restart of stalled or collapsed replication forks (Petermann *et al.*, 2010; Thangavel *et al.*, 2015; Zellweger *et al.*, 2015). BRCA2 has also been shown to protect stalled replication forks, preventing replication fork collapse (Lomonosov, 2003; Mijic *et al.*, 2017). Therefore, an inability to complete centromeric DNA replication could theoretically impede CENP-A deposition in the subsequent mitosis or early G1. The observation that CENP-A levels are disrupted upon RAD51 depletion in non-dividing, quiescent hTERT-RPE1 cells argues against a DNA replication hypothesis (**Figure 5.2**). However, it should be noted that this was only observed in hTERT-RPE1 cells, in which CENP-A protein and mRNA levels are also disrupted upon RAD51 depletion (**Figure 5.3**). Therefore, the possible contribution of incomplete DNA replication to CENP-A deposition defects in the absence of RAD51 cannot be entirely excluded.

Third, both RAD51 and BRCA2 have proposed roles in chromatin remodelling. RAD51 can form nucleoprotein filaments on double-stranded DNA to promote DNA unwrapping to disassemble nucleosomes (Dupaigne *et al.*, 2008; North *et al.*, 2013). In addition, BRCA2 physically interacts with several chromatin remodelling factors including P/CAF and EMSY (Fuks *et al.*, 1998; Hughes-Davies *et al.*, 2003). Indeed, transient depletion of BRCA2 results in large chromatin state changes associated with histone

acetylation (Gruber *et al.*, 2019). In this model, BRCA2 and RAD51-mediated chromatin remodelling at centromeres may contribute to a permissive chromatin environment for CENP-A deposition, although the precise chromatin landscape required has not yet been determined.

Fourth, a physical interaction between RAD51 and CENP-A could theoretically explain the observed inability to deposit new CENP-A at centromeres. Previously, it has been found that *C. albicans* Rad51 and Rad52 physically associate with CENP-A (Cse4), and Rad51-Rad52 was subsequently proposed to directly recruit CENP-A (Cse4) to sites of DNA replication fork stalling for deposition. Indeed, initial observations of reciprocal protein level disruptions between RAD51 and CENP-A supported a notion that they may physically interact (**Figure 5.3**). However, this protein level disruption is reflected by upstream transcription disruptions, and not observed to the same extent in p53-negative cell lines (**Figure 5.4**). Despite this, there was still ~20% reduction in proteins levels upon RAD51 depletion, when normalizing to the input control, even in p53-negative cell lines. Further investigation will therefore be required to determine whether a RAD51-CENP-A interaction occurs in humans, as has been observed in *C. albicans*.

Finally, recent studies have revealed a potential role for RNA-DNA hybrids as a mediator of CENP-A reduction in the absence of HR factors. BRCA1 was shown to maintain CENP-A levels at centromeres by preventing the accumulation of RNA-DNA hybrids (Racca *et al.*, 2021). Since accumulated RNA-DNA hybrids have been shown to de-repress centromeric transcription, which can interfere with CENP-A deposition, it was proposed that centromeric RNA-DNA hybrids may negatively impact CENP-A centromere occupancy (Bergmann *et al.*, 2011; Giunta *et al.*, 2021). Given that BRCA2 has been shown to have similar roles in preventing the accumulation of RNA-DNA hybrids (D'Alessandro *et al.*, 2018; Shivji *et al.*, 2018), it is possible that RNA-DNA hybrids accumulate at centromeres in the absence of BRCA2. However, the observed disruption, rather than de-repression, of centromeric transcription in the absence of BRCA2 does not support this hypothesis (**Figure 5.7**). In addition, the RNA-DNA hybrids found to accumulate in the absence of BRCA2 are due to PPP stalling which are not endogenously found at human centromeres (Shivji *et al.*, 2018). However, once again this possibility also does not explain the phenotypes observed upon RAD51 depletion.

5.3.4 Physiological significance of CENP-A reduction in the absence of RAD51-BRCA2

Regardless of the potential mechanisms underlying this relationship, the reduction of CENP-A upon RAD51-BRCA2 depletion still holds physiological significance. For example, BRCA2- and BRCA1-defective cell lines exhibit abnormal karyotypes harbouring both structural rearrangements as a function of defective DNA repair, as well as numerical chromosome changes (Grigorova *et al.*, 2004). The mech-

anisms resulting in numerical chromosome changes in BRCA2-deficient cells, usually an indication of chromosome mis-segregation, are poorly understood. However, the reduction of CENP-A upon BRCA2 depletion could theoretically explain such chromosome segregation defects. The connection between CENP-A reduction and whole-chromosome mis-segregation in the absence of RAD51-BRCA2 therefore warrants further investigation in the understanding of human health and disease.

The data presented in this chapter remain largely descriptive, and further work will be required to fully deduce the mechanistic basis of these observations. The results presented here will hopefully guide future work to elucidate how RAD51 promotes CENP-A deposition and, in turn, the faithful inheritance of centromere specification.

Chapter 6

Discussion

6.1 Thesis summary and future perspectives

In recent years, attention has increasingly focused on centromeres and pericentromeres as potential sources of genome instability (Barra and Fachinetti, 2018; Black and Giunta, 2018). This has led to the proposal that (peri)centromeres, either despite or because of their physiologically essential function, are intrinsically prone to DNA breaks and rearrangements. However, we are only just beginning to investigate DNA fragility originating at (peri)centromeres and the subsequent implications for human disease.

In this work, I explore the inherent DNA fragility proposed to occur at human centromeres. First, I describe and investigate a striking enrichment of spontaneous DNA strand breaks at human centromeres, which appear to be predominantly single-stranded, and potential suppression of canonical DNA damage signalling. Investigating the role of the repair pathway homologous recombination (HR), I find that the mammalian recombinase RAD51 catalyses strand exchange at human centromeres in unperturbed conditions, providing the first mechanistic insight into pathways driving frequent centromeric recombination. Finally, I uncover a new and unanticipated role for RAD51 in promoting centromere specification through the deposition of the centromere-specific histone CENP-A at centromeres, adding to a growing body of evidence linking DNA repair factors and centromere specification. Together, these findings lead to a model wherein innate DNA fragility at centromeres drives frequent RAD51-mediated recombination, further enabling a chromatin environment conducive to CENP-A deposition (**Figure 6.1**). While the mechanistic links between recombination activity and CENP-A occupancy remain to be elucidated, preliminary evidence suggests that local recombination enables centromeric transcription, thereby directly promoting CENP-A deposition. As discussed further below, these observations lead to several outstanding questions in our pursuit of understanding the structure and stability of the human genome.

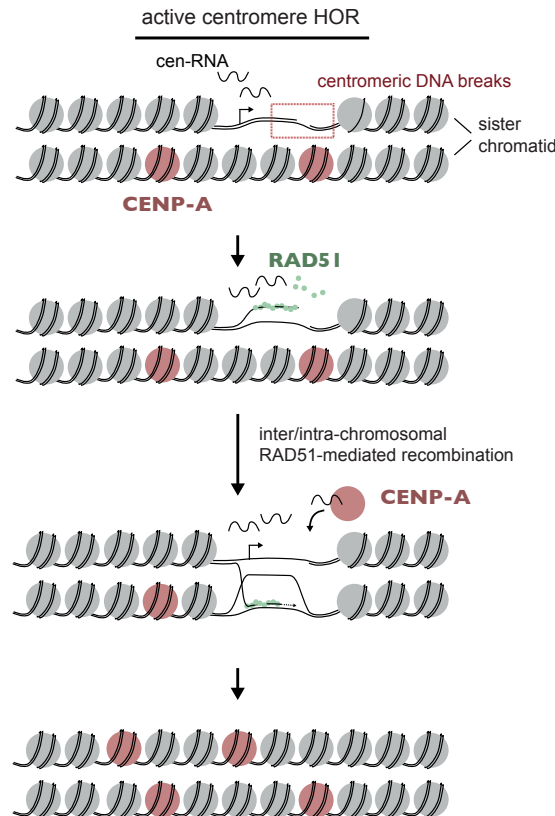


Figure 6.1: **Model for RAD51-mediated recombination at spontaneous centromeric DNA breaks.** An accumulation of DNA strand breaks at human HOR arrays in unperturbed conditions drives RAD51-mediated recombination between inter/intra-chromosomal satellite repeats. CENP-A localization at centromeres is promoted by the recombination capacity (i.e., presence of RAD51) of cells, supporting the notion that CENP-A positioning may be reinforced by active break-and-repair cycles at human centromeres.

6.1.1 Spontaneous DNA fragility at human centromeres

Centromeres and flanking pericentromeres have been proposed to act as genome instability hotspots (section 1.3.5), a notion primarily based on (peri)centromere-derived rearrangements frequently observed in human tumors. However, it has remained unclear whether (peri)centromere-derived rearrangements are an indirect outcome of the highly mutagenic cancer genome environment, or whether (peri)centromere-derived DNA fragility can drive genome instability. In **Chapter 3**, I provide evidence that chromosomal arm copy-number alterations (arm-CNAs), a direct outcome of (peri)centromeric DNA breaks, are the most common mutation class driving early carcinogenesis. Notably, this implies that (peri)centromeres are highly prone to DNA breaks even in non-cancerous cells and with significant ramifications for human health, highlighting the importance of understanding (peri)centromeric DNA fragility.

Supporting the notion of (peri)centromere fragility, analyses of publicly available NGS datasets reveal a striking enrichment of spontaneous DNA breaks across inner centromere cores as well human satel-

lite repeats found in pericentromeric regions. These breaks appear to be predominantly single-stranded. Therefore, I propose that SSBs incurred at centromeres, of which a variable proportion are converted to DSBs, underlie innate centromere fragility and frequent rearrangements implicated in carcinogenesis. The observation that DNA breaks are enriched within both centromeric cores as well as pericentromeric-associated satellite repeats suggests that both centromeres and pericentromeres can act as potential sources of genome instability. These results therefore shed light onto the ambiguity regarding the respective roles of centromeres and pericentromeres in driving genome instability (**section 3.3.1.3**).

An important next step will be identifying the source of centromeric DNA breaks for two principal reasons. First, to interrogate the relationship between spontaneous DNA breaks and downstream structural rearrangements originating at centromeres (e.g., arm-CNAs). Indeed, while it seems plausible that centromeric DNA breaks drive the innate fragility reflected in arm-CNA formation, supporting experimental evidence will be required. Second, to identify whether centromeric DNA breaks are an indirect outcome of some physiological process, or whether they serve a cellular function. For example, centromeric DNA single-stranded breaks may reflect replication or repair intermediates, but it is also feasible that they are directed at centromeres to resolve topological stress or even act in some signaling capacity (a notion explored in **Chapter 5**). Identifying and subsequently modulating the source of centromeric DNA breaks could help address these critical questions.

In **Chapter 3**, I also develop and validate *exo-FISH*, a new technique with which to detect DNA breaks at repetitive regions of the genome. *exo-FISH* is subsequently applied to further investigate the source of DNA fragility at human centromeres. Although the source of DNA breaks could not be identified, this work reveals that spontaneous centromeric DNA breaks are largely independent of active cellular proliferation and CENP-A chromatin occupancy. It seems probable that some feature of centromeric DNA renders centromeres inherently susceptible to SSBs. As further explored in **section 3.3.1.4**, likely candidates include structure-specific endonucleases such as DNA2 or Topoisomerase II α , transcription, DNA repair intermediates, or even centromeric proteins with domesticated transposons (i.e., CENP-B). The development of *exo-FISH* as a high-throughput and direct single-cell DNA break detection method will hopefully enable future studies into the source of spontaneous centromere fragility.

6.1.2 DNA damage signalling at human centromeres

Genome stability is globally maintained through efficient DNA damage recognition, signaling and repair pathways, collectively termed the DNA damage response (DDR). Surprisingly, the work done in **Chapter 3** led to the observation that canonical DDR signals, more specifically γ H2A.X and phosphorylated RPA, are unresponsive to centromeric DNA breaks. Indeed, both spontaneous SSBs as well as induced DSBs (i.e., replication stress treatment or Topoisomerase II α poisoning) are insufficient to elicit detectable

H2A.X or RPA phosphorylation at centromeres. While these observations will require more comprehensive validation, the observed DDR suppression is notable for two reasons. First, DDR suppression at centromeres could explain why such a striking innate fragility at human centromeres has not previously been reported, as DNA instability is frequently assessed using markers of DDR activation. Second, these data raise important questions in understanding centromere-derived genome instability. Indeed, it is tempting to speculate that arm-CNAs are so frequently observed in early carcinogenesis partly because centromeric DNA breaks are not adequately signalled for efficient repair.

Potential mechanisms for DNA damage response suppression are further explored in **section 3.3.2.2**, and it will be important to both verify and further investigate these possibilities in future studies. For example, one pressing question is whether upstream DDR factors (e.g., MRN complex or PARP1) are recruited to centromeres upon DNA break induction, to determine whether this suppression is universal to all DDR factors. However, regardless of the mechanism, such a phenomenon could provide significant insight into understanding centromere-derived genome instability.

The notion of DDR suppression is not entirely novel, as telomeres have similarly been shown to evade canonical DDR recognition (Carneiro *et al.*, 2010), but it remains unclear why centromeres should be refractory to DNA break detection. One possibility is that centromere functionality itself is inherently harmful to DNA integrity, and so centromeres require a higher DNA break tolerance to allow continued cellular proliferation. Alternatively, it is plausible that repetitive regions have as-of-yet uncharacterized DNA damage signalling mechanisms. Indeed, DNA repair at repetitive regions is prone to mishaps due to the abundance of repair templates and may require specialized mechanisms to prevent excessive chromosomal rearrangements. Finally, due to the high abundance of repair templates, DNA damage signalling may not be necessarily required for efficient repair, and so centromeric DNA breaks could avoid signalling by the canonical DDR to prevent cell toxicity induced by excessive DDR signalling.

6.1.3 RAD51-driven recombination in the repair of centromeric DNA breaks

Given the innate centromere fragility and DDR suppression described in **Chapter 3**, a large portion of this work is dedicated to investigating how centromeric DNA breaks are repaired in human cells. More specifically, this thesis explores the role of the repair pathway homologous recombination (HR) in the repair of centromeric breaks, as previous works have observed high rates of sister chromatid exchanges (SCEs), one of two major HR outcomes, at mammalian centromeres (Giunta and Funabiki, 2017; Jaco *et al.*, 2008). To this end, I investigate the role of the central mammalian recombinase RAD51, which catalyzes HR in mammalian cells, at human centromeres (Baumann *et al.*, 1996).

Through exo-FISH analyses in non-proliferating hTERT-RPE1 cells, it is evident that RAD51 actively

repairs centromeric DNA breaks in a recombination-dependent manner (i.e., requiring RAD51 strand exchange functionality). Notably, specificity for centromeres is demonstrated as other repetitive regions of the genomes (i.e., telomeres), are not similarly targeted by RAD51. The specificity of RAD51 is further supported by a localization study that reveal a consistent enrichment of RAD51 at artificial centromeric alpha satellites, although curiously not at endogenous centromeres. One compelling hypothesis is that the centromeric DNA breaks identified in **Chapter 3** drive RAD51-mediated recombination, explaining the striking rates of basal SCEs observed at human centromeres in the absence of exogenous perturbation. While some ambiguity remains as to whether SSBs or DSBs specifically drive RAD51-mediated recombination, both forms of lesions can theoretically initiate HR (**section 1.1.1.4**). To my knowledge, this is the first experimental evidence supporting the role of canonical HR factors in mammalian centromeric recombination, highlighting a previously unappreciated role for HR in genome stability.

Curiously, RAD51 recruitment to artificial centromeres appears to be largely independent of its canonical recruiter and regulator BRCA2. BRCA2 also does not appear to be important for the prevention of DNA breaks at centromeres, suggesting that RAD51 acts independently of the canonical HR mediator complex during DNA break repair at centromeres.

Notably, this work does not systematically address all repair pathways which may contribute to the protection of centromeric DNA, which undoubtedly requires further investigation. However, the observed role for HR in repairing centromeric breaks is still an important breakthrough on at least two fronts. First, as centromeric breaks can act as substrates of HR-mediated repair, this provides a compelling mechanism by which arm-CNAs are so frequently observed. Indeed, non-allelic HR (NAHR) between satellite repeats could feasibly drive arm-CNAs observed in early carcinogenesis. Second, this observation also holds significant implications for our understanding of centromeric satellite DNA evolution. Many decades of evolutionary modelling of centromeric arrays have been based on the assumption of frequent expansions, contractions, and homogenization of satellite arrays through genetic exchanges and unequal crossovers. While some studies have identified high gene exchange frequencies at centromeres, the mechanisms driving these rearrangements have remained largely unexplored. The identification of RAD51-driven centromeric recombination in generally physiological (i.e., unperturbed) conditions provides some mechanistic insight into why and how centromeres are among the most rapidly evolving regions of the genome (**Figure 6.2**).

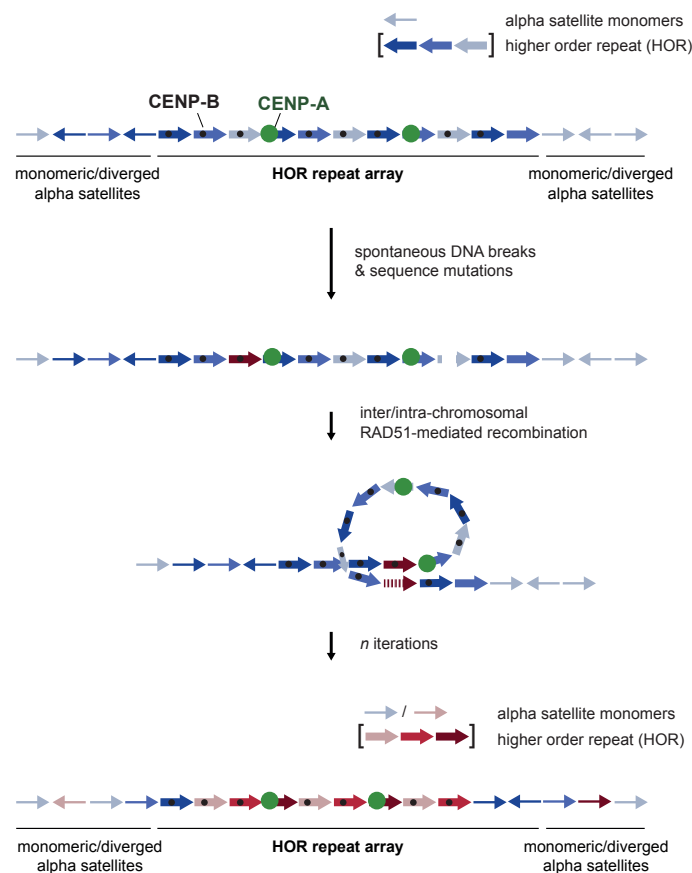


Figure 6.2: **Model for RAD51-mediated recombination in the evolution of centromeric satellite arrays.** Active centromeres contain alpha satellites repeats organized into large repeat units known as higher-order repeats (HORs), with more divergent and monomeric alpha satellites towards centromere peripheries. Spontaneous DNA strand breaks promote frequent recombination between centromeric alpha satellite arrays, either inter- or intra-chromosomally, such that random sequence mutations are homogenized efficiently, resulting the rapid evolution observed with centromeric repeat arrays.

6.1.4 Proposed relationship between DNA repair factors and centromere specification

The final chapter in this thesis explores the relationship between DNA repair factors and centromere specification. Centromeres exhibit remarkable spatial plasticity, as they are primarily defined by the presence of the centromere-specific histone variant CENP-A. However, the processes facilitating permissive chromatin environments for CENP-A positioning remain to be fully identified.

In this work, I show that the central recombinase RAD51 as well its canonical regulator BRCA2 promote CENP-A occupancy at centromeres by enabling new CENP-A deposition at centromeres. An important caveat remains regarding the specificity of RAD51-BRCA2 in CENP-A loading ([section 5.3.2](#)). Indeed,

it is unclear whether RAD51-BRCA2 promotes the deposition of nucleosomes genome-wide, or specifically CENP-A-containing nucleosomes at centromeres. To my knowledge, the former possibility has not previously been observed, but warrants consideration. Canonical SNAP-tagged H3 histones should be used as a direct control in future experiments to determine whether there are any defects in H3 chromatin loading at CENP-B-defined centromeres in the absence of RAD51-BRCA2.

As thoroughly discussed in **section 5.3.3**, there are several potential mechanisms by which RAD51 could promote CENP-A deposition. An immediate question is whether CENP-A deposition is dependent on the recombination role of RAD51 or some HR-independent function. This question could be addressed by making use of the RAD51 separation-of-function variants generated and validated in **Chapter 4**, to determine whether RAD51 strand exchange activity is required for promoting CENP-A chromatin occupancy. If recombination activity is indeed involved, the link between HR and CENP-A deposition could be indirect, through the disruption of centromeric transcription upon DNA break accumulation for example, or direct, through recombination intermediates (**Figure 5.9**). An intriguing hypothesis is that centromeric recombination induced by spontaneous centromeric DNA breaks helps initiate centromeric transcription, thereby enabling CENP-A deposition to highly recombinogenic alpha satellite repeats (**Figure 6.1**).

If recombination intermediates are required for CENP-A deposition, there are at least two highly intriguing implications concerning the link between DNA damage repair and centromere specification. First, such a link might imply that part of the centromeric specification paradigm is an innate propensity for centromeres to undergo DNA damage and repair, notions explored in **Chapters 3** and **Chapter 4**. For example, if centromere activity were inherently damaging to DNA, a feedback loop would be created ensuring that centromere-induced DNA damage only occurs within regions already susceptible to DNA damage, protecting more regions of the genome where sequence fidelity is more important (i.e., protein-coding loci). It would be, to my knowledge, one of the first instances in which cells have evolved to capitalize on localized DNA fragility in such a significant sense. Second, already established centromere specification mechanisms (as described in **section 1.3.4**) are primarily based on pre-existing CENP-A levels, such that CENP-A levels can only be maintained at equivalence. It is therefore intriguing to speculate whether under certain conditions, plasticity in centromere ‘strength’ might enable adaptation. For example, situations in which stronger microtubule-chromatin associations are required because sister chromatid are more physically intertwined might necessitate more CENP-A molecules per centromere, thereby initiating larger kinetochores and increased microtubule attachment. Intermediates of HR, as one of the mechanisms likely intertwining sister chromatid, might be a useful mechanism by which cells sense whether the levels of CENP-A deposition should be modulated.

6.2 Concluding remarks

Centromeres, having been identified as early as 1882 (Flemming, 1882), still remain largely enigmatic today. As we move forward in understanding these complex regions of the genome, we uncover new and surprising insights into genome biology. This thesis provides evidence that centromeres harbor a striking and previously unrecognized DNA fragility, the purpose for which remains unknown. However, it is evident that centromeric DNA breaks can be targeted by the recombination machinery, providing mechanistic evidence for the recombinational activities thought to occur at centromeres for several decades. Emerging evidence also opens the possibility that this intrinsic fragility is somehow implicated in a signaling capacity, given the connection between DNA break repair and centromere specification described in **Chapter 5**. Overall, this work will hopefully guide new avenues of research to further our understanding of the structure and stability of the human genome.

Bibliography

- Adeyinka, A., Wei, S., and Sanchez, J. (2007). Loss of 17p is a major consequence of whole-arm chromosome translocations in hematologic malignancies. *Cancer Genetics and Cytogenetics* 173, 136-143.
- Ahmad, K., and Henikoff, S. (2001). Centromeres are specialized replication domains in heterochromatin. *Journal Cell Biology* 153, 101-110.
- Alexandrov, I., Alexei, K., Irina, T., Valery, S., and Yurov, Y. (2001). Alpha-satellite DNA of primates: old and new families. *Chromosoma* 110, 253-266.
- Alsop, A.E., Teschendorff, A., and Edwards, P.A. (2006). Distribution of breakpoints on chromosome 18 in breast, colorectal, and pancreatic carcinoma cell lines. *Cancer Genetics Cytogenetics* 164, 97-109.
- Altemose, N., Logsdon, G.A., Bzikadze, A.V., Sidhwani, P., Langley, S.A., Caldas, G.V., Hoyt, S.J., Uralsky, L., Ryabov, F.D., Shew, C.J., *et al.* (2021). Complete genomic and epigenetic maps of human centromeres. *bioRxiv*.
- Amitai, A., Seeber, A., Gasser, S.M., and Holcman, D. (2017). Visualization of Chromatin Decompaction and Break Site Extrusion as Predicted by Statistical Polymer Modeling of Single-Locus Trajectories. *Cell Reports* 18, 1200-1214.
- Arias-Lopez, C., Lazaro-Trueba, I., Kerr, P., Lord, C.J., Dexter, T., Iravani, M., Ashworth, A., and Silva, A. (2006). p53 modulates homologous recombination by transcriptional regulation of the RAD51 gene. *EMBO Reports* 7, 219-224.
- Arinkin, V., Smyshlyaev, G., and Barabas, O. (2019). Jump ahead with a twist: DNA acrobatics drive transposition forward. *Current Opinion in Structural Biology* 59, 168-177.
- Arnould, C., Rocher, V., Clouaire, T., Caron, P., Mangeot, P.E., Ricci, E.P., Mourad, R., Noordermeer, D., and Legube, G. (2020). Loop extrusion as a mechanism for DNA Double-Strand Breaks repair foci formation. *bioRxiv*
- Aze, A., Sannino, V., Soffientini, P., Bachi, A., and Costanzo, V. (2016). Centromeric DNA replication reconstitution reveals DNA loops and ATR checkpoint suppression. *Nature Cell Biology* 18, 684-691.
- Backsch, C., Pauly, B., Liesenfeld, M., Scheungraber, C., Gajda, M., Mrasek, K., Liehr, T., Clad, A., Schrock, E., Runnebaum, I.B., *et al.* (2011). Two novel unbalanced whole arm translocations are frequently detected in cervical squamous cell carcinoma. *Cancer Genetics* 204, 646-653.
- Bacolla, A., Wojciechowska, M., Kosmider, B., Larson, J.E., and Wells, R.D. (2006). The involvement of non-B DNA structures in gross chromosomal rearrangements. *DNA Repair* 5, 1161-1170.
- Bailey, S.M., Goodwin, E.H., and Cornforth, M.N. (2004). Strand-specific fluorescence in situ hybridization: the CO-FISH family. *Cytogenetic and Genome Research* 107, 14-17.
- Balachandar, V., Lakshman Kumar, B., Sasikala, K., Manikantan, P., Sangeetha, R., and Mohana Devi, S. (2007). Identification of a high frequency of chromosomal rearrangements in the centromeric regions of prostate cancer patients. *Journal of Zhejiang University Science B* 8, 638-646.
- Baltimore, D. (1981). Gene conversion: some implications for immunoglobulin genes. *Cell* 24, 592-594.
- Balzano, E., and Giunta, S. (2020). Centromeres under Pressure: Evolutionary Innovation in Conflict with Conserved Function. *Genes (Basel)* 11, 912.
- Barbouti, A., Stankiewicz, P., Nusbaum, C., Cuomo, C., Cook, A., Höglund, M., Johansson, B., Hagemeyer, A., Park, S.-S., Mitelman, F., *et al.* (2004). The Breakpoint Region of the Most Common Isochromosome, i(17q), in Human Neoplasia Is Characterized by a Complex Genomic Architecture with Large, Palindromic, Low-Copy Repeats. *American Journal of Human Genetics* 74, 1-10.
- Barnhart, M.C., Kuich, P.H.J.L., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *Journal of Cell Biology* 194, 229-243.

- Barra, V., and Fachinetti, D. (2018). The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. *Nature Communications* 9, 4340.
- Bassing, C.H., and Alt, F.W. (2004). The cellular response to general and programmed DNA double strand breaks. *DNA Repair (Amst)* 3, 781-796.
- Baumann, P., Benson, F.E., and West, S.C. (1996). Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions *in vitro*. *Cell* 87, 757-766.
- Beadle, G.W. (1932). A Possible Influence of the Spindle Fibre on Crossing-Over in *Drosophila*. *Proceedings of the National Academy of Sciences* 18, 160.
- Beishline, K., Vladimirova, O., Tutton, S., Wang, Z., Deng, Z., and Lieberman, P.M. (2017). CTCF driven TERRA transcription facilitates completion of telomere DNA replication. *Nature Communications* 8, 2114.
- Benson, F.E., Stasiak, A., and West, S.C. (1994). Purification and characterization of the human Rad51 protein, an analogue of *E. coli* RecA. *EMBO Journal* 13, 5764-5771.
- Bergmann, J.H., Rodríguez, M.G., Martins, N.M.C., Kimura, H., Kelly, D.A., Masumoto, H., Larionov, V., Jansen, L.E.T., and Earnshaw, W.C. (2011). Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *EMBO Journal* 30, 328-340.
- Bernard, P., Maure, J.F., Partridge, J.F., Genier, S., Javerzat, J.P., and Allshire, R.C. (2001). Requirement of heterochromatin for cohesion at centromeres. *Science* 294, 2439-2442.
- Beroukhi, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., *et al.* (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899-905.
- Bhogal, N., Jalali, F., and Bristow, R.G. (2009). Microscopic imaging of DNA repair foci in irradiated normal tissues. *International Journal of Radiation Biology* 85, 732-746.
- Bi, X. (2015). Mechanism of DNA damage tolerance. *World Journal of Biological Chemistry* 6, 48-56.
- Bièche, I., Khodja, A., Driouch, K., and Lidereau, R. (1997). Genetic alteration mapping on chromosome 7 in primary breast cancer. *Clinical Cancer Research* 3, 1009.
- Black, E.M., and Giunta, S. (2018). Repetitive Fragile Sites: Centromere Satellite DNA As a Source of Genome Instability in Human Diseases. *Genes (Basel)* 9, 615.
- Blat, Y., and Kleckner, N. (1999). Cohesins Bind to Preferential Sites along Yeast Chromosome III, with Differential Regulation along Arms versus the Centric Region. *Cell* 98, 249-259.
- Bloom, K.S., and Carbon, J. (1982). Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. *Cell* 29, 305-317.
- Blower, M.D. (2016). Centromeric Transcription Regulates Aurora-B Localization and Activation. *Cell Reports* 15, 1624-1633.
- Bobkov, G.O.M., Gilbert, N., and Heun, P. (2018). Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. *Journal of Cell Biology*. 217, 1957-1972.
- Bobkov, G.O.M., Huang, A., van den Berg, S.J.W., Mitra, S., Anselm, E., Lazou, V., Schunter, S., Feederle, R., Imhof, A., Lusser, A., *et al.* (2020). Spt6 is a maintenance factor for centromeric CENP-A. *Nature Communications* 11, 2919.
- Bodor, D.L., Mata, J.F., Sergeev, M., David, A.F., Salimian, K.J., Panchenko, T., Cleveland, D.W., Black, B.E., Shah, J.V., and Jansen, L.E. (2014). The quantitative architecture of centromeric chromatin. *Elife* 3, e02137.
- Bodor, D.L., Rodriguez, M.G., Moreno, N., and Jansen, L.E. (2012). Analysis of protein turnover by quantitative SNAP-based pulse-chase imaging. *Current Protocols in Cell Biology* Chapter 8, Unit 8.8.
- Bodor, D.L., Valente, L.P., Mata, J.F., Black, B.E., and Jansen, L.E. (2013). Assembly in G1 phase and long-term stability are unique intrinsic features of CENP-A nucleosomes. *Molecular Biology of the Cell* 24, 923-932.

- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.
- Bouwman, B.A.M., and Crosetto, N. (2018). Endogenous DNA Double-Strand Breaks during DNA Transactions: Emerging Insights and Methods for Genome-Wide Profiling. *Genes* 9, 632.
- Boveri T, F.F. (1914). Zur Frage der Entstehung maligner Tumoren. *Science* 40, 857-859.
- Brandsma, I., and Gent, D.C. (2012). Pathway choice in DNA double strand break repair: observations of a balancing act. *Genome Integrity* 3, 9.
- Budd, M.E., Choe, W.C., and Campbell, J.L. (1995). DNA2 encodes a DNA helicase essential for replication of eukaryotic chromosomes. *Journal of Biological Chemistry* 270, 26766-26769.
- Burma, S., Chen, B.P., Murphy, M., Kurimasa, A., and Chen, D.J. (2001). ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *Journal of Biological Chemistry* 276, 42462-42467.
- Bury, L., Moodie, B., Ly, J., McKay, L.S., Miga, K.H., and Cheeseman, I.M. (2020). Alpha-satellite RNA transcripts are repressed by centromere-nucleolus associations. *Elife* 9.
- Byrne, B.M., and Oakley, G.G. (2019). Replication protein A, the laxative that keeps DNA regular: The importance of RPA phosphorylation in maintaining genome stability. *Seminars in Cell and Developmental Biology* 86, 112-120.
- Caldecott, K.W. (2008). Single-strand break repair and genetic disease. *Nature Reviews Genetics* 9, 619-631.
- Canela, A., Maman, Y., Huang, S.-Y.N., Wutz, G., Tang, W., Zagnoli-Vieira, G., Callen, E., Wong, N., Day, A., Peters, J.-M., *et al.* (2019). Topoisomerase II-Induced Chromosome Breakage and Translocation Is Determined by Chromosome Architecture and Transcriptional Activity. *Molecular Cell* 75, 252-266.e258.
- Canela, A., Maman, Y., Jung, S., Wong, N., Callen, E., Day, A., Kieffer-Kwon, K.-R., Pekowska, A., Zhang, H., Rao, S.S.P., *et al.* (2017). Genome Organization Drives Chromosome Fragility. *Cell* 170, 507-521.e518.
- Carneiro, T., Khair, L., Reis, C.C., Borges, V., Moser, B.A., Nakamura, T.M., and Ferreira, M.G. (2010). Telomeres avoid end detection by severing the checkpoint signal transduction pathway. *Nature* 467, 228-232.
- Caron, P., Aymard, F., Iacovoni, J.S., Briois, S., Canitrot, Y., Bugler, B., Massip, L., Losada, A., and Legube, G. (2012). Cohesin protects genes against γ H2AX Induced by DNA double-strand breaks. *PloS Genetics* 8, e1002460.
- Carroll, C.W., Milks, K.J., and Straight, A.F. (2010). Dual recognition of CENP-A nucleosomes is required for centromere assembly. *Journal of Cell Biology* 189, 1143-1155.
- Chan, F.L., Marshall, O.J., Saffery, R., Won Kim, B., Earle, E., Choo, K.H.A., and Wong, L.H. (2012). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. *Proceedings of the National Academy of Sciences* 109, 1979-1984.
- Chappell, W.H., Gautam, D., Ok, S.T., Johnson, B.A., Anacker, D.C., and Moody, C.A. (2015). Homologous Recombination Repair Factors Rad51 and BRCA1 Are Necessary for Productive Replication of Human Papillomavirus 31. *Journal of Virology* 90, 2639-2652.
- Chardon, F., Japaridze, A., Witt, H., Velikovskiy, L., Chakraborty, C., Wilhelm, T., Dumont, M., Yang, W., Kikuti, C., Gangnard, S., *et al.* (2022). CENP-B-mediated DNA loops regulate activity and stability of human centromeres. *Molecular Cell*.
- Chastain, M., Zhou, Q., Shiva, O., Fadri-Moskwik, M., Whitmore, L., Jia, P., Dai, X., Huang, C., Ye, P., and Chai, W. (2016). Human CST Facilitates Genome-wide RAD51 Recruitment to GC-Rich Repetitive Sequences in Response to Replication Stress. *Cell Reports* 16, 2048.
- Chen, C.C., Bowers, S., Lipinski, Z., Palladino, J., Trusiak, S., Bettini, E., Rosin, L., Przewłoka, M.R., Glover, D.M., O'Neill, R.J., *et al.* (2015). Establishment of Centromeric Chromatin by the CENP-A Assembly Factor CAL1 Requires FACT-Mediated Transcription. *Developmental Cell* 34, 73-84.
- Chi, P., Van Komen, S., Sehorn, M.G., Sigurdsson, S., and Sung, P. (2006). Roles of ATP binding and ATP hydrolysis in human Rad51 recombinase function. *DNA Repair (Amsterdam)* 5, 381-391.

- Chintapalli, S.V., Bhardwaj, G., Babu, J., Hadjiyianni, L., Hong, Y., Todd, G.K., Boosalis, C.A., Zhang, Z., Zhou, X., Ma, H., *et al.* (2013). Reevaluation of the evolutionary events within recA/RAD51 phylogeny. *BMC Genomics* 14, 240.
- Clark, A.J. (1973). Recombination deficient mutants of *E. coli* and other bacteria. *Annual Review of Genetics* 7, 67-86.
- Clark, A.J., and Margulies, A.D. (1965). Isolation and Characterization of Recombination-Deficient Mutants of *Escherichia Coli* K12. *Proceedings of National Academy of Sciences* 53, 451-459.
- Cloud, V., Chan, Y.I., Grubb, J., Budke, B., and Bishop, D.K. (2012). Rad51 is an accessory factor for Dmc1-mediated joint molecule formation during meiosis. *Science* 337, 1222-1225.
- Collins, A.R. (2004). The comet assay for DNA damage and repair: principles, applications, and limitations. *Molecular Biotechnology* 26, 249-261.
- Cong, K., Peng, M., Kousholt, A.N., Lee, W.T.C., Lee, S., Nayak, S., Krais, J., VanderVere-Carozza, P.S., Pawelczak, K.S., Calvo, J., *et al.* (2021). Replication gaps are a key determinant of PARP inhibitor synthetic lethality with BRCA deficiency. *Mol Cell* 81, 3227.
- Cooke, S.L., Pole, J.C.M., Chin, S.F., Ellis, I., Caldas, C., and Edwards, P.A. (2008). High-resolution array CGH clarifies events occurring on 8p in carcinogenesis. *BMC Cancer* 8, 288.
- Creighton, H.B., and McClintock, B. (1931). A Correlation of Cytological and Genetical Crossing-Over in *Zea Mays*. *Proceedings of National Academy of Sciences* 17, 492-497.
- Cristini, A., Ricci, G., Britton, S., Salimbeni, S., Huang, S.-y.N., Marinello, J., Calsou, P., Pommier, Y., Favre, G., Capranico, G., *et al.* (2019). Dual Processing of R-Loops and Topoisomerase I Induces Transcription-Dependent DNA Double-Strand Breaks. *Cell Reports* 28, 3167-3181.e3166.
- Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., *et al.* (2013). Nucleotide-resolution DNA double-strand breaks mapping by next-generation sequencing. *Nature Methods* 10, 361-365.
- D'Alessandro, G., Whelan, D.R., Howard, S.M., Vitelli, V., Renaudin, X., Adamowicz, M., Iannelli, F., Jones-Weinert, C.W., Lee, M., Matti, V., *et al.* (2018). BRCA2 controls DNA:RNA hybrid level at DSBs by mediating RNase H2 recruitment. *Nature Communications* 9, 5376.
- Darai-Ramqvist, E., Sandlund, A., Muller, S., Klein, G., Imreh, S., and Kost-Alimova, M. (2008). Segmental duplications and evolutionary plasticity at tumor chromosome break-prone regions. *Genome Research* 18, 370-379.
- Davis, L., and Maizels, N. (2014). Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proceedings of National Academy of Sciences* 111, 924-932.
- Depinet, T.W., Zackowski, J.L., Earnshaw, W.C., Kaffe, S., Sekhon, G.S., Stallard, R., Sullivan, B.A., Vance, G.H., Van Dyke, D.L., Willard, H.F., *et al.* (1997). Characterization of neo-centromeres in marker chromosomes lacking detectable alpha-satellite DNA. *Human Molecular Genetics* 6, 1195-1204.
- Dion, V., and Gasser, S.M. (2013). Chromatin movement in the maintenance of genome stability. *Cell* 152, 1355-1364.
- Dion, V., Kalck, V., Horigome, C., Towbin, B.D., and Gasser, S.M. (2012). Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery. *Nature Cell Biology* 14, 502-509.
- Dunleavy, E.M., Roche, Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani Y., and Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* 137, 485-497.
- Dupaigne, P., Lavelle, C., Justome, A., Lafosse, S., Mirambeau, G., Lipinski, M., Pietrement, O., and Le Cam, E. (2008). Rad51 polymerization reveals a new chromatin remodeling mechanism. *PLoS One* 3, e3643.
- Earnshaw, W.C., and Rothfield, N. (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* 91, 313-321.
- Eder, P.S., Walder, R.Y., and Walder, J.A. (1993). Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. *Biochimie* 75, 123-126.

- Egel, R. (1981). Intergenic conversion and reiterated genes. *Nature* 290, 191-192.
- Elvers, I., Johansson, F., Groth, P., Erixon, K., and Helleday, T. (2011). UV stalled replication forks restart by re-priming in human fibroblasts. *Nucleic Acids Research* 39, 7049-7057.
- Erliandri, I., Fu, H., Nakano, M., Kim, J.-H., Miga, K.H., Liskovych, M., Earnshaw, W.C., Masumoto, H., Kouprina, N., Aladjem, M.I., *et al.* (2014). Replication of alpha-satellite DNA arrays in endogenous human centromeric regions and in human artificial chromosome. *Nucleic Acids Research* 42, 11502-11516.
- Esashi, F., Christ, N., Gannon, J., Liu, Y., Hunt, T., Jasin, M., and West, S.C. (2005). CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature* 434, 598-604.
- Fachinetti, D., Folco, H.D., Nechemia-Arbely, Y., Valente, L.P., Nguyen, K., Wong, A.J., Zhu, Q., Holland, A.J., Desai, A., Jansen, L.E., *et al.* (2013). A two-step mechanism for epigenetic specification of centromere identity and function. *Nature Cell Biology* 15, 1056-1066.
- Falk, S.J., Guo, L.Y., Sekulic, N., Smoak, E.M., Mani, T., Logsdon, G.A., Gupta, K., Jansen, L.E., Van Duyne, G.D., Vinogradov, S.A., *et al.* (2015). Chromosomes. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science* 348, 699-703.
- Feng, W., Di Rienzi, S.C., Raghuraman, M.K., and Brewer, B.J. (2011). Replication stress-induced chromosome breakage is correlated with replication fork progression and is preceded by single-stranded DNA formation. *G3 (Bethesda)* 1, 327-335.
- Fernandez, J.L., Vazquez-Gundin, F., Rivero, M.T., Genesca, A., Gosalvez, J., and Goyanes, V. (2001). DBD-fish on neutral comets: simultaneous analysis of DNA single- and double-strand breaks in individual cells. *Experimental Cell Research* 270, 102-109.
- Fernandez-Banet, J., Lee, N.P., Chan, K.T., Gao, H., Liu, X., Sung, W.-K., Tan, W., Fan, S.T., Poon, R.T., Li, S., *et al.* (2014). Decoding complex patterns of genomic rearrangement in hepatocellular carcinoma. *Genomics* 103, 189-203.
- Ferri, F., Bouzinba-Segard, H., Velasco, G., Hube, F., and Francastel, C. (2009). Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. *Nucleic Acids Research* 37, 5071-5080.
- Figuroa-Gonzalez, G., and Perez-Plasencia, C. (2017). Strategies for the evaluation of DNA damage and repair mechanisms in cancer. *Oncology Letters* 13, 3982-3988.
- Fitzgerald-Hayes, M., Clarke, L., and Carbon, J. (1982). Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell* 29, 235-244.
- Flemming, W. (1882). *Zellsubstanz, kern und zelltheilung*. Vogel.
- Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell* 137, 472-484.
- Frankenberg-Schwager, M. (1989). Review of repair kinetics for DNA damage induced in eukaryotic cells *in vitro* by ionizing radiation. *Radiotherapy Oncology* 14, 307-320.
- Froelich-Ammon, S.J., Gale, K.C., and Osheroff, N. (1994). Site-specific cleavage of a DNA hairpin by topoisomerase II. DNA secondary structure as a determinant of enzyme recognition/cleavage. *Journal of Biological Chemistry* 269, 7719-7725.
- Fujita, Y., Hayashi, T., Kiyomitsu, T., Toyoda, Y., Kokubu, A., Obuse, C., and Yanagida, M. (2007). Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. *Developmental Cell* 12, 17-30.
- Fuks, F., Milner, J., and Kouzarides, T. (1998). BRCA2 associates with acetyltransferase activity when bound to P/CAF. *Oncogene* 17, 2531-2534.
- Furuyama, S., and Biggins, S. (2007). Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proceedings of National Academy of Sciences* 104, 14706-14711.
- Furuyama, T., Dalal, Y., and Henikoff, S. (2006). Chaperone-mediated assembly of centromeric chromatin *in vitro*. *Proceedings of National Academy of Sciences* 103, 6172-6177.

- Gallego, J., Chou, S.H., and Reid, B.R. (1997). Centromeric pyrimidine strands fold into an intercalated motif by forming a double hairpin with a novel T:G:G:T tetrad: solution structure of the d(TCCCGTTTCCA) dimer. *Journal of Molecular Biology* 273, 840-856.
- Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L., and Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes and Development* 25, 2041-2056.
- Gangloff, S., Zou, H., and Rothstein, R. (1996). Gene conversion plays the major role in controlling the stability of large tandem repeats in yeast. *EMBO Journal* 15, 1715-1725.
- Ganly, P., McDonald, M., Spearing, R., and Morris, C.M. (2004). Constitutional t(5;7)(q11;p15) rearranged to acquire monosomy7q and trisomy 1q in a patient with myelodysplastic syndrome transforming to acute myelocytic leukemia. *Cancer Genetics and Cytogenetics* 149, 125-130.
- Garavís, M., Escaja, N., Gabelica, V., Villasante, A., and González, C. (2015). Centromeric Alpha-Satellite DNA Adopts Dimeric i-Motif Structures Capped by AT Hoogsteen Base Pairs. *Chemistry* 21, 9816-9824.
- Garavís, M., Méndez-Lago, M., Gabelica, V., Whitehead, S.L., González, C., and Villasante, A. (2015). The structure of an endogenous *Drosophila* centromere reveals the prevalence of tandemly repeated sequences able to form i-motifs. *Science Reports* 5, 13307.
- Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* 109, E2579.
- Gavrieli, Y., Sherman, Y., and Ben-Sasson, S.A. (1992). Identification of programmed cell death *in situ* via specific labeling of nuclear DNA fragmentation. *Journal of Cell Biology* 119, 493-501.
- Gerstung, M., Jolly, C., Leshchiner, I., Dentre, S.C., Gonzalez, S., Rosebrock, D., Mitchell, T.J., Rubanova, Y., Anur, P., Yu, K., *et al.* (2020). The evolutionary history of 2,658 cancers. *Nature* 578, 122-128.
- Giunta, S., and Funabiki, H. (2017). Integrity of the human centromere DNA repeats is protected by CENP-A, CENP-C, and CENP-T. *Proceedings of National Academy of Sciences* 114, 1928-1933.
- Giunta, S., Hervé, S., White, R.R., Wilhelm, T., Dumont, M., Scelfo, A., Gamba, R., Wong, C.K., Rancati, G., Smogorzewska, A., *et al.* (2021). CENP-A chromatin prevents replication stress at centromeres to avoid structural aneuploidy. *Proceedings of National Academy of Sciences* 118, e2015634118.
- Glynn, E.F., Megee, P.C., Yu, H.-G., Mistrot, C., Unal, E., Koshland, D.E., DeRisi, J.L., Gerton, J.L., and Stillman, B. (2004). Genome-wide mapping of the cohesin complex in the yeast *Saccharomyces cerevisiae*. *PLoS biology* 2, e259.
- Gravholt, C.H., Friedrich, U., Caprani, M., and Jørgensen, A.L. (1992). Breakpoints in Robertsonian translocations are localized to satellite III DNA by fluorescence *in situ* hybridization. *Genomics* 14, 924-930.
- Grigorova, M., Staines, J.M., Ozdag, H., Caldas, C., and Edwards, P.A. (2004). Possible causes of chromosome instability: comparison of chromosomal abnormalities in cancer cell lines with mutations in BRCA1, BRCA2, CHK2 and BUB1. *Cytogenetics and Genome Research* 104, 333-340.
- Grimaldi, K.A., McGurk, C.J., McHugh, P.J., and Hartley, J.A. (2002). PCR-based methods for detecting DNA damage and its repair at the sub-gene and single nucleotide levels in cells. *Molecular Biotechnology* 20, 181-196.
- Gruber, J.J., Chen, J., Geller, B., Jager, N., Lipchik, A.M., Wang, G., Kurian, A.W., Ford, J.M., and Snyder, M.P. (2019). Chromatin Remodeling in Response to BRCA2-Crisis. *Cell Reports* 28, 2182-2193 e2186.
- Gu, W., Zhang, F., and Lupski, J.R. (2008). Mechanisms for human genomic rearrangements. *Pathogenetics* 1, 4.
- Guerrero, A.A., Gamero, M.C., Trachana, V., Fütterer, A., Pacios-Bras, C., Díaz-Concha, N.P., Cigudosa, J.C., Martínez-A, C., and van Wely, K.H.M. (2010). Centromere-localized breaks indicate the generation of DNA damage by the mitotic spindle. *Proceedings of National Academy of Science* 107, 4159-4164.

- Haber, J.E., Rogers, D.T., and McCusker, J.H. (1980). Homothallic conversions of yeast mating-type genes occur by intrachromosomal recombination. *Cell* 22, 277-289.
- Hanada, K., Budzowska, M., Davies, S.L., van Drunen, E., Onizawa, H., Beverloo, H.B., Maas, A., Essers, J., Hickson, I.D., and Kanaar, R. (2007). The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. *Nature Structural and Molecular Biology* 14, 1096-1104.
- Hanseemann, D. (1890). Ueber asymmetrische Zelltheilung in Epithelkrebsen und deren biologische Bedeutung. *Archiv fur Pathologische Anatomie* 119, 299-326.
- Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., *et al.* (2020). Array programming with NumPy. *Nature* 585, 357-362.
- Hasson, D., Panchenko, T., Salimian, K.J., Salman, M.U., Sekulic, N., Alonso, A., Warburton, P.E., and Black, B.E. (2013). The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nature Structural Molecular Biology* 20, 687-695.
- Hauer, M.H., Seeber, A., Singh, V., Thierry, R., Sack, R., Amitai, A., Kryzhanovska, M., Eglinger, J., Holcman, D., Owen-Hughes, T., *et al.* (2017). Histone degradation in response to DNA damage enhances chromatin dynamics and recombination rates. *Nature Structural Molecular Biology* 24, 99-107.
- Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. *Cell* 118, 715-729.
- Hazan, I., Monin, J., Bouwman, B.A.M., Crosetto, N., and Aqeilan, R.I. (2019). Activation of Oncogenic Super-Enhancers Is Coupled with DNA Repair by RAD51. *Cell Reports* 29, 560-572.e564.
- Heeres, J.T., and Hergenrother, P.J. (2007). Poly(ADP-ribose) makes a date with death. *Current Opinions in Chemical Biology* 11, 644-653.
- Hegde, M.L., Hazra, T.K., and Mitra, S. (2008). Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Research* 18, 27-47.
- Hegedűs, É., Kókai, E., Nánási, P., Imre, L., Halász, L., Jossé, R., Antunovics, Z., Webb, M.R., El-Hage, A., Pommier, Y., *et al.* (2018). Endogenous single-strand DNA breaks at RNA polymerase II promoters in *Saccharomyces cerevisiae*. *Nucleic Acids Research* 46, 10649-10668.
- Henikoff, J.G., Thakur, J., Kasinathan, S., and Henikoff, S. (2015). A unique chromatin complex occupies young alpha-satellite arrays of human centromeres. *Science Advances* 1, e1400234.
- Henikoff, S., Ahmad, K., and Malik, H.S. (2001). The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. *Science* 293, 1098-1102.
- Herbert, S., Brion, A., Arbona, J.M., Lelek, M., Veillet, A., Lelandais, B., Parmar, J., Fernandez, F.G., Almayrac, E., Khalil, Y., *et al.* (2017). Chromatin stiffening underlies enhanced locus mobility after DNA damage in budding yeast. *EMBO Journal* 36, 2595-2608.
- Hermesen, M., Snijders, A., Guervós, M.A., Tænzer, S., Koerner, U., Baak, J., Pinkel, D., Albertson, D., van Diest, P., Meijer, G., *et al.* (2005). Centromeric chromosomal translocations show tissue-specific differences between squamous cell carcinomas and adenocarcinomas. *Oncogene* 24, 1571-1579.
- Hermesen, M.A.J.A., Joenje, H., Arwert, F., Welters, M.J.P., Braakhuis, B.J.M., Bagnay, M., Westerveld, A., and Slater, R. (1996). Centromeric breakage as a major cause of cytogenetic abnormalities in oral squamous cell carcinoma. *Genes, Chromosomes and Cancer* 15, 1-9.
- Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D., and Karpen, G.H. (2006). Mislocalization of the *Drosophila* centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Developmental Cell* 10, 303-315.
- Hinde, E., Kong, X., Yokomori, K., and Gratton, E. (2014). Chromatin dynamics during DNA repair revealed by pair correlation analysis of molecular flow in the nucleus. *Biophysics Journal* 107, 55-65.
- Ho, K.H., Tsuchiya, D., Oliger, A.C., and Lacefield, S. (2014). Localization and function of budding yeast CENP-A depends upon kinetochore protein interactions and is independent of canonical centromere sequence. *Cell Reports* 9, 2027-2033.
- Hoffmann, S., Dumont, M., Barra, V., Ly, P., Nechemia-Arbely, Y., McMahon, M.A., Herve, S., Cleve-

- land, D.W., and Fachinetti, D. (2016). CENP-A Is Dispensable for Mitotic Centromere Function after Initial Centromere/Kinetochore Assembly. *Cell Reports* 17, 2394-2404.
- Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genetical Research* 5, 282-304.
- Hori, T., Shang, W.H., Takeuchi, K., and Fukagawa, T. (2013). The CCAN recruits CENP-A to the centromere and forms the structural core for kinetochore assembly. *Journal of Cell Biology* 200, 45-60.
- Huang, F., Motlekar, N.A., Burgwin, C.M., Napper, A.D., Diamond, S.L., and Mazin, A.V. (2011). Identification of specific inhibitors of human RAD51 recombinase using high-throughput screening. *ACS Chemical Biology* 6, 628-635.
- Hudson, D.F., Fowler, K.J., Earle, E., Saffery, R., Kalitsis, P., Trowell, H., Hill, J., Wreford, N.G., de Kretser, D.M., Cancilla, M.R., *et al.* (1998). Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights. *Journal of Cell Biology* 141, 309-319.
- Hughes-Davies, L., Huntsman, D., Ruas, M., Fuks, F., Bye, J., Chin, S.F., Milner, J., Brown, L.A., Hsu, F., Gilks, B., *et al.* (2003). EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. *Cell* 115, 523-535.
- Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering* 9, 90-95.
- Iacovoni, J.S., Caron, P., Lassadi, I., Nicolas, E., Massip, L., Trouche, D., and Legube, G. (2010). High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. *EMBO Journal* 29, 1446-1457.
- Jackson, A.L., and Loeb, L.A. (2001). The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutation Research* 477, 7-21.
- Jackson, J.A., and Fink, G.R. (1981). Gene conversion between duplicated genetic elements in yeast. *Nature* 292, 306-311.
- Jaco, I., Canela, A., Vera, E., and Blasco, M.A. (2008). Centromere mitotic recombination in mammalian cells. *Journal of Cell Biology* 181, 885-892.
- Jansen, L.E., Black, B.E., Foltz, D.R., and Cleveland, D.W. (2007). Propagation of centromeric chromatin requires exit from mitosis. *Journal of Cell Biology* 176, 795-805.
- Jensen, R.B., Carreira, A., and Kowalczykowski, S.C. (2010). Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature* 467, 678-683.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337, 816.
- Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nature Reviews Molecular Cell Biology* 7, 335-346.
- Jonstrup, A.T., Thomsen, T., Wang, Y., Knudsen, B.R., Koch, J., and Andersen, A.H. (2008). Hairpin structures formed by alpha satellite DNA of human centromeres are cleaved by human topoisomerase IIalpha. *Nucleic Acids Research* 36, 6165-6174.
- Kabeche, L., Nguyen, H.D., Buisson, R., and Zou, L. (2018). A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. *Science* 359, 108-114.
- Kamath-Loeb, A.S., Loeb, L.A., Johansson, E., Burgers, P.M.J., and Fry, M. (2001). Interactions between the Werner Syndrome Helicase and DNA Polymerase δ Specifically Facilitate Copying of Tetraplex and Hairpin Structures of the d(CGG)_n Trinucleotide Repeat Sequence. *Journal of Biological Chemistry* 276, 16439-16446.
- Kapoor, M., Montes de Oca Luna, R., Liu, G., Lozano, G., Cummings, C., Mancini, M., Ouspenski, I., Brinkley, B.R., and May, G.S. (1998). The cenpB gene is not essential in mice. *Chromosoma* 107, 570-576.
- Karagiannis, T.C., Harikrishnan, K., and El-Osta, A. (2007). Disparity of histone deacetylase inhibition on repair of radiation-induced DNA damage on euchromatin and constitutive heterochromatin compartments. *Oncogene* 26, 3963-3971.

- Kasinathan, S., and Henikoff, S. (2018). Non-B-Form DNA Is Enriched at Centromeres. *Molecular Biology and Evolution* 35, 949-962.
- Kathe, S.D., Shen, G., and Wallace, S.S. (2004). Single-stranded breaks in DNA but not oxidative DNA base damages block transcriptional elongation by RNA polymerase II in HeLa cell nuclear extracts. *Journal of Biological Chemistry* 279, 18511-18520.
- Kato, T., Sato, N., Hayama, S., Yamabuki, T., Ito, T., Miyamoto, M., Kondo, S., Nakamura, Y., and Daigo, Y. (2007). Activation of Holliday Junction-Recognizing Protein Involved in the Chromosomal Stability and Immortality of Cancer Cells. *Cancer Research* 67, 8544.
- Kawashima, Y., Yamaguchi, N., Teshima, R., Narahara, H., Yamaoka, Y., Anai, H., Nishida, Y., and Hanada, K. (2017). Detection of DNA double-strand breaks by pulsed-field gel electrophoresis. *Genes Cells* 22, 84-93.
- Keeney, S. (2008). Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. *Genome Dynamics and Stability* 2, 81-123.
- Kennedy, L., and Shelbourne, P.F. (2000). Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? *Human Molecular Genetics* 9, 2539-2544.
- Khan, F.A., and Ali, S.O. (2017). Physiological Roles of DNA Double-Strand Breaks 6439169. *Journal of Nucleic Acids*.
- Khristich, A.N., and Mirkin, S.M. (2020). On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *Journal of Biological Chemistry* 295, 4134-4170.
- Kim, T.M., Xi, R., Luquette, L.J., Park, R.W., Johnson, M.D., and Park, P.J. (2013). Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. *Genome Research* 23, 217-227.
- Klein, H.L. (2020). Stressed DNA replication generates stressed DNA. *Proceedings of National Academy of Sciences* 117, 10108-10110.
- Klein, H.L., and Petes, T.D. (1981). Intrachromosomal gene conversion in yeast. *Nature* 289, 144-148.
- Kononenko, A.V., Ebersole, T., Vasquez, K.M., and Mirkin, S.M. (2018). Mechanisms of genetic instability caused by (CGG)_n repeats in an experimental mammalian system. *Nature Structural Molecular Biology* 25, 669-676.
- Kordon, M.M., Zarebski, M., Solarczyk, K., Ma, H., Pederson, T., and Dobrucki, J.W. (2020). STRIDE-a fluorescence method for direct, specific *in situ* detection of individual single- or double-strand DNA breaks in fixed cells. *Nucleic Acids Research* 48, e14.
- Koren, A., Tsai, H.J., Tirosh, I., Burrack, L.S., Barkai, N., and Berman, J. (2010). Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase. *PLoS Genetics* 6, e1001068.
- Kouprina, N., Samoshkin, A., Erliandri, I., Nakano, M., Lee, H.S., Fu, H., Iida, Y., Aladjem, M., Oshimura, M., Masumoto, H., *et al.* (2012). Organization of Synthetic Alphoid DNA Array in Human Artificial Chromosome (HAC) with a Conditional Centromere. *ACS Synthetic Biology* 1, 590-601.
- Krumm, A., Hickey, L.B., and Groudine, M. (1995). Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. *Genes and Development* 9, 559-572.
- Kuzminov, A. (2001). Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proceedings of National Academy of Sciences* 98, 8241-8246.
- Lacle, M.M., Moelans, C.B., Kornegoor, R., van der Pol, C., Witkamp, A.J., van der Wall, E., Rueschoff, J., Buerger, H., and van Diest, P.J. (2015). Chromosome 17 copy number changes in male breast cancer. *Cell Oncology* 38, 237-245.
- Le Baccon, P., Leroux, D., Dascalescu, C., Duley, S., Marais, D., Esmenjaud, E., Sotto, J.J., and Callanan, M. (2001). Novel evidence of a role for chromosome 1 pericentric heterochromatin in the pathogenesis of B-cell lymphoma and multiple myeloma. *Genes, Chromosomes and Cancer* 32, 250-264.
- Lee, C., Wevrick, R., Fisher, R.B., Ferguson-Smith, M.A., and Lin, C.C. (1997). Human centromeric DNAs. *Human Genetics* 100, 291-304.

- Lee, C.-S., Haber James, E., Gellert, M., and Craig, N. (2015). Mating-type Gene Switching in *Saccharomyces cerevisiae*. *Microbiology Spectrum* 3, 3.2.29.
- Lee, C.-S., Lee, K., Legube, G., and Haber, J.E. (2014). Dynamics of yeast histone H2A and H2B phosphorylation in response to a double-strand break. *Nature Structural and Molecular Biology* 21, 103-109.
- Lee, C.H., Lee, M., Kang, H.J., Kim, D.H., Kang, Y.H., Bae, S.H., and Seo, Y.S. (2013). The N-terminal 45-kDa domain of Dna2 endonuclease/helicase targets the enzyme to secondary structure DNA. *Journal of Biological Chemistry* 288, 9468-9481.
- Lee, H.R., Zhang, W., Langdon, T., Jin, W., Yan, H., Cheng, Z., and Jiang, J. (2005). Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. *Proceedings of National Academy of Sciences* 102, 11793-11798.
- Lee, J.A., Carvalho, C.M., and Lupski, J.R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 131, 1235-1247.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv* 1303.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Li, H., Xiao, J., Li, J., Lu, L., Feng, S., and Droge, P. (2009). Human genomic Z-DNA segments probed by the Z alpha domain of ADAR1. *Nucleic Acids Research* 37, 2737-2746.
- Li, X., and Heyer, W.-D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell Research* 18, 99-113.
- Li, Y., Zhu, Z., Zhang, S., Yu, D., Yu, H., Liu, L., Cao, X., Wang, L., Gao, H., and Zhu, M. (2011). ShRNA-targeted centromere protein A inhibits hepatocellular carcinoma growth. *PLoS One* 6, e17794.
- Li, Z., Liu, B., Jin, W., Wu, X., Zhou, M., Liu, V.Z., Goel, A., Shen, Z., Zheng, L., and Shen, B. (2018). hDNA2 nuclease/helicase promotes centromeric DNA replication and genome stability. *EMBO Journal* 37.
- Liang, F., Han, M., Romanienko, P.J., and Jasin, M. (1998). Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proceedings of the National Academy of Sciences* 95, 5172-5177.
- Lieber, M.R., Ma, Y., Pannicke, U., and Schwarz, K. (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nature Reviews Molecular Cell Biology* 4, 712-720.
- Lin, W., Sampath, S., Dai, H., Liu, C., Zhou, M., Hu, J., Huang, Q., Campbell, J., Shin-Ya, K., Zheng, L., *et al.* (2013). Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *EMBO Journal* 32, 1425-1439.
- Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* 362, 709-715.
- Lindahl, T., and Barnes, D.E. (2000). Repair of endogenous DNA damage. *Cold Spring Harbour Symposium on Quantitative Biology* 65, 127-133.
- Lindegren, C.C. (1955). Non-Mendelian Segregation in a Single Tetrad of *Saccharomyces* Ascribed to Gene Conversion. *Science* 121, 605-607.
- Lindsey, J.C., and Leach, D.R.F. (1989). Slow replication of palindrome-containing DNA. *Journal of Molecular Biology* 206, 779-782.
- Liskay, R.M., and Stachelek, J.L. (1983). Evidence for intrachromosomal gene conversion in cultured mouse cells. *Cell* 35, 157-165.
- Liu, H., Qu, Q., Warrington, R., Rice, A., Cheng, N., and Yu, H. (2015). Mitotic Transcription Installs Sgo1 at Centromeres to Coordinate Chromosome Segregation. *Molecular Cell* 59, 426-436.
- Liu, P., Carvalho, C.M., Hastings, P.J., and Lupski, J.R. (2012). Mechanisms for recurrent and complex human genomic rearrangements. *Current Opinions Genetics and Development* 22, 211-220.

- Liu, S.T., Rattner, J.B., Jablonski, S.A., and Yen, T.J. (2006). Mapping the assembly pathways that specify formation of the trilaminar kinetochore plates in human cells. *Journal of Cell Biology* 175, 41-53.
- Liu, Y., Liu, Q., Su, H., Liu, K., Xiao, X., Li, W., Sun, Q., Birchler, J.A., and Han, F. (2021). Genome-wide mapping reveals R-loops associated with centromeric repeats in maize. *Genome Research* 31, 1409-1418.
- Logsdon, G.A., Gambogi, C.W., Liskovych, M.A., Barrey, E.J., Larionov, V., Miga, K.H., Heun, P., and Black, B.E. (2019). Human Artificial Chromosomes that Bypass Centromeric DNA. *Cell* 178, 624-639.e619.
- Logsdon, G.A., Vollger, M.R., Hsieh, P., Mao, Y., Liskovych, M.A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P.C., Rhie, A., *et al.* (2021). The structure, function and evolution of a complete human chromosome 8. *Nature* 593, 101-107.
- Lomonosov, M. (2003). Stabilization of stalled DNA replication forks by the BRCA2 breast cancer susceptibility protein. *Genes and Development* 17, 3017-3022.
- Lopes, M., Foiani, M., and Sogo, J.M. (2006). Multiple Mechanisms Control Chromosome Integrity after Replication Fork Uncoupling and Restart at Irreparable UV Lesions. *Molecular Cell* 21, 15-27.
- Lopez-Delisle, L., Rabbani, L., Wolff, J., Bhardwaj, V., Backofen, R., Grüning, B., Ramírez, F., and Manke, T. (2021). pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* 37, 422-423.
- Lorkovic, Z.J., Park, C., Goiser, M., Jiang, D., Kurzbauer, M.T., Schlögelhofer, P., and Berger, F. (2017). Compartmentalization of DNA Damage Response between Heterochromatin and Euchromatin Is Mediated by Distinct H2A Histone Variants. *Current Biology* 27, 1192-1199.
- Luijsterburg, M.S., de Krijger, I., Wiegant, W.W., Shah, R.G., Smeenk, G., de Groot, A.J.L., Pines, A., Vertegaal, A.C.O., Jacobs, J.J.L., Shah, G.M., *et al.* (2016). PARP1 Links CHD2-Mediated Chromatin Expansion and H3.3 Deposition to DNA Repair by Non-homologous End-Joining. *Molecular Cell* 61, 547-562.
- Lukas, C., Melander, F., Stucki, M., Falck, J., Bekker-Jensen, S., Goldberg, M., Lerenthal, Y., Jackson, S.P., Bartek, J., and Lukas, J. (2004). Mdc1 couples DNA double-strand break recognition by Nbs1 with its H2AX-dependent chromatin retention. *EMBO J* 23, 2674-2683.
- Lyu, X., Chastain, M., and Chai, W. (2019). Genome-wide mapping and profiling of γ H2AX binding hotspots in response to different replication stress inducers. *BMC Genomics* 20, 579.
- Ma, W., Halweg, C.J., Menendez, D., and Resnick, M.A. (2012). Differential effects of poly(ADP-ribose) polymerase inhibition on DNA break repair in human cells are revealed with Epstein-Barr virus. *Proceedings of National Academy of Sciences* 109, 6590-6595.
- Ma, W., Westmoreland, J.W., and Resnick, M.A. (2013). Homologous recombination rescues ssDNA gaps generated by nucleotide excision repair and reduced translesion DNA synthesis in yeast G2 cells. *Proceedings of National Academy of Sciences* 110, E2895-2904.
- Maddox, P.S., Hyndman, F., Monen, J., Oegema, K., and Desai, A. (2007). Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. *Journal of Cell Biology* 176, 757-763.
- Maehara, K., Takahashi, K., and Saitoh, S. (2010). CENP-A reduction induces a p53-dependent cellular senescence response to protect cells from executing defective mitoses. *Molecular Cell Biology* 30, 2090-2104.
- Mah, L.J., El-Osta, A., and Karagiannis, T.C. (2010). γ H2AX: a sensitive molecular marker of DNA damage and repair. *Leukemia* 24, 679-686.
- Mahaney, B.L., Meek, K., and Lees-Miller, S.P. (2009). Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochemical Journal* 417, 639-650.
- Maizels, N., and Davis, L. (2018). Initiation of homologous recombination at DNA nicks. *Nucleic acids research* 46, 6962-6973.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science* 339, 823.

- Malik, H.S., and Bayes, J.J. (2006). Genetic conflicts during meiosis and the evolutionary origins of centromere complexity. *Biochemical Society Transactions* 34, 569-573.
- Malik, I., Kelley, C.P., Wang, E.T., and Todd, P.K. (2021). Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nature Reviews Molecular Cell Biology* 22, 589-607.
- Marshall, O.J., Chueh, A.C., Wong, L.H., and Choo, K.H. (2008). Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *American Journal of Human Genetics* 82, 261-282.
- Marsischky, G.T., Lee, S., Griffith, J., and Kolodner, R.D. (1999). *Saccharomyces cerevisiae* MSH2/6 complex interacts with Holliday junctions and facilitates their cleavage by phage resolution enzymes. *Journal of Biological Chemistry* 274, 7200-7206.
- Martínez, J.G., Pérez-Escuredo, J., Llorente, J.L., Suárez, C., and Hermesen, M.A. (2012). Localization of centromeric breaks in head and neck squamous cell carcinoma. *Cancer Genetics* 205, 622-629.
- Martins, C., Jin, Y., Jin, C., Wennerberg, J., Höglund, M., and Mertens, F. (1999). Fluorescent in situ hybridisation (FISH) characterisation of pericentromeric breakpoints on chromosome 5 in head and neck squamous cell carcinomas. *European Journal of Cancer* 35, 498-501.
- Mason, J.M., Chan, Y.-L., Weichselbaum, R.W., and Bishop, D.K. (2019). Non-enzymatic roles of human RAD51 at stalled replication forks. *Nature Communications* 10, 4410.
- Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., and Okazaki, T. (1989). A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *Journal of Cell Biology* 109, 1963-1973.
- Mateo, L., and González, J. (2014). Pogo-like Transposases Have Been Repeatedly Domesticated into CENP-B-Related Proteins. *Genome Biology and Evolution* 6, 2008-2016.
- Mather, K. (1939). Crossing over and Heterochromatin in the X Chromosome of *Drosophila Melanogaster*. *Genetics* 24, 413-435. McCarroll, R.M., and Fangman, W.L. (1988). Time of replication of yeast centromeres and telomeres. *Cell* 54, 505-513.
- McKinley, K.L., and Cheeseman, I.M. (2014). Polo-like kinase 1 licenses CENP-A deposition at centromeres. *Cell* 158, 397-411.
- McKinley, K.L., and Cheeseman, I.M. (2016). The molecular basis for centromere identity and function. *Nature Reviews Molecular Cell Biology* 17, 16-29.
- McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*. McNulty, S.M., and Sullivan, B.A. (2018). Alpha satellite DNA biology: finding function in the recesses of the genome. *Chromosome Research* 26, 115-138.
- McNulty, S.M., Sullivan, L.L., and Sullivan, B.A. (2017). Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. *Developmental Cell* 42, 226-240.
- Mellone, B.G., Grive, K.J., Shteyn, V., Bowers, S.R., Oderberg, I., and Karpen, G.H. (2011). Assembly of *Drosophila* centromeric chromatin proteins during mitosis. *PLoS Genet* 7, e1002068.
- Melters, D.P., Bradnam, K.R., Young, H.A., Telis, N., May, M.R., Ruby, J.G., Sebra, R., Peluso, P., Eid, J., Rank, D., *et al.* (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biology* 14, R10.
- Mendez-Bermudez, A., Lototska, L., Bauwens, S., Giraud-Panis, M.J., Croce, O., Jamet, K., Irizar, A., Mowinkel, M., Koundrioukoff, S., Nottet, N., *et al.* (2018). Genome-wide Control of Heterochromatin Replication by the Telomere Capping Protein TRF2. *Molecular Cell* 70, 449-461.
- Mendez-Bermudez, A., Pousse, M., Lototska, L., Tessier, F., Croce, O., Latrick, C., Cherdyntseva, V., Nassour, J., Jiang, X., Yiming, L.U., *et al.* (2021). Selective pericentromeric heterochromatin dismantling caused by TP53 activation during senescence. *Nature Portfolio*.
- Mendiburo, M.J., Padeken, J., Fülöp, S., Schepers, A., and Heun, P. (2011). *Drosophila* CENH3 is sufficient for centromere formation. *Science* 334, 686-690.
- Merta, L., Gandalovicova, A., Cermak, V., Dibus, M., Gutschner, T., Diederichs, S., Rosel, D., and

- Brabek, J. (2020). Increased Level of Long Non-Coding RNA MALAT1 is a Common Feature of Amoeboid Invasion. *Cancers (Basel)* 12.
- Meselson, M.S., and Radding, C.M. (1975). A general model for genetic recombination. *Proceedings of National Academy of Sciences* 72, 358-361.
- Miga, K.H. (2019). Centromeric Satellite DNAs: Hidden Sequence Variation in the Human Population. *Genes (Basel)* 10, 352.
- Miga, K.H., and Alexandrov, I.A. (2021). Variation and Evolution of Human Centromeres: A Field Guide and Perspective. *Annual Reviews Genetics* 55, 583-602.
- Mijic, S., Zellweger, R., Chappidi, N., Berti, M., Jacobs, K., Mutreja, K., Ursich, S., Ray Chaudhuri, A., Nussenzweig, A., Jancsak, P., *et al.* (2017). Replication fork reversal triggers fork degradation in BRCA2-defective cells. *Nature Communications* 8 859.
- Milligan, J.R., Ng, J.Y., Wu, C.C., Aguilera, J.A., Fahey, R.C., and Ward, J.F. (1995). DNA repair by thiols in air shows two radicals make a double-strand break. *Radiation Research* 143, 273-280.
- Milner, J., Ponder, B., Hughes-Davies, L., Seltmann, M., and Kouzarides, T. (1997). Transcriptional activation functions in BRCA2. *Nature* 386, 772-773.
- Mine-Hattab, J., and Rothstein, R. (2012). Increased chromosome mobility facilitates homology search during recombination. *Nature Cell Biology* 14, 510-517.
- Mitra, S., Gomez-Raja, J., Larriba, G., Dubey, D.D., and Sanyal, K. (2014). Rad51-Rad52 mediated maintenance of centromeric chromatin in *Candida albicans*. *PLoS Genetics* 10, e1004344.
- Molina, O., Kouprina, N., Masumoto, H., Larionov, V., and Earnshaw, W.C. (2017). Using human artificial chromosomes to study centromere assembly and function. *Chromosoma* 126, 559-575.
- Monckton, D.G., Cayuela, M.L., Gould, F.K., Brock, G.J., Silva, R., and Ashizawa, T. (1999). Very large (CAG)(n) DNA repeat expansions in the sperm of two spinocerebellar ataxia type 7 males. *Human Molecular Genetics* 8, 2473-2478.
- Moree, B., Meyer, C.B., Fuller, C.J., and Straight, A.F. (2011). CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. *Journal of Cell Biology* 194, 855-871.
- Moroni, F. (2008). Poly(ADP-ribose)polymerase 1 (PARP-1) and postischemic brain damage. *Current Opinions Pharmacology* 8, 96-103.
- Morrison, C., Shinohara, A., Sonoda, E., Yamaguchi-Iwai, Y., Takata, M., Weichselbaum Ralph, R., and Takeda, S. (1999). The Essential Functions of Human Rad51 Are Independent of ATP Hydrolysis. *Molecular and Cellular Biology* 19, 6891-6897.
- Moudry, P., Watanabe, K., Wolanin, K.M., Bartkova, J., Wassing, I.E., Watanabe, S., Strauss, R., Troelsgaard Pedersen, R., Oestergaard, V.H., Lisby, M., *et al.* (2016). TOPBP1 regulates RAD51 phosphorylation and chromatin loading and determines PARP inhibitor sensitivity. *Journal of Cell Biology* 212, 281-288.
- Muris, D.F., Vreeken, K., Carr, A.M., Broughton, B.C., Lehmann, A.R., Lohman, P.H., and Pastink, A. (1993). Cloning the RAD51 homologue of *Schizosaccharomyces pombe*. *Nucleic Acids Research* 21, 4586-4591.
- Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M., and Okazaki, T. (1992). Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. *Journal of Cell Biology* 116, 585-596.
- Nagylaki, T., and Petes, T.D. (1982). Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* 100, 315-337.
- Nakajima, K., Zhou, Y., Tomita, A., Hirade, Y., Gurumurthy, C.B., and Nakada, S. (2018). Precise and efficient nucleotide substitution near genomic nick via noncanonical homology-directed repair. *Genome Research* 28, 223-230.
- Nakano, M., Cardinale, S., Noskov, V., Gassmann, R., Vagnarelli, P., Kandels-Lewis, S., Larionov, V., Earnshaw, W., and Masumoto, H. (2008). Inactivation of a human kinetochore by specific targeting of chromatin modifiers. *Developmental Cell* 14, 507-522.

- Nechemia-Arbely, Y., Fachinetti, D., Miga, K.H., Sekulic, N., Soni, G.V., Kim, D.H., Wong, A.K., Lee, A.Y., Nguyen, K., Dekker, C., *et al.* (2017). Human centromeric CENP-A chromatin is a homotypic, octameric nucleosome at all cell cycle points. *Journal of Cell Biology* 216, 607-621.
- Nick McElhinny, S.A., Watts, B.E., Kumar, D., Watt, D.L., Lundström, E.B., Burgers, P.M.J., Johansson, E., Chabes, A., and Kunkel, T.A. (2010). Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proceedings of the National Academy of Sciences* 107, 4949.
- Nonaka, N., Kitajima, T., Yokobayashi, S., Xiao, G., Yamamoto, M., Grewal, S.I., and Watanabe, Y. (2002). Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. *Nature Cell Biology* 4, 89-93.
- Nonin-Lecomte, S., and Leroy, J.L. (2001). Structure of a C-rich strand fragment of the human centromeric satellite III: a pH-dependent intercalation topology. *Journal of Molecular Biology* 309, 491-506.
- North, J.A., Amunugama, R., Klajner, M., Bruns, A.N., Poirier, M.G., and Fishel, R. (2013). ATP-dependent nucleosome unwrapping catalyzed by human RAD51. *Nucleic Acids Res* 41, 7302-7312.
- Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., *et al.* (2021). The complete sequence of a human genome. *bioRxiv*.
- O'Donnell, L., Panier, S., Wildenhain, J., Tkach, J.M., Al-Hakim, A., Landry, M.C., Escibano-Diaz, C., Szilard, R.K., Young, J.T., Munro, M., *et al.* (2010). The MMS22L-TONSL complex mediates recovery from replication stress and homologous recombination. *Molecular Cell* 40, 619-631.
- Ohzeki, J., Nakano, M., Okada, T., and Masumoto, H. (2002). CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *Journal of Cell Biology* 159, 765-775.
- Ohzeki, J.I., Bergmann, J.H., Kouprina, N., Noskov, V.N., Nakano, M., Kimura, H., Earnshaw, W.C., Larionov, V., and Masumoto, H. (2012). Breaking the HAC barrier: histone H3K9 acetyl/methyl balance regulates CENP-A assembly. *EMBO Journal* 31, 2391-2402.
- Okada, M., Okawa, K., Isobe, T., and Fukagawa, T. (2009). CENP-H-containing complex facilitates centromere deposition of CENP-A in cooperation with FACT and CHD1. *Molecular Biology of the Cell* 20, 3986-3995.
- Okano, S., Lan, L., Caldecott, K.W., Mori, T., and Yasui, A. (2003). Spatial and temporal cellular responses to single-strand breaks in human cells. *Molecular Cell Biology* 23, 3974-3981.
- Olson, E., Nievera, C.J., Klimovich, V., Fanning, E., and Wu, X. (2006). RPA2 is a direct downstream target for ATR to regulate the S-phase checkpoint. *Journal of Biological Chemistry* 281, 39517-39533.
- Onaka, A.T., Toyofuku, N., Inoue, T., Okita, A.K., Sagawa, M., Su, J., Shitanda, T., Matsuyama, R., Zafar, F., Takahashi, T.S., *et al.* (2016). Rad51 and Rad54 promote noncrossover recombination between centromere repeats on the same chromatid to prevent isochromosome formation. *Nucleic Acids Research* 44, 10744-10757.
- Ou, Z., Stankiewicz, P., Xia, Z., Breman, A.M., Dawson, B., Wiszniewska, J., Szafranski, P., Cooper, M.L., Rao, M., Shao, L., *et al.* (2011). Observation and prediction of recurrent human translocations mediated by NAHR between nonhomologous chromosomes. *Genome Research* 21, 33-46.
- Owen, B.A., Yang, Z., Lai, M., Gajec, M., Badger, J.D., 2nd, Hayes, J.J., Edelman, W., Kucherlapati, R., Wilson, T.M., and McMurray, C.T. (2005). (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nature Structural Molecular Biology* 12, 663-670.
- Palmer, D.K., O'Day, K., Wener, M.H., Andrews, B.S., and Margolis, R.L. (1987). A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *Journal of Cell Biology* 104, 805-815.
- Panzarino, N.J., Kraus, J., Peng, M., Mosqueda, M., Nayak, S., Bond, S., Calvo, J., Cong, K., Doshi, M., Bere, M., *et al.* (2019). Replication gaps underlie BRCA-deficiency and therapy response. *bioRxiv*, 781955.
- Park, S.H., Kang, N., Song, E., Wie, M., Lee, E.A., Hwang, S., Lee, D., Ra, J.S., Park, I.B., Park, J., *et al.* (2019). ATAD5 promotes replication restart by regulating RAD51 and PCNA in response to replication stress. *Nature Communications* 10, 5718.
- Paull, T.T., Rogakou, E.P., Yamazaki, V., Kirchgessner, C.U., Gellert, M., and Bonner, W.M. (2000). A

critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Current Biology* 10, 886-895.

Peak, M.J., and Peak, J.G. (1982). Single-strand breaks induced in *Bacillus subtilis* DNA by ultraviolet light: action spectrum and properties. *Photochemistry and Photobiology* 35, 675-680.

Perez-Castro, A.V., Shamanski, F.L., Meneses, J.J., Lovato, T.L., Vogel, K.G., Moyzis, R.K., and Pedersen, R. (1998). Centromeric protein B null mice are viable with no apparent abnormalities. *Developmental Biology* 201, 135-143.

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

Piberger, A.L., Bowry, A., Kelly, R.D.W., Walker, A.K., Gonzalez-Acosta, D., Bailey, L.J., Doherty, A.J., Mendez, J., Morris, J.R., Bryant, H.E., *et al.* (2020). PrimPol-dependent single-stranded gap formation mediates homologous recombination at bulky DNA adducts. *Nature Communications* 11, 5863.

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

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

Promonet, A., Padioleau, I., Liu, Y., Sanz, L., Biernacka, A., Schmitz, A.L., Skrzypczak, M., Sarrazin, A., Mettling, C., Rowicka, M., *et al.* (2020). Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. *Nature Communications* 11, 3940.

Quenet, D., and Dalal, Y. (2014). A long non-coding RNA is required for targeting centromeric protein A to the human centromere. *Elife* 3, e03254.

Rabiau, N., Dechelotte, P., Adjakly, M., Kemeny, J.L., Guy, L., Boiteux, J.P., Kwiatkowski, F., Bignon, Y.J., and Bernard-Gallon, D. (2011). BRCA1, BRCA2, AR and IGF-I expression in prostate cancer: correlation between RT-qPCR and immunohistochemical detection. *Oncology Reports* 26, 695-702.

Racca, C., Britton, S., Hedouin, S., Francastel, C., Calsou, P., and Larminat, F. (2021). BRCA1 prevents R-loop-associated centromeric instability. *Cell Death and Disease* 12, 896.

Radding, C.M. (1982). Homologous pairing and strand exchange in genetic recombination. *Annual Review of Genetics* 16, 405-437.

Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Grüning, B.A., Villaveces, J., Habermann, B., Akhtar, A., and Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature Communications* 9, 189.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* 44, W160-W165.

Rasmussen, E.B., and Lis, J.T. (1993). *In vivo* transcriptional pausing and cap formation on three *Drosophila* heat shock genes. *Proceedings of National Academy of Sciences* 90, 7923-7927.

Rattner, J.B., Hendzel, M.J., Furbee, C.S., Muller, M.T., and Bazett-Jones, D.P. (1996). Topoisomerase II alpha is associated with the mammalian centromere in a cell cycle- and species-specific manner and is required for proper centromere/kinetochore structure. *Journal of Cell Biology* 134, 1097-1107.

Regnier, V., Vagnarelli, P., Fukagawa, T., Zerjal, T., Burns, E., Trouche, D., Earnshaw, W., and Brown, W. (2005). CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1. *Molecular Cell Biology* 25, 3967-3981.

Roberts, S.A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L.J., Kryukov, G.V., Malc, E., Mieczkowski, P.A., *et al.* (2012). Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Molecular Cell* 46, 424-435.

Rogakou, E.P., Boon, C., Redon, C., and Bonner, W.M. (1999). Megabase chromatin domains involved in DNA double-strand breaks *in vivo*. *Journal of Cell Biology*. 146, 905-916.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded

- breaks induce histone H2AX phosphorylation on serine 139. *Journal of Biological Chemistry* 273, 5858-5868.
- Rosic, S., Kohler, F., and Erhardt, S. (2014). Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. *Journal of Cell Biology* 207, 335-349.
- Rouet, P., Smih, F., and Jasin, M. (1994). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences* 91, 6064.
- Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell* 54, 795-804.
- Rudd, M.K., Wray, G.A., and Willard, H.F. (2006). The evolutionary dynamics of alpha-satellite. *Genome Res* 16, 88-96.
- Rydberg, B., and Game, J. (2002). Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proceedings of National Academy of Sciences* 99, 16654-16659.
- Saayman, X., and Esashi, F. (2021). Breaking the paradigm: early insights from mammalian DNA breakomes. *The FEBS Journal*.
- Sassa, A., Yasui, M., and Honma, M. (2019). Current perspectives on mechanisms of ribonucleotide incorporation and processing in mammalian DNA. *Genes and Environment* 41, 3.
- Sax, K. (1938). Chromosome Aberrations Induced by X-Rays. *Genetics* 23, 494-516.
- Scherer, S., and Davis, R.W. (1980). Recombination of dispersed repeated DNA sequences in yeast. *Science* 209, 1380-1384.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9, 676-682.
- Schlacher, K., Christ, N., Siaud, N., Egashira, A., Wu, H., and Jasin, M. (2011). Double-Strand Break Repair-Independent Role for BRCA2 in Blocking Stalled Replication Fork Degradation by MRE11. *Cell* 145, 529-542.
- Schlacher, K., Wu, H., and Jasin, M. (2012). A Distinct Replication Fork Protection Pathway Connects Fanconi Anemia Tumor Suppressors to RAD51-BRCA1/2. *Cancer Cell* 22, 106-116.
- Schueler, M.G., Dunn, J.M., Bird, C.P., Ross, M.T., Viggiano L., NISC Comparative Sequencing Program, Rocchi, M., Willard, H.F., and Green, E. (2005). Progressive proximal expansion of the primate X chromosome centromere. *Proceedings of National Academy of Sciences* 102, 10563-10568.
- Scott, K.C., and Sullivan, B.A. (2014). Neocentromeres: a place for everything and everything in its place. *Trends in Genetics* 30, 66-74.
- Sedelnikova, O.A., Rogakou, E.P., Panyutin, I.G., and Bonner, W.M. (2002). Quantitative detection of (125)IdU-induced DNA double-strand breaks with gamma-H2AX antibody. *Radiation Research* 158, 486-492.
- Shi, J., Wolf, S.E., Burke, J.M., Presting, G.G., Ross-Ibarra, J., and Dawe, R.K. (2010). Widespread gene conversion in centromere cores. *PLoS Biol* 8, e1000327.
- Shibata, T., DasGupta, C., Cunningham, R.P., Williams, J.G., Osber, L., and Radding, C.M. (1981). Homologous pairing in genetic recombination. The pairing reaction catalyzed by *Escherichia coli* recA protein. *Journal of Biological Chemistry* 256, 7565-7572.
- Shinohara, A., Ogawa, H., and Ogawa, T. (1992). Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* 69, 457-470.
- Shivji, M.K.K., Renaudin, X., Williams, C.H., and Venkitaraman, A.R. (2018). BRCA2 Regulates Transcription Elongation by RNA Polymerase II to Prevent R-Loop Accumulation. *Cell Reports* 22, 1031-1039.
- Shuaib, M., Ouarrhni, K., Dimitrov, S., and Hamiche, A. (2010). HJURP binds CENP-A via a highly conserved N-terminal domain and mediates its deposition at centromeres. *Proceedings of National Academy of Sciences* 107, 1349-1354.

- Silva, M.C., Bodor, D.L., Stellfox, M.E., Martins, N.M., Hocheegger, H., Foltz, D.R., and Jansen, L.E. (2012). Cdk activity couples epigenetic centromere inheritance to cell cycle progression. *Developmental Cell* 22, 52-63.
- Smedley, D., Sidhar, S., Birdsall, S., Bennett, D., Herlyn, M., Cooper, C., and Shipley, J. (2000). Characterization of chromosome 1 abnormalities in malignant melanomas. *Genes, Chromosomes and Cancer* 28, 121-125.
- Smith, A.M., Takeuchi, R., Pellenz, S., Davis, L., Maizels, N., Monnat, R.J., and Stoddard, B.L. (2009). Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proceedings of National Academy of Sciences* 106, 5099-5104.
- Smith, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* 191, 528-535.
- Smith, K.D., Fu, M.A., and Brown, E.J. (2009). Tim-Tipin dysfunction creates an indispensable reliance on the ATR-Chk1 pathway for continued DNA synthesis. *Journal of Cell Biology* 187, 15-23.
- Sollier, J., Stork, C.T., García-Rubio, M.L., Paulsen, R.D., Aguilera, A., and Cimprich, K.A. (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Molecular Cell* 56, 777-785.
- Song, K.Y., Chekuri, L., Rauth, S., Ehrlich, S., and Kucherlapati, R. (1985). Effect of double-strand breaks on homologous recombination in mammalian cells and extracts. *Molecular Cell Biology* 5, 3331-3336.
- Sonoda, E., Sasaki, M.S., Buerstedde, J.M., Bezzubova, O., Shinohara, A., Ogawa, H., Takata, M., Yamaguchi-Iwai, Y., and Takeda, S. (1998). Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death. *EMBO Journal* 17, 598-608.
- Sparks, J.L., Chon, H., Cerritelli, S.M., Kunkel, T.A., Johansson, E., Crouch, R.J., and Burgers, P.M. (2012). RNase H2-initiated ribonucleotide excision repair. *Molecular Cell* 47, 980-986.
- Sriramachandran, A.M., Petrosino, G., Méndez-Lago, M., Schäfer, A.J., Batista-Nascimento, L.S., Zilio, N., and Ulrich, H.D. (2020). Genome-wide Nucleotide-Resolution Mapping of DNA Replication Patterns, Single-Strand Breaks, and Lesions by GLOE-Seq. *Molecular Cell* 78, 975-985.
- Stankiewicz, P., and Lupski, J.R. (2002). Genome architecture, rearrangements and genomic disorders. *Trends in Genetics* 18, 74-82.
- Strehler, B.L. (1986). Genetic instability as the primary cause of human aging. *Experimental Gerontology* 21, 283-319.
- Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* 123, 1213-1226.
- Sung, P. (1994). Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science* 265, 1241-1243.
- Surova, O., and Zhivotovsky, B. (2013). Various modes of cell death induced by DNA damage. *Oncogene* 32, 3789-3797.
- Svoboda, D.L., and Vos, J.M. (1995). Differential replication of a single, UV-induced lesion in the leading or lagging strand by a human cell extract: fork uncoupling or gap formation. *Proceedings of National Academy of Sciences* 92, 11975-11979.
- Sy, S.M.H., Huen, M.S.Y., and Chen, J. (2009). PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proceedings of National Academy of Sciences* 106, 7155-7160.
- Symington, L.S. (2002). Role of RAD52 Epistasis Group Genes in Homologous Recombination and Double-Strand Break Repair. *Microbiology Molecular Biology Reviews* 66, 630-670.
- Szillard, R.K., Jacques, P.-É., Laramée, L., Cheng, B., Galicia, S., Bataille, A.R., Yeung, M., Mendez, M., Bergeron, M., Robert, F., *et al.* (2010). Systematic identification of fragile sites via genome-wide location analysis of γ -H2AX. *Nature structural and molecular biology* 17, 299-305.
- Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., and Stahl, F.W. (1983). The double-strand-break

repair model for recombination. *Cell* 33, 25-35.

Szostak, J.W., and Wu, R. (1980). Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. *Nature* 284, 426-430.

Taccioli, G.E., Rathbun, G., Oltz, E., Stamato, T., Jeggo, P.A., and Alt, F.W. (1993). Impairment of V(D)J recombination in double-strand break repair mutants. *Science* 260, 207-210.

Taglialatela, A., Alvarez, S., Leuzzi, G., Sannino, V., Ranjha, L., Huang, J.-W., Madubata, C., Anand, R., Levy, B., Rabadan, R., *et al.* (2017). Restoration of Replication Fork Stability in BRCA1- and BRCA2-Deficient Cells by Inactivation of SNF2-Family Fork Remodelers. *Molecular Cell* 68, 414-430.e418.

Takahashi, K., Murakami, S., Chikashige, Y., Funabiki, H., Niwa, O., and Yanagida, M. (1992). A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. *Molecular Biology of the Cell* 3, 819-835.

Takizawa, Y., Ho, C.H., Tachiwana, H., Matsunami, H., Kobayashi, W., Suzuki, M., Arimura, Y., Hori, T., Fukagawa, T., Ohi, M.D., *et al.* (2020). Cryo-EM Structures of Centromeric Tri-nucleosomes Containing a Central CENP-A Nucleosome. *Structure* 28, 44-53 e44.

Talbert, P.B., Kasinathan, S., and Henikoff, S.A. (2018). Simple and Complex Centromeric Satellites in *Drosophila* Sibling Species. *Genetics* 208, 977-990.

Tan-Wong, S.M., Dhir, S., and Proudfoot, N.J. (2019). R-Loops Promote Antisense Transcription across the Mammalian Genome. *Molecular Cell* 76, 600-616 e606.

Tanaka, Y., Nureki, O., Kurumizaka, H., Fukai, S., Kawaguchi, S., Ikuta, M., Iwahara, J., Okazaki, T., and Yokoyama, S. (2001). Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. *EMBO Journal* 20, 6612-6618.

Ten Hagen, K.G., Gilbert, D.M., Willard, H.F., and Cohen, S.N. (1990). Replication timing of DNA sequences associated with human centromeres and telomeres. *Molecular and cellular biology* 10, 6348-6355.

Teng, Y., Yadav, T., Duan, M., Tan, J., Xiang, Y., Gao, B., Xu, J., Liang, Z., Liu, Y., Nakajima, S., *et al.* (2018). ROS-induced R loops trigger a transcription-coupled but BRCA1/2-independent homologous recombination pathway through CSB. *Nature Communications* 9, 4115.

Thakur, J., and Henikoff, S. (2018). Unexpected conformational variations of the human centromeric chromatin complex. *Genes and Development* 32, 20-25.

Thakur, J., Packiaraj, J., and Henikoff, S. (2021). Sequence, Chromatin and Evolution of Satellite DNA. *International Journal of Molecular Sciences* 22, 4309.

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

Thomas, B.J., and Rothstein, R. (1989). Elevated recombination rates in transcriptionally active DNA. *Cell* 56, 619-630.

Thompson, F.H., Taetle, R., Trent, J.M., Liu, Y., Massey-Brown, K., Scott, K.M., Weinstein, R.S., Emerson, J.C., Alberts, D.S., and Nelson, M.A. (1997). Band 1p36 Abnormalities and t(1;17) in Ovarian Carcinoma. *Cancer Genetics and Cytogenetics* 96, 106-110.

Tirman, S., Quinet, A., Wood, M., Meroni, A., Cybulla, E., Jackson, J., Pegoraro, S., Simoneau, A., Zou, L., and Vindigni, A. (2021). Temporally distinct post-replicative repair mechanisms fill PRIMPOL-dependent ssDNA gaps in human cells. *Molecular Cell* 81, 4026-4040 e4028.

Trujillo, K.M., and Sung, P. (2001). DNA structure-specific nuclease activities in the *Saccharomyces cerevisiae* Rad50-Mre11 complex. *Journal of Biological Chemistry* 276, 35458-35464.

Tsouroula, K., Furst, A., Rogier, M., Heyer, V., Maglott-Roth, A., Ferrand, A., Reina-San-Martin, B., and Soutoglou, E. (2016). Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Molecular Cell* 63, 293-305.

Tsuzuki, T., Fujii, Y., Sakumi, K., Tominaga, Y., Nakao, K., Sekiguchi, M., Matsushiro, A., Yoshimura, Y., and Morita T (1996). Targeted disruption of the Rad51 gene leads to lethality in embryonic mice.

Proceedings of National Academy of Sciences 93, 6236-6240.

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

Tudor, M., Lobočka, M., Goodell, M., Pettitt, J., and O'Hare, K. (1992). The pogo transposable element family of *Drosophila melanogaster*. *Molecular Genetics and Genomics* 232, 126-134.

Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., Pantescio, V., De Vos, J., Thomas, A., Theillet, C., *et al.* (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature Cell Biology* 11, 1315-1324.

Tyler-Smith, C., Oakey, R.J., Larin, Z., Fisher, R.B., Crocker, M., Affara, N.A., Ferguson-Smith, M.A., Muenke, M., Zuffardi, O., and Jobling, M.A. (1993). Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. *Nature Genetics* 5, 368-375.

Uziel, T., Lerenthal, Y., Moyal, L., Andegeko, Y., Mittelman, L., and Shiloh, Y. (2003). Requirement of the MRN complex for ATM activation by DNA damage. *EMBO Journal* 22, 5612-5621.

Van, C., Yan, S., Michael, W.M., Waga, S., and Cimprich, K.A. (2010). Continued primer synthesis at stalled replication forks contributes to checkpoint activation. *Journal of Cell Biology* 189, 233-246.

Van Hooser, A.A., Ouspenski, I.I., Gregson, H.C., Starr, D.A., Yen, T.J., Goldberg M.L., Yokomori K.Y., Earnshaw W.C., Sullivan K.F., and Brinkley, B.R. (2001). Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. *Journal of Cell Science* 114, 3529-3542.

van Mameren, J., Modesti, M., Kanaar, R., Wyman, C., Peterman, E.J.G., and Wuite, G.J. (2009). Counting RAD51 proteins disassembling from nucleoprotein filaments under tension. *Nature* 457, 745-748.

Van Rossum, G., and Drake Jr, F.L. (1995). Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam.

Vance, G.H., Curtis, C.A., Heerema, N.A., Schwartz, S., and Palmer, C.G. (1997). An apparently acentric marker chromosome originating from 9p with a functional centromere without detectable alpha and beta satellite sequences. *American Journal of Medical Genetics* 71, 436-442.

Ventura, M., Mudge, J.M., Palumbo, V., Burn, S., Blennow, E., Pierluigi, M., Giorda, R., Zuffardi, O., Archidiacono, N., Jackson, M.S., *et al.* (2003). Neocentromeres in 15q24-26 map to duplicons which flanked an ancestral centromere in 15q25. *Genome Res* 13, 2059-2068.

Vilenchik, M.M., and Knudson, A.G. (2003). Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proceedings of National Academy of Sciences* 100, 12871-12876.

Vincenten, N., Kuhl, L.M., Lam, I., Oke, A., Kerr, A.R., Hochwagen, A., Fung, J., Keeney, S., Vader, G., and Marston, A.L. (2015). The kinetochore prevents centromere-proximal crossover recombination during meiosis. *Elife* 4.

Voullaire, L.E., Slater, H.R., Petrovic, V., and Choo, K.H. (1993). A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *American Journal of Human Genetics* 52, 1153-1163.

Vriend, L.E., and Krawczyk, P.M. (2017). Nick-initiated homologous recombination: Protecting the genome, one strand at a time. *DNA Repair (Amsterdam)* 50, 1-13.

Wang, A.T., Kim, T., Wagner, J.E., Conti, B.A., Lach, F.P., Huang, A.L., Molina, H., Sanborn, E.M., Zierhut, H., Cornes, B.K., *et al.* (2015). A Dominant Mutation in Human RAD51 Reveals Its Function in DNA Interstrand Crosslink Repair Independent of Homologous Recombination. *Molecular Cell* 59, 478-490.

Wang, G., and Vasquez, K.M. (2006). Non-B DNA structure-induced genetic instability. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 598, 103-119.

Wang, J., Liu, X., Dou, Z., Chen, L., Jiang, H., Fu, C., Fu, G., Liu, D., Zhang, J., Zhu, T., *et al.* (2014). Mitotic regulator Mis18beta interacts with and specifies the centromeric assembly of molecular chaperone holliday junction recognition protein (HJURP). *Journal of Biological Chemistry* 289, 8326-8336.

- Warburton, P.E., Cooke, C.A., Bourassa, S., Sullivan B.A., Stetten G., Gimelli G., Warburton D., Tyler-Smith C., Sullivan K.F., Poirier G.G., and Earnshaw, W.C. (1997). Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Current Biology* 7, 901-904.
- Warburton, P.E., and Willard, H.F. (1995). Interhomologue sequence variation of alpha satellite DNA from human chromosome 17: evidence for concerted evolution along haplotypic lineages. *Journal of Molecular Evolution*. 41, 1006-1015.
- Ward, I.M., and Chen, J. (2001). Histone H2AX is phosphorylated in an ATR-dependent manner in response to replicational stress. *Journal of Biological Chemistry* 276, 47759-47762.
- Ward, I.M., Minn, K., and Chen, J. (2004). UV-induced ataxia-telangiectasia-mutated and Rad3-related (ATR) activation requires replication stress. *Journal of Biological Chemistry* 279, 9677-9680.
- Waskom, M.L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software* 6, 3021.
- Wassing, I.E., Saayman, X., Rampazzo, L., Ralf, C., Bassett, A., and Esashi, F. (2020). The RAD51 recombinase protects mitotic chromatin in human cells. *Nature Communications* 12, 5380.
- Wei, H., and Yu, X. (2016). Functions of PARylation in DNA Damage Repair Pathways. *Genomics Proteomics Bioinformatics* 14, 131-139.
- Willard, H.F. (1985). Chromosome-specific organization of human alpha satellite DNA. *American Journal of Human Genetics* 37, 524-532.
- Willard, H.F. (1989). The genomics of long tandem arrays of satellite DNA in the human genome. *Genome* 31, 737-744.
- Willard, H.F., and Wayne, J.S. (1987). Chromosome-specific subsets of human alpha satellite DNA: analysis of sequence divergence within and between chromosomal subsets and evidence for an ancestral pentameric repeat. *Journal of Molecular Evolution*. 25, 207-214.
- Wolff, D.J., Miller, A.P., Van Dyke, D.L., Schwartz, S., and Willard, H.F. (1996). Molecular definition of breakpoints associated with human Xq isochromosomes: implications for mechanisms of formation. *American Journal of Human Genetics* 58, 154-160.
- Wong, L.H., Brettingham-Moore, K.H., Chan, L., Quach, J.M., Anderson, M.A., Northrop, E.L., Hannan, R., Saffery, R., Shaw, M.L., Williams, E., *et al.* (2007). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Research* 17, 1146-1160.
- Wong, N., John, S., Nussenzweig, A., and Canela, A. (2021). END-seq: An Unbiased, High-Resolution, and Genome-Wide Approach to Map DNA Double-Strand Breaks and Resection in Human Cells. *Methods Molecular Biology* 2153, 9-31.
- Wu, J.C., and Manuelidis, L. (1980). Sequence definition and organization of a human repeated DNA. *Journal of Molecular Biology* 142, 363-386.
- Yata, K., Bleuyard, J.-Y., Nakato, R., Ralf, C., Katou, Y., Schwab, Rebekka, A., Niedzwiedz, W., Shirahige, K., and Esashi, F. (2014). BRCA2 Coordinates the Activities of Cell-Cycle Kinases to Promote Genome Stability. *Cell Reports* 7, 1547-1559.
- Yata, K., Lloyd, J., Maslen, S., Bleuyard, J.-Y., Skehel, M., Smerdon, Stephen, J., and Esashi, F. (2012). Plk1 and CK2 Act in Concert to Regulate Rad51 during DNA Double Strand Break Repair. *Molecular Cell* 45, 371-383.
- Yilmaz, D., Furst, A., Meaburn, K., Lezaja, A., Wen, Y., Altmeyer, M., Reina-San-Martin, B., and Soutoglou, E. (2021). Activation of homologous recombination in G1 preserves centromeric integrity. *Nature* 600, 748-753.
- Yuan, S.S.F., Lee, S.Y., Chen, G., Song, M., Tomlinson, G.E., and Lee, E.Y.H.P. (1999). BRCA2 Is Required for Ionizing Radiation-induced Assembly of Rad51 Complex in Vivo. *Cancer Res* 59, 3547-3551.
- Zafar, F., Okita, A.K., Onaka, A.T., Su, J., Katahira, Y., Nakayama, J.-i., Takahashi, T.S., Masukata, H., and Nakagawa, T. (2017). Regulation of mitotic recombination between DNA repeats in centromeres. *Nucleic Acids Research* 45, 11222-11235.

Zeitlin, S.G., Baker, N.M., Chapados, B.R., Soutoglou, E., Wang, J.Y., Berns, M.W., and Cleveland, D.W. (2009). Double-strand DNA breaks recruit the centromeric histone CENP-A. *Proceedings of National Academy of Sciences* 106, 15762-15767.

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

Zhou, W., and Doetsch, P.W. (1993). Effects of abasic sites and DNA single-strand breaks on prokaryotic RNA polymerases. *Proceedings of National Academy of Sciences* 90, 6601-6605.

Zhou, W., and Doetsch, P.W. (1994). Transcription bypass or blockage at single-strand breaks on the DNA template strand: effect of different 3' and 5' flanking groups on the T7 RNA polymerase elongation complex. *Biochemistry* 33, 14926-34.

6.3 Sequencing: NGS alignment pipeline [bash]

```
# run the following file from terminal as: sh ./file.sh SRR695402
# where accession of interest is only input

#!/usr/bin/env bash
FILE=${1?Error: no file given.}
echo -e "working with file:" $FILE " \n "

# downloading FASTQ file
../packages/sratoolkit/bin/fasterq-dump ./ $FILE # input

# trimming FASTQ file
# HEADCROP and CROP parameters should be manually determined following FASTQC on
un-trimmed file
java -jar ../packages/Trimmomatic-0.39/trimmomatic-0.39.jar SE -phred33 ./f$FILE.
fastq "$FILE"_trimmed.fastq ILLUMINACLIP:../packages/Trimmomatic-0.39/adapters
/NexteraPE-PE.fa:2:30:10 LEADING:5 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:20

# performing QC
fastqc "$FILE"_trimmed.fastq

# aligning to hg38 reference genome with bwa aln
bwa aln -t 16 -n 0 ../refseq/chm13/chm13-v1.1.fasta "$FILE"_trimmed.fastq > "$FILE
"_chm13.sai

# converting to sam file
```

```

bwa samse ../refseq/chm13/chm13_v1.1.fasta "$FILE"_chm13.sai "$FILE"_trimmed.fastq
    > "$FILE"_chm13.sam

# converting to bam file
samtools view -bSu "$FILE"_chm13.sam -o "$FILE"_chm13.bam

echo "removing any duplicate reads to remove PCR duplicates"
samtools fixmate -m "$FILE"_chm13.bam "$FILE"_chm13_fixmate.bam
samtools sort "$FILE"_chm13_fixmate.bam -o "$FILE"_chm13_fixmate.sorted.bam
samtools markdup -r "$FILE"_chm13_fixmate.sorted.bam "$FILE"_chm13_nodup.sorted.
    bam

# indexing sorted & unique bam file
samtools view -bu -F 2308 "$FILE"_chm13_nodup.sorted.bam -o "$FILE"
    _chm13_nodup.2038.sorted.bam
samtools index "$FILE"_chm13_nodup.2038.sorted.bam
samtools flagstat "$FILE"_chm13_nodup.2038.sorted.bam

# converting to bw, normalizing using RPKM.
# bw can be used to visualize sample in reader e.g., Integrated Genome Browser
bamCoverage -b ./"$FILE"_chm13_nodup.2038.sorted.bam -o ./"$FILE"_chm13.sorted.bw
    --normalizeUsing RPKM

# converting to bedgraph
bigWigToBedGraph ./"$FILE"_chm13.sorted.bw "$FILE"_chm13.bedgraph

# converting to bed file
awk '{ \
    if ($1 ~ /^chr/) { \
        print $1"\t"$2"\t"$3"\tid—"NR"\t"$4; \
    } \
}' "$FILE"_chm13.bedgraph > "$FILE"_chm13.bed
sort-bed "$FILE"_chm13.bed > "$FILE"_chm13.sorted.bed

# finally , create bed files spanning read scores over centromeres as annotated by
    T2T

# cen_merged.bed must be bed filing listing all centromere coordinates
bedmap --echo --sum ./cen_merged.bed "$FILE"_chm13.sorted.bed > "$FILE"_cen.bed

```

6.4 NGS: per-chromosome enrichment score analysis [python]

```
from matplotlib import pyplot as plt
from matplotlib import rc
import pandas as pd
import numpy as np
import os
import sys

tx=str(sys.argv[1])
ctrl=str(sys.argv[2])

# set working directory containing your sequence analysis files:
os.chdir('.././.././../')

# set plotting parameters
plt.style.use('ggplot')
#plt.rcParams['font.sans-serif'] = 'Gill Sans'
rc('font',**{'family':'sans-serif','serif':['Helvetica']})
plt.rcParams['pdf.fonttype'] = 42

plt.rcParams['axes.edgecolor']='#333F4B'
plt.rcParams['axes.linewidth']=0.8
plt.rcParams['xtick.color']='#333F4B'
plt.rcParams['ytick.color']='#333F4B'

# import BED files denoting reads at centromeric regions
tx_df = pd.read_csv(tx, sep='[\t]', header=None, engine='python')
header = ['chrom', 'chromStart', 'chromEnd', 'score']
tx_df.columns = header[:len(tx_df.columns)]
tx_df['score']=pd.to_numeric(tx_df['score'], errors='coerce')

ctrl_df = pd.read_csv(ctrl, sep='[\t]', header=None, engine='python')
header = ['chrom', 'chromStart', 'chromEnd', 'score']
ctrl_df.columns = header[:len(ctrl_df.columns)]
ctrl_df['score']=pd.to_numeric(ctrl_df['score'], errors='coerce')

# calculate enrichment of treatment vs. control samples
```

```

columns=['chr ', 'enrichment score ']
enrichment_df=pd.DataFrame(columns=columns)
enrichment_df['chr']=tx_df['chrom']
enrichment_df['enrichment score']=tx_df['score']/ctrl_df['score']

enrichment_df.replace([np.inf, -np.inf], np.nan, inplace=True)

enrichment_df['enrichment score']=enrichment_df['enrichment score'].fillna(0)

enrichment_df.columns=['chrom', 'enrichment score ']

enrichment_df=enrichment_df.sort_values(by=['enrichment score'])

# set a numeric placeholder for the y axis
y_range=list(range(1,len(enrichment_df.index)+1))

fig, ax = plt.subplots(figsize=(4,8))
ax.set_facecolor('white')

# option to export data for further processing
enrichment_df.to_csv('./final_summary/'+str(tx)+'_'+str(ctrl)+'.csv')

plt.hlines(y=y_range, xmin=0, xmax=enrichment_df['enrichment score'], color='#2
a6f52', alpha=0.2, linewidth=5)

plt.plot(enrichment_df['enrichment score'], y_range, "o", markersize=5, color='#2
a6f52', alpha=0.6)

# set labels
ax.set_xlabel('enrichment score', fontsize=20, fontweight='black', color = 'black
')
ax.set_ylabel('chromosome', fontsize=20, fontweight='black', color = 'black')
ax.set_xlim([-1,10])
fig.suptitle('Read enrichment at '+tx.split('_')[0]+' vs. '+ctrl.split('_')[0],
            fontsize=30, color='black')

# set axis
ax.tick_params(axis='both', which='major', labelsize=12)

```

```

plt.yticks(y_range, enrichment_df.index)
y_labels = enrichment_df['chrom'].to_list()
ax.set_yticklabels(y_labels)

# change the style of the axis spines
ax.spines['top'].set_color('none')
ax.spines['right'].set_color('none')

# set the spines position
ax.spines['bottom'].set_position(('axes', -0.04))
ax.spines['left'].set_position(('axes', 0.015))

# set a vertical line at x=1 to show no enrichment
plt.axvline(x=1, ymin=0.02, ymax=0.98, dashes=(4,4), color='black', alpha=0.5)

plt.savefig('../chm13_'+str(tx)+'_'+str(ctrl)+'.pdf')
plt.show()

```

6.5 Automated microscopy analysis [Macro, Fiji]

```

// save this file as a .java and run from within Fiji
// this script is written for analysis of two channels, using each channel to
    create its own boxes
// i.e., as is done during cenFISH / telFISH analysis
// modified to use one channel for boxing by changing which channel foci are used

// outputs of this analysis:
// 1. boxes for each focus within each cell, which will then be further processed
// 2. DAPI and all channel intensities per cell in 'dapi_int'

// manual settings required:
// 1. 'prominence' for foci selection, should be defined manually before running
    macro
// 2. working directory to specificity path to files of interest
// 3. box size, once again following manual inspection (can be adjusted later)

// set the directory for image analysis containing only the images you want

```

```
    analysed

dir = ".../"

files = getFileList(dir)

// make folders that will contain the downstream analysis

foci_intensity = dir+"foci_count"+File.separator;
File.makeDirectory(foci_intensity);

foci_coloc_C2 = dir+"C2_foci_coloc"+File.separator;
File.makeDirectory(foci_coloc_C2);
C2_med_focus_per_cell = dir+"C2_med_focus_per_cell"+File.separator;
File.makeDirectory(C2_med_focus_per_cell);
C2_med_foci_images = dir+"C2_med_foci_images"+File.separator;
File.makeDirectory(C2_med_foci_images);

foci_coloc_C3 = dir+"C3_foci_coloc"+File.separator;
File.makeDirectory(foci_coloc_C3);
C3_med_focus_per_cell = dir+"C3_med_focus_per_cell"+File.separator;
File.makeDirectory(C3_med_focus_per_cell);
C3_med_foci_images = dir+"C3_med_foci_images"+File.separator;
File.makeDirectory(C3_med_foci_images);

dapi_int = dir+"dapi_int"+File.separator;
File.makeDirectory(dapi_int);

final_summary = dir+"final_summary"+File.separator;
File.makeDirectory(final_summary);

for(f = 0; f < files.length; f++){
    if (endsWith(files[f], "/" ))
        continue;
    name = files[f];
    if (File.exists(dir+"/foci_count/"+name+"-C2-0.csv")) {
        print(files[f]+" has already been analyzed.");
    }
}
```



```

        continue;
    } else {
        open(dir+files[f], "OK");
        selectWindow(name);
        //
        // if multiple slices were imaged, include the following:
        //run("Z Project...", "projection=[Sum Slices]");
        //
        run("Split Channels");
        //
        // define cell regions of interest (ROIs) to analyze
        //
        selectWindow("C1-"+name);
        run("Duplicate...", "title=dup");
        selectWindow("dup");
        run("8-bit");
        run("Auto Threshold", "method=Li white");
        run("Convert to Mask");
        run("Dilate");
        run("Fill Holes");
        run("Erode");
        //
        // if cells are aggregated, including the following:
        //run("Watershed");
        //
        run("Analyze Particles...", "size=30-Infinity display
            exclude clear summarize add");
        selectWindow("C1-"+name);
        run("Grays");
        //
        // manual inspection to verify that cell segmentation was
        // done properly:
        //
        waitForUser("Choose cells.");
        run("Clear Results");
        //
        // measure number C2 and C3 foci in each cell
        //

```

```

nROIs = roiManager("count");
selectWindow("C2-"+name);
for (x=0; x<nROIs;x++){
    roiManager("Select", x); //newArray(1)?
    run("Find Maxima...", "prominence=400 output=List
    ");
    saveAs("Results", dir+"/foci_count/"+name+"-C2-"+x
    +".csv");
    run("Clear Results");
}
selectWindow("C3-"+name);
for (x=0; x<nROIs;x++){
    roiManager("Select", x); //newArray(1)?
    run("Find Maxima...", "prominence=400 output=List
    ");
    saveAs("Results", dir+"/foci_count/"+name+"-C3-"+x
    +".csv");
    run("Clear Results");
}

//
// export each box surrounding foci within C2 and C3
    channels
// set box size below (i.e., 4, 8, 6, 12, etc.)
// use single channel foci if using box defined with one
    channel
//
for (x=0; x<nROIs;x++){
    selectWindow("C1-"+name);
    run("Clear Results");
    roiManager("Select", x);
    run("Measure");
    saveAs("Results", dir+"/dapi_int/"+name+"-"+x+"
    csv");
    run("Clear Results");
    filestring=File.openAsString(dir+"/foci_count/"+
    name+"-C3-"+x+".csv");
    xlines=split(filestring, "\n");
    for (i=1;i<xlines.length;i++){

```

```

        line=split ( xlines [ i ] , " , " );
        index=line [ 0 ];
        x_coord=parseInt ( line [ 1 ] ) -4;
        y_coord=parseInt ( line [ 2 ] ) -4;
        print ( x_coord , y_coord );
        selectWindow ( " C3—" +name );
        makeRectangle ( x_coord , y_coord , 8 , 8 );
        run ( " Duplicate ..." , " title=" +name + " _" +i + "
            duplicate channels=3" );
        selectWindow ( name + " _" +i );
        saveAs ( " Text Image" , dir + " / C3_foci_coloc
            /" +name + " _C3_" +x + " _" +i + ".txt" );
        close ( );
    }
run ( " Clear Results" );
filestring=File.openAsString ( dir + " / foci_count /" +
    name + " -C2-" +x + ".csv" );
xlines=split ( filestring , " \n" );
for ( i=1; i<xlines.length; i++){
    line=split ( xlines [ i ] , " , " );
    index=line [ 0 ];
    x_coord=parseInt ( line [ 1 ] ) -6;
    y_coord=parseInt ( line [ 2 ] ) -6;
    print ( x_coord , y_coord );
    selectWindow ( " C2—" +name );
    makeRectangle ( x_coord , y_coord , 12 , 12 );
    run ( " Duplicate ..." , " title=" +name + " _" +i + "
        duplicate channels=2" );
    selectWindow ( name + " _" +i );
    saveAs ( " Text Image" , dir + " / C2_foci_coloc
        /" +name + " _C2_" +x + " _" +i + ".txt" );
    close ( );
}
run ( " Clear Results" );
}
roiManager ( " Select All" );
roiManager ( " Delete" );
run ( " Close All" );

```

```

    }
}

```

6.6 Down-stream processing of microscopy data [Python]

```

# analyze all foci boxes generated using the Macro java-based program (section
    2.6.4.3)
# these should be within subfolders i.e. C2-foci-coloc where the C2 denotes the
    boxing channel, and it should contain a matrix for each channel within each
    focus box.
# this script will then generate an 'average' (actually, median) focus for each
    cell and save it under the 'med.focus-per-cell' subfolder.
# importantly, np.nanmedian has to be used instead of np.median to prevent some
    anomalous data points from the image processing.

# import dependencies

import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import os

# set the boxing channel and size
# box trimming can be done by changing the desired_size
channel = 'C2'
size = 12
desired_size=12

# input the following into terminal.
cwd = os.getcwd()

list = os.listdir(cwd+'/'+str(channel)+'_foci_coloc/')
dapi_list=os.listdir(cwd+'/dapi_int/')

for file in os.listdir(cwd):

```

```

if file.endswith( '.oib' ):
    treatment="_".join( file.split( '-' )[:-1] )
    sample=file.split( '.' )[-2]
    # identify number of cells per sample
    dapi_sample_matching = [s for s in dapi_list if s.startswith(
        sample)]
    cell_num = max(int(x.split( '-' )[-1].split( '.' )[-2]) for x in
        dapi_sample_matching)
    #
    sample_matching = [s for s in list if s.startswith(sample)]
    # parse through cells , identifying number of foci and parsing
    through foci
    for cell in range(0,cell_num+1):
        #
        cum_data_C2 = np.empty((size , size , 1))
        cum_data_C3 = np.empty((size , size , 1))
        #
        cell_matching = [c for c in sample_matching if c.
            startswith( file+'_'+str(channel)+'_'+str( cell)+'_' )]
        if len(cell_matching)==0:
            print( 'Assigning ' +str(sample)+' cell ' +str( cell)
                + ' with 0 due to insufficient foci.' )
np.savetxt( './'+str(channel)+'_med_focus_per_cell/'+str(sample)+'_'+str( cell)+'_'+
    str(channel)+'_med_focus.txt ', np.zeros((size , size)))
        continue
    else:
        foci_num = max(int(x.split( '.' )[-2].split( '-' )
            [-1]) for x in cell_matching)
        foci_size_cell=pd.DataFrame(columns=['Sample', '
            Cell', 'Foci', 'C2 Foci Size', 'C3 Foci Size', 'C4
            Foci Size'])
        # parse through all foci and generate heatmaps
        for foci in range(1,foci_num+1):
            data_C3 = np.loadtxt( './'+str(channel)+'
                _foci_coloc/'+file+'_C3_'+str( cell)+"_"+str(
                    foci)+"_foci.txt")
            #f data_C3.shape==(size , size):
                cum_data_C3 = np.dstack((cum_data_C3,

```

```

        data_C3))

    data_C2 = np.loadtxt('./'+str(channel)+'
        _foci_coloc/'+file+'_C2_'+str(cell)+'_'+str(
        foci)+'_'.txt")

    if data_C2.shape==(size , size):
        cum_data_C2 = np.dstack((cum_data_C2,
            data_C2))

    #
np.savetxt('./'+str(channel)+'_med-focus-per-cell/'+str(sample)+'_'+str(cell)+'
    _C2_med-focus.txt ', np.nanmedian(cum_data_C2 , axis=2))
np.savetxt('./'+str(channel)+'_med-focus-per-cell/'+str(sample)+'_'+str(cell)+'
    _C3_med-focus.txt ', np.nanmedian(cum_data_C3 , axis=2))
    else:
        continue

# collect the cell data for each treatment and generate a heatmap for the 'average
    ' (actually , median) focus.
# a background subtraction step will be done on the cellular level.
# set the samples here , images names should begin with these strings:

txs = [ ' ', ' ' ]
tx_list = os.listdir(cwd+'/' +str(channel)+'_med-focus-per-cell/')

def crop_center(array):
    y,x = array.shape
    startx = x//2-(desired_size//2)
    starty = y//2-(desired_size//2)
    return array[starty:starty+desired_size , startx:startx+desired_size]

for tx in txs:
    tx_sample_matching = [s for s in tx_list if s.startswith(tx)]
    tx_num = len(tx_sample_matching)
    #
    data_C2_all = np.empty((desired_size , desired_size , 1))
    data_C2_all_BG_sub = np.empty((desired_size , desired_size , 1))
    #

```

```

data_C3_all = np.empty((desired_size , desired_size , 1))
data_C3_all_BG_sub = np.empty((desired_size , desired_size , 1))
#
foci_size_sample=pd.DataFrame(columns=['Sample ', 'Cell ', 'Foci ', 'C2 Foci
    Size ', 'C3 Foci Size ', 'C4 Foci Size '])
#
for file in tx_sample_matching:
    #
    if file.split('_')[-3]=='C2':
        data_C2 = crop_center(np.loadtxt('./'+str(channel)+'
            _med_focus_per_cell/'+file))
        background_C2 = np.nanmedian([np.nanmedian(data_C2[:1]),np
            .nanmedian(data_C2[-1:]),np.nanmedian(data_C2[:,0]),np
            .nanmedian(data_C2[:, -1])])
        net_data_C2 = data_C2 - background_C2
        #
        data_C2_all = np.dstack((data_C2_all , data_C2))
        data_C2_all_BG_sub = np.dstack((data_C2_all_BG_sub ,
            net_data_C2))
        #
    if file.split('_')[-3]=='C3':
        data_C3 = crop_center(np.loadtxt('./'+str(channel)+'
            _med_focus_per_cell/'+file))
        background_C3 = np.nanmedian([np.nanmedian(data_C3[:1]),np
            .nanmedian(data_C3[-1:]),np.nanmedian(data_C3[:,0]),np
            .nanmedian(data_C3[:, -1])])
        net_data_C3 = data_C3 - background_C3
        #
        data_C3_all = np.dstack((data_C3_all , data_C3))
        data_C3_all_BG_sub = np.dstack((data_C3_all_BG_sub ,
            net_data_C3))
        #
#
# plot each median focus , as well as background-subtracted focus
#
final_C2 = np.nanmedian(data_C2_all , axis=2)
final_C2_BG_sub = np.nanmedian(data_C2_all_BG_sub , axis=2)
#

```

```

sns.heatmap(final_C2 , vmin=0,vmax=1000)
plt.savefig('./'+str(channel)+'_med_foci_images/'+tx+'_C2.pdf')
plt.close()
sns.heatmap(final_C2_BG_sub , vmin=0,vmax=1000)
plt.savefig('./'+str(channel)+'_med_foci_images/'+tx+'_C2_BG_sub.pdf')
plt.close()
#
final_C3 = np.nanmedian(data_C3_all , axis=2)
final_C3_BG_sub = np.nanmedian(data_C3_all_BG_sub , axis=2)
#
sns.heatmap(final_C3 , vmin=0,vmax=2000)
plt.savefig('./'+str(channel)+'_med_foci_images/'+tx+'_C3.pdf')
plt.close()
sns.heatmap(final_C3_BG_sub , vmin=0,vmax=2000)
plt.savefig('./'+str(channel)+'_med_foci_images/'+tx+'_C3_BG_sub.pdf')
plt.close()
#

# finally , collect all the data in a csv spreadsheet for plotting in prism and
    visualizing the cell-cell variation in the average focus per cell , as well as
    the background-subtracted focus per cell .
# of note , the background subtraction is done on a cellular level .

tx_list = os.listdir(cwd+'/' + str(channel) + '_med_focus_per_cell/')
data_all=pd.DataFrame(columns=['Sample' , 'Cell' , 'Channel' , 'Raw Foci Size' , 'BG-sub
    Foci Size'])

for tx in txs:
    tx_sample_matching = [s for s in tx_list if s.startswith(tx) and "MPS1i"
        not in s]
    tx_num = len(tx_sample_matching)
    #
    for file in tx_sample_matching:
        #
        # data collection depending on channel identity .
        #
        if file.split('_')[-3]=='C2':

```



```

data_C2 = crop_center(np.loadtxt('./'+str(channel)+'
    _med_focus_per_cell/'+file))
background_C2 = np.nanmedian([np.nanmedian(data_C2[:1]),np
    .nanmedian(data_C2[-1:]),np.nanmedian(data_C2[:,0]),np
    .nanmedian(data_C2[:, -1])])
final_data_C2 = data_C2 - background_C2
if file.split('_')[-3]=='C3':
    data_C3 = crop_center(np.loadtxt('./'+str(channel)+'
        _med_focus_per_cell/'+file))
    background_C3 = np.nanmedian([np.nanmedian(data_C3[:1]),np
        .nanmedian(data_C3[-1:]),np.nanmedian(data_C3[:,0]),np
        .nanmedian(data_C3[:, -1])])
    final_data_C3 = data_C3 - background_C3
if file.split('_')[-3]=='C4':
    data_C4 = crop_center(np.loadtxt('./'+str(channel)+'
        _med_focus_per_cell/'+file))
    background_C4 = np.nanmedian([np.nanmedian(data_C4[:1]),np
        .nanmedian(data_C4[-1:]),np.nanmedian(data_C4[:,0]),np
        .nanmedian(data_C4[:, -1])])
    final_data_C4 = data_C4 - background_C4

#
# compile data into one table, depending on channel identity.
#
sample_data=pd.DataFrame(columns=['Sample', 'Cell', 'Channel', 'Raw
    Foci Size', 'BG-sub Foci Size'])
#
if file.split('_')[-3]=='C2':
    sample_data=sample_data.append({
        "Sample": tx+'_'+file.split('_')[-5],
        "Cell": file.split('_')[-4],
        "Channel": file.split('_')[-3],
        "Raw Foci Size": sum(sum(data_C2)),
        "BG-sub Foci Size": sum(sum(final_data_C2)),
    }, ignore_index=True)

#
if file.split('_')[-3]=='C3':
    sample_data=sample_data.append({
        "Sample": tx+'_'+file.split('_')[-5],

```

```

        "Cell": file.split('_')[-4],
        "Channel": file.split('_')[-3],
        "Raw Foci Size": sum(sum(data_C3)),
        "BG-sub Foci Size": sum(sum(final_data_C3)),
        }, ignore_index=True)

#
if file.split('_')[-3]=='C4':
    sample_data=sample_data.append({
        "Sample": tx+'_'+file.split('_')[-5],
        "Cell": file.split('_')[-4],
        "Channel": file.split('_')[-3],
        "Raw Foci Size": sum(sum(data_C4)),
        "BG-sub Foci Size": sum(sum(final_data_C4)),
        }, ignore_index=True)

#
data_all = data_all.append(sample_data)

data_all.to_csv('./final_summary/signal_per_cell_'+str(channel)+'.csv')

```