Expansion of human centromeric arrays in cells undergoing break-induced replication

Soyeon Showman

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Steven Henikoff, Chair

Toshio Tsukiyama

Stephen Tapscott

Program authorized to offer degree:

Molecular and Cellular Biology

University of Washington

Abstract

Expansion of human centromeric arrays in cells undergoing break-induced replication

Soyeon Showman

Chair of the Supervisory Committee:

Steven Henikoff

Genome Sciences

Human centromeres are located within highly homogeneous mega-sized α-satellite arrays and evolve rapidly, which can lead to variation in array lengths and sequences. Proposed mechanisms for such alterations are homology directed repair mechanisms including unequal cross-over between sister chromatids, gene conversion, and break-induced replication. However, the underlying molecular mechanisms responsible for the massive, complex, and rapid sequence turn over, length variation, and homogeneous organization of centromeric arrays have not been experimentally validated. This dissertation project investigates whether centromeric array expansion and contraction can occur within limited somatic cell divisions and the molecular mechanisms responsible for this change.

This thesis work has demonstrated that centromeric array length can change in somatic cells (in ~20 cell divisions) of different cell lines, with various magnitudes, in a chromosome specific manner. In addition, centromeric arrays expand more frequently than contract, which can counteract the loss of SSA of DSBs at centromeres leading to an overall increase in array length. Large contractions can occur, but usually only when the array length is significantly longer than the population average. Finally, this array length change does not occur without the BIR essential proteins RAD52 and PIF1, indicating that BIR can drive centromere sequence evolution in cells undergoing BIR. This project provided key insights into a longstanding fundamental question: how centromere sequences evolve.

ACKNOWLEDGEMENTS

Science cannot be performed in a vacuum. Therefore, I have so many people who have guided me to be where I am now. The most important individual in my scientific career is Steve Henikoff, my current PI and role model. I learned perseverance, resilience, excitement with the unknown, and to not be afraid of innovation from Steve. I cannot wait to apply all the lessons that I learned from Steve to my next position. I truly appreciate all the help that my committee members provided for me. (Toshio Tsukiyama, Stephen Tapscott, Brian Beliveau, and Andrew Stergachis). I also want to thank my collaborator Richard Adeyemi for mentoring me through my project.

I would like to thank my mentors and lab members in the Henikoff lab along with my previous lab mates for their support throughout my scientific careers. In particular, my good friend and former lab mate Renee who checked my manuscripts, fellowship applications, and even read my thesis. I appreciate her kind friendship and support.

I especially want to thank my husband Jamie, daughter Sophia, brother Dongkyu, sister-in-law Myeongsung, sister Heather, and my mother-in-law Melanie.

Of all the people, my brother Dongkyu understands me the best. I would not be able to do this without his and my sister-in-law's support. I started my undergraduate when Sophia was 8 months old. Now she is 12. I would not be able to even dream about getting a PhD if she was not as mature as she is. She is my friend, cheers for me, and even shares her precocious wisdom when I struggle. I appreciate that she motivates me to be a good human and scientist. The person who should own the letter 'H' from my PhD degree is my husband, Jamie. Everyone including even the security officer calls him my personal taxi driver because Jamie dropped me off and picked me up at work every day throughout my entire my PhD. Jamie has helped me with everything including editing my manuscripts, fellowships, and thesis. He might be able to answers some of the questions related to centromeres at this point. He relocated his job to Seattle, took care of most of the parental duties and house chores so that I can truly focus on my academics. I would not be able to have a PhD without his support and sacrifice.

I especially want to thank my mother-in-law, Melanie, who has passed away last week. I started my academics in the US at a community college in New Mexico. My mother-in-law was the one who promoted me to pursue a higher degree and trusted me more than I did myself. She treated me just like her own kid. I appreciate mom who told me that I've made her proud in her last message to me. I dedicate my thesis to her.

Table of Contents

Chapter 1: Introduction

For two decades since the Human Genome Project was initiated, highly repetitive heterochromatic regions such as satellites have prevented complete genome assembly and made research into centromere biology challenging[1]. However, in 2022, the Telomere-to-Telomere (T2T) Consortium created a gapless human genome assembly for all chromosomes which finally provided the basis for studying rapid centromere sequence evolution[2]. Follow up comparative and evolutionary studies confirmed that the centromere sequence evolves rapidly as a result of molecular evolution. Chromosome specific diversifications as a result of molecular evolution include extreme length, homogenization, variation in sequence, complex structure, and organization[3-7]. However, the underlying mechanisms that cause these genetic changes that drive centromere sequence evolution have not been experimentally tested.

1.1 Molecular mechanisms for the evolution of centromeric α-satellite arrays

Centromeres are crucial genomic regions that accurately separate genetic materials into two daughter cells during mitosis and meiosis[8]. Human centromeres are located within α-satellite arrays that are composed of multi-monomeric higher-order repeat units (HORs) of 171 base-pair (bp) AT-rich head-to-tail tandemly oriented monomers expanding from 30 kb up to 8 Mb[9-12]. While centromere function is conserved across species, centromere sequences evolve rapidly and play an important role in genomic and karyotypic evolution[13-16]. Although the active centromere is epigenetically defined by the enrichment of the centromere-specific histone H3 variant CENP-A, the assembly of the kinetochore occurs at the core of the active centromere which contains evolutionarily young and homogeneous

HORs, suggesting the influence of genetic components in centromere specification[3,8,17]. The genetic variations between species, populations, individuals, and even two homologous chromosomes are chromosome specific and include divergence in length, high sequence turnover, HOR structure and composition[3,4,6,7,11,18]. The proposed underlying evolutionary molecular mechanisms that are responsible for these variations are unequal crossover[19], gene conversion[20], and break-induced replication (BIR)[21,22]. The most commonly cited mechanism is based on the unequal crossover of tandem repeats between sister chromatids that are misaligned during HR[19]. While this process is a neutral evolution model that should produce random sequence arrays, centromeric arrays are composed of highly structured HORs[21]. In addition, the unequal crossover mechanism was proposed before the discovery of single stranded annealing (SSA)[23], which is the predominant mechanism used in the repair of repetitive regions[24]. SSA causes the deletion of one HOR unit[25] and, thus, gradually shrinks the centromere over time suggesting that there are other mechanisms that counteract this loss and, in turn, leads to the extremely large α-satellite arrays that have been observed in human populations[2,3,5,11,26]. An alternative mechanism proposed is based on BIR, which is the main repair mechanism of one-sided double stranded DNA breaks (DSB) that are generated from fork collapses during replication[21,22]. BIR is a highly mutagenetic repair mechanism that favors the expansion of array[21,27]. Centromeric arrays that are enriched with various replication obstacles provide a high level of BIR substrates[28-31]. The possible outcomes of BIR match with observed human centromeres include mega-base size centromeric arrays due to its bias toward expansion, sequence turn-over due to a high substitution mutation rate, and complex organization in a

replication-dependent manner[21,32,33]. However, these mechanisms had not been experimentally validated due to technical challenges in researching variations in centromeres that occur outside of the evolutionary time scale.

1.2 Aims

The goal of my dissertation project is to test whether BIR acts as an evolutionary molecular process that led to centromere sequence evolution. To do this, I first generated and optimized a method to quantify HOR copy number variations that occur during clonal evolution. I confirmed whether centromeric array changes can occur within somatic cell divisions and then determined the rate of change that can serve as a basis to test the BIR mechanism. Second, I characterized array changes, such as expansion bias. Finally, I validated the BIR mechanism as a driver for centromere sequence evolution using BIR inactive mutants.

1.3 Dissertation overview

Chapter 2 offers an overview of centromere biology and α-satellite arrays including structure, organization, specialized replication and repair programs that are linked to evolutionary processes.

Chapter 3 describes a protocol to measure array changes throughout clonal evolution among subclones that are derived from a monoclonal cell line with minimal pre-existing genetic background using the ddPCR-based assay. This method allows us to accurately measure the array changes that occur within somatic cell divisions in a cell line with a

sensitized background despite the intrinsically high rate of technical error that is associated with the ddPCR.

Chapter 4 demonstrates that HOR arrays can expand and contract within ~20 somatic divisions and favor expansion over contraction using the method described in Chapter 3. The array alteration only occurs in the presence of RAD52 and PIF1, that are essential for BIR, validating BIR as a molecular driver that leads to rapid centromere sequence evolution.

Chapter 5 discusses the relevance of this work to evolution and proposes future studies that could provide more insight into the evolutionary molecular processes that are mediated by the repair of satellite arrays. These include elucidating the prevalence of BIR substrate in centromeric arrays, length homeostasis, and the timing of BIR during the cell cycle.

Chapter 2: Overview of centromere biology

2.1 Cell cycle

In both prokaryotes and eukaryotes, a highly complex network of cell cycle regulations ensures the faithful transmission of genetic materials from mother cells to daughter cells by mediating the progress of genomic DNA replication and its segregation in a timely manner[34,35]. The cell cycle has four sequential phases: G_1 (gap phase 1), S (synthesis phase), G_2 (gap phase 2), and M (mitosis). The main events in the cell cycle are the S phase, in which timely DNA replication occurs at a high fidelity, and the M phase, in which DNA is segregated into two genetically identical daughter cells. These two phases are separated by gap phases G_1 and G_2 that serve as checkpoints in cell cycle progression[35,36].

Most human cells do not cycle and stay in either a terminally differentiated state, such neurons and myocytes, or a quiescent state (G_0), which has entered a reversible non-proliferate state by exiting the cell cycle[37-39]. The cyclin-dependent kinase (CDK) family, whose kinase activity is dependent on their regulatory partner cyclins, promotes G_0 cells to restart the cell cycle[40,41]. During each cell phase, a specific subset of cyclins is accumulated or depleted based on cell cycle-regulated transcription and proteolytic degradation[35,42,43]. The phosphorylation of cyclins by CDK is a critical step needed for cells to pass the multiple checkpoints that are either within or border the cell phases[44].

The cells that have passed the commitment point of G_1 can initiate DNA replication in the S-phase, whose fidelity of action is strictly controlled by the intra-S checkpoint and DNA damage response[45]. Upon completion of the replication, the postreplicative G_2 phase creates another decision point for the cell to enter mitosis, depending on CDK1 activity[46].

The M phase is further divided into five sub-phases based on the physical state of the chromosomes and spindle: prophase (chromosome condensation), metaphase (chromosome alignment), anaphase (chromatid separation), telophase (re-formation of the nuclear membrane and chromosome decondensation), and cytokinesis (cellular separation)[47]. Entry into mitosis is initiated with a wave of mitotic phosphorylation activated by CDK1, including the mitotic kinases PLK1, Aurora A, and Aurora B. This massive phosphorylation leads to structural changes in the cell, such as mitotic cell rounding and chromosome condensation, that are well known characteristics of mitotic cells and necessary for mitotic progression[48,49]. In the prometaphase, the disintegration of the nuclear membrane permits the binding of microtubules to kinetochores that are assembled at a special genomic locus of each siter chromatid, leading to chromosome alignment in the middle of the bipolar spindle in the metaphase[50,51]. The sister chromatids are separated and pulled to opposite poles by spindle microtubules, and finally the cell divides into two daughter cells. The kinetochore, a multiprotein complex that is assembled in the centromere, is critical for accurate cell division during mitosis and meiosis[52].

2.2 Centromere function and specification

Centromeres are essential chromosomal loci needed for cell division and defects in the region can cause chromosome mis-segregation, aneuploidy, and structural abnormalities that are a hallmark of cancer and developmental diseases[8,53]. The centromeres cytologically appear as a primary constriction on condensed chromosomes where two sister chromatids, one replicated from the other, are joined[54]. While the molecular features

that identify centromeres are complex and diverse across eukaryotes, the centromeres of monocentric chromosomes in eukaryotes are often established at highly repetitive DNA sequences such as satellite DNAs and transposons, with a few exceptions such as the point centromeres of *Saccharomyces cerevisiae,* known as budding yeast[55,56]. Point centromeres are established within a specific centromere DNA sequence and are able to undergo chromosome segregation with functionally assembled kinetochores[57]. However, most plant and animal centromeres are found within satellite arrays that vary dramatically in length and sequence[56,58]. Despite differences in the function of centromeric DNA sequences among species, a common characteristic among them is that the centromeres position and function are epigenetically defined by the enrichment of centromere specific histone H3 variant centromeric protein A (CENP-A)[3,59,60].

CENP-A, a key epigenetic marker, is required for the formation of centromeres and neocentromeres (that are established at different sites without the canonical repetitive sequences found in most organisms)[61,62]. In addition, it is a docking site for components of the kinetochore[52]. Barnhart et al., found that targeting the Holliday Junction Recognition Protein (HJURP), which is the histone chaperone that deposits CENP-A at centromeres in late mitosis and early G_1, to a non-centromeric locus is sufficient to recruit the constitutive centromere-associated network (CCAN), a subcomplex of the kinetochore, and assemble the kinetochore, highlighting the vital role of CENP-A for centromere identification[63,64].

2.3 Alpha satellite array structure and organization

Human centromeres are located within α-satellite arrays that represent ~6% of the human genome and are composed of multi-monomeric higher-order repeat units (HORs) of 171 base-pair (bp) AT-rich head-to-tail tandemly oriented monomers (Figure 2.1a)[3,9,65,66]. Homogenous HORs repeat continuously, without disruption, hundreds to thousands of times to create a mega-base sized array with nearly identical HOR tandem copies[2,3,67]. This extreme repetitiveness of the α-satellite array had excluded centromere sequences from the human genome reference for decades[2,68,69]. While several human chromosomes contain multiple HOR arrays on the same chromosome, only one array is likely active[9,70]. The active centromere, defined by the enrichment of CENP-A, contains more homogenous and evolutionally young HOR sequences[3,5,6,70]. Unlike HORs in the active centromere, peri-centromeres are more divergent α-satellite HORs that flank the active centromere, are less oriented, but still maintain their repetitiveness with the DNA monomers that are intermixed with transposable elements[3,17,71-73]. This centromere structure is shared across many taxa including plants, mice, and primates[74]. HOR sequences among HORs within the same chromosomal array or different chromosomal arrays are almost identical, but polymorphisms among monomers within a HOR is (50-70%)[9]. The composition of the HOR is chromosome-specific ranging from 2 to 34 monomers[16]. HORs that share similar sequences and a linear monomer order but a distinctive total number of monomers can be grouped into suprachromosomal families (SFs); including five major (SF1-5) and 15 minor (SF6-18, SF01 and SF02) groups (Figure 2.1b-e)[9,39,75].

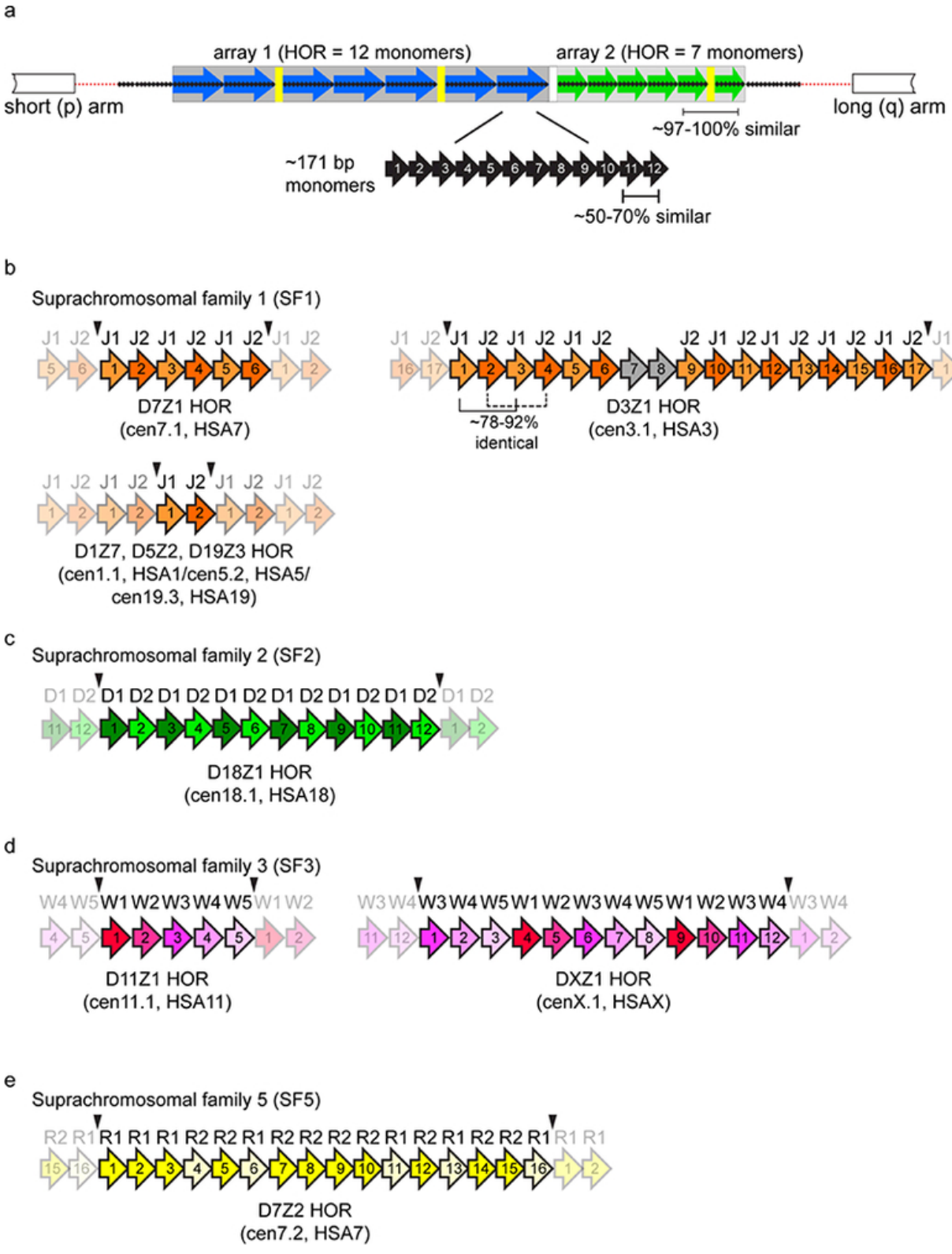

a
array 1 (HOR = 12 monomers)
array 2 (HOR = 7 monomers)
short (p) arm
long (q) arm
~97-100% similar
~171 bp monomers
~50-70% similar
b
Suprachromosomal family 1 (SF1)
D7Z1 HOR (cen7.1, HSA7)
~78-92% identical
D3Z1 HOR (cen3.1, HSA3)
D1Z7, D5Z2, D19Z3 HOR (cen1.1, HSA1/cen5.2, HSA5/ cen19.3, HSA19)
c
Suprachromosomal family 2 (SF2)
D18Z1 HOR (cen18.1, HSA18)
d
Suprachromosomal family 3 (SF3)
D11Z1 HOR (cen11.1, HSA11)
DXZ1 HOR (cenX.1, HSAX)
e
Suprachromosomal family 5 (SF5)
D7Z2 HOR (cen7.2, HSA7)

Figure 1. 1: Examples of human alpha satellite array organization, the five major suprachromosomal family structures, and their linear orders. a) Characteristics of human alpha satellite arrays including HOR structures (black arrow: monomers, blue arrow :12mer, green arrow: 7mer, yellow bar: transposable element), linear organizations (in grey boxes), and sequence similarities. b-e) Structure, sequence identity, and linear order of the five major suprachromosomal families (SFs) and examples of chromosome specific arrays for each group.

The chromosomes in the same family share similar HOR configurations using the 12 consensus monomers: J1, J2, D1, D2, W1, W2, W3, W4, W5, M1, R1, and R2. The majority of functional HORs that are found at active centromeres fall into one of three new or modern families (SF1-3) made of either two dimeric (J1-J2 and D1-D2) or one pentameric monomer organization (W1-5). The divergent monomers that flank the active centromere fall into SF4 (M1) and SF5 (R1-R2) lack higher-order periodicity[17,76,77]. For example, HORs in the SF1 family, which includes nine human chromosomes (1,3,5,6,7,10,12,16, and 19), have an alternating J1-J2 monomer organization with a chromosome specific total repeat number. While the canonical HOR found in the chromosome 1 specific array D1Z7 (chromosome number followed by D) is created with perfect J1-J2 dimers (340bp), the canonical HOR in D7Z1 which is 6-mer (1020bp) is constituted of a J1-J2- J1-J2- J1-J2 monomer configuration that shares the same sequence as D1Z7 but a different length of HOR unit. The canonical DXZ1 HOR, which is 12-mer (2040bp), is made of two W1-W5 pentameric monomers plus a duplication of the W3 and W4 monomers[9]. The difference in the number of monomers, the configuration comprising HOR units, and the presence of

structural HOR variants create chromosome specific α-satellite arrays even though some chromosomes share the same HORs[9]. Every chromosome-specific array contains a different number of structural HOR variants that are present along with the canonical HOR of the same array[3,5].

All α-satellite monomers can be broadly divided into two types of monomers based on the presence of a binding site for either CENP-B, the only known centromere sequence specific binding protein, or pJα whose function is not well understood[74]. The CENP-B box, a 17-bp sequence motif (5'-T/CTCGTTGGAAA/GCGGGA-3') is found in all human centromeres with a unique distribution to each chromosome except the Y chromosome which lacks a CENP-B box motif and belongs to the divergent SF4 family[78,79]. Because CENP-B is found in both active and inactive centromeres and CENP-B knock-out mice remained viable, its role in the maintenance and regulation of centromere chromatin was underestimated[9,74,80]. Indeed, CENP-B is important in CENP-A *de novo* deposition, the positioning of CENP-A containing nucleosomes, and the stabilization of CENP-A and CENP-C within centromeric chromatin[81,82]. The Y chromosome and neocentromere lack CENP-B boxes but enrich non-B DNA and enable the assembly of the kinetochore, suggesting that alternative mechanisms are present at different α-satellite arrays and highlighting the plasticity in centromere organization and kinetochore function[81].

2.4 Centromere heterogeneity in length, sequence, and organization

Centromeres evolve rapidly which leads to extreme divergence in sequence, array length, and organization among different species (species-specific differences), individuals, and even between two homologous chromosomes[3,5,76]. Logsdon et al., assembled a second human genome using the complete hydatidiform mole (CHM)1 cell line which has two nearly identical homologous chromosomes and compared the centromere sequence to the CHM13 genome known as T2T assembly[2]. The sequence homology between two haplotypes is only 63-71.5% (>90% identity). This rate further decreases to ~53-55% when the centromere sequences of 56 diverse human genomes from the Human Pangenome Reference Consortium (HPRC) and Human Genome Structural Variation Consortium (HGSVC) were compared. In addition, neither of the assemblies are a better match for the diverse human genome samples, indicating substantial variation in α-satellite sequences and new HOR structure occurrences among centromeric arrays of diverse individuals[5,18].

Chromosome specific variation manifests as different forms of centromere diversity such as sequence, length, and epigenetic differences due to different evolutionary trajectories[3,5]. While the chromomere X α-satellite array shows the highest sequence and structural conservation, chromosome 5 shows the most divergence in sequence, and chromosome 4 shows the greatest length variation among human haplotypes[5,6]. The α-satellite array length comparison among diverse human genomes shows a great degree of variation across all centromeric arrays, ranging from 0.03 Mb on chromosome 4 to 6.5 Mb

on chromosome 11, mostly due to differences in HOR array organization and size. As an example of an array expansion that originated from the variation in HOR organization, a novel 6-mer layer instead of the 5-mer canonical HOR array of the D11Z1 in CHM13 arose within the active centromere of chromosome 11 in the CHM1, resulting in a 1.2 mb expansion in array size[5].

Comparative analysis of ape centromeric arrays from multiple studies have shown species-specific divergence in HOR organization and structure but all share evidence of layered expansion (Figure 2.2)[3,5,6].

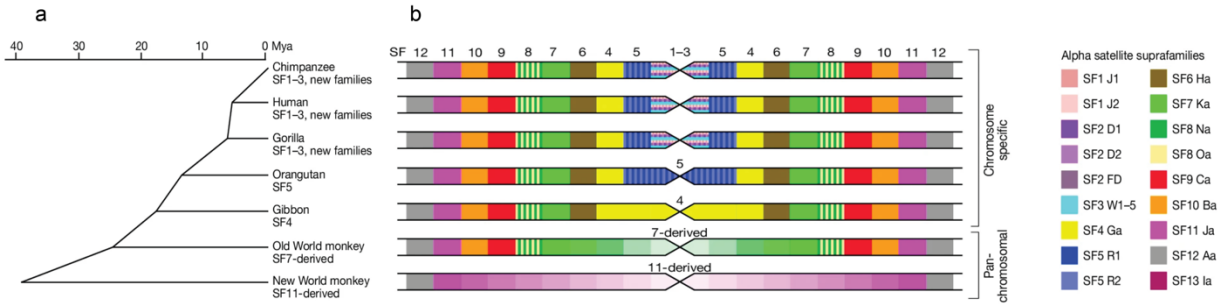


Figure 2.2: a) Phylogenetic tree of suprachromasomal families (SFs) at active centromeric regions among primate which shows the differences in SF organization between chromosome specific and pan-chromosomal groups. b) Linear configuration of SFs on centromeres of primates (match with Figure 2.2a). The active centromeric region is located at the center of the chromosome and is flanked by older pre-existing layers.

All primate centromeres contain new evolutionary symmetrical layers of highly homogenous HORs that emerged from the active centromere and expanded within the pre-exiting arrays. The older displaced arrays in the flanks quickly became inactive and divergent over time due to the accumulation of mutations[2,3,5,6]. The age of the α-satellite

arrays can be estimated by the increase in intra-array divergence from the core to the periphery of the array[2,3,5,6]. While the HOR structure of active centromeres of all chromosomes of apes that evolutionarily predate human lineage are more uniform and less than dimer length (pan-chromosomal HORs), chromosome specific HORs emerged in great apes, with the exception of the Y chromosome[3,6]. For example, while the younger SF group SF1-3 are present in all active centromeric arrays of Humans, Chimpanzees, and Gorillas, the active centromeres in Old and New World Monkeys are composed of the older SF groups that have been pushed into the periphery of the centromere in African Apes (Figure 2.2)[6]. Some of the non-homologous chromosomes such as 1/5/19, 13/21, and 14/22 share nearly identical or very similar HORs[9]. It is still enigmatic whether these arrays did not have a sufficient amount of time to diverge or are still undergoing a nonstop homogenization process between these chromosomal arrays.

The underlying molecular mechanisms that drive this extreme sequence diversification, homogeneity, and length variation at centromeres across species have not been fully elucidated. The proposed mechanisms for this diversification are unequal crossover[19], gene conversion[20], and break-induced replication (BIR)[21,22]. These molecular mechanisms are subsets of Homology Directed Repair (HDR) which underscores the role of DNA breaks to drive centromere sequence evolution via repair-dependent recombination events[19,20,83].

2.5 Centromeres: difficult-to replicate regions

Despite the essentiality of centromere function for cell viability, centromeres are recombination hotspots and are heavily enriched with DNA breaks in comparison with

other repeats such as telomeres in physiological somatic cells and even in quiescent cells[84]. DNA replication in centromeres is challenging due to the presence of complex secondary structures such as single-stranded DNA, hairpins, cruciforms, and DNA: RNA hybrid R-loops that act as physical replication barriers[85-88]. Centromeres that are replicated in the mid-late S-phase in mammalian cells have a slower replication fork speed compared to the bulk genome[29]. While most genomic regions only utilize a fraction of their replication origins during the S-phase and leave the unused origins as backups in case complications arise during replication, in the centromere, most origins are actively used and therefore the region is more susceptible to misregulation[29]. One of the mechanisms that ensures frequent origin firing and timely completion of DNA replication at the centromere during S-phase is mediated by the formation of replication and condensin-dependent double stranded DNA loops[89]. These loops, that are positively supercoiled and accumulate behind the replication fork, inhibit the binding of replication protein A (RPA) to replication forks and, in turn, limit the activation of the ATM and Rad3-related (ATR) replication checkpoint kinase that suppresses origin firing at the centromere[90]. Both CENP-A and CENP-B prevent replication stress (RS) at the centromere[28].

CENP-A safeguards α-satellites from recombination and RS by preventing the formation of R-loops that are generated due to replication-transcription conflicts[91]. Despite their low level of expression, α-satellites that contain transcriptionally permissive markers H3K4me2 and H3K36me2 are actively transcribed by RNA polymerase II[3,92,93]. Depletion of centromeric transcripts reduces CENP-A deposition, indicating that it has a beneficial role in centromere identification at the expense of replication dynamic alteration[94]. CENP-B

regulates the replication dynamic at the centromere by maintaining slow replication forks via CENP-B-mediated loops and by influencing origin selection via nucleosome positioning or heterochromatin formation[95]. Multiple origin firings were observed in the centromere which indicates a relatively high degree of converging replication forks that can induce replication-dependent topological stress which needs to be resolved by topoisomerases[29,96]. The depletion of Topoisomerase II in vertebrates causes replication fork stalling[97]. These results highlight the importance in maintaining correct fork speed and frequency of origin firing to avoid RS which can cause centromere under-replication and ultimately centromere breaks[29]. The mechanisms that regulate replication dynamics within different chromosome specific α-satellite arrays that have varied HOR copy numbers and sequences in turn influences the level of centromere protein recruitment such as CENP-A and CENP-B are not well understood.

2.6 DNA damage responses at centromeres

Centromeres are enriched with various DNA damage response proteins including such as RAD51 (HR), MRE11/RAD50 (HR), Ku80 (NHEJ), PARP1 (base excision repair), XRCC6, and the mismatch repair MSH2-6 complex throughout cell cycles without DNA damage, indicating that a strong surveillance system is acting on these special regions to minimize harm[89,98]. DNA damage response (DDR) factors, whose recruitment is spatiotemporally regulated, senses different types of lesions such single stranded DNA break (SSB) or double stranded DNA break (DSB) and initiates specific DNA repair pathways to physically remove the damage[99].

Although it is not clear what level of *de novo* DNA breaks are present in the centromere in G_1, DNA breaks can be induced by the formation of R-loops and secondary structures of displaced ssDNA that can arise from the transcription-associated torsional stress[100]. The pathway decision for DSB is made based on competition between 5'–3' end resection and lesion repair[30,99]. NHEJ is the main mechanism used to join the two broken ends when end resection is limited[101]. When end resection occurs, the error-free conservative HR, Single-Strand Annealing (SSA) and Alt-EJ further compete[102,103]. HR utilizes the homology in sister chromatids DNA templates for DNA synthesis by strand invasion and contains two sub-pathways: synthesis-dependent strand annealing (SDSA) and break-induced repair (BIR)[104]. SSA and Alt-EJ employ annealing at either long interspersed nucleotide repeats or microhomologies respectively[18]. The genetic outcomes of each mechanism differ and can lead to various mutations including LOH, deletions, and insertions[105]. NHEJ is the main DSB repair mechanism in the G_1 phase because HR is predominant in S/G_2 when homologous sister chromatid templates are available[99,102]. However, exogenously induced DSBs in the centromeric DNA, created by either ionizing radiation or Cas9, were repaired without NHEJ, suggesting HR is active by intra-strand invasion of α-satellite repeats in the G_1 phase[106,107]. Yilmaz et al. investigated the mechanism of H3K4me2 and R-loops mediated RAD51 recruitment at Cas9 induced DSB in centromeres and found that Ubiquitin Specific Peptidase 11 (USP11) facilitates CENP-A deposition by HJURP deubiquitylation. This damage dependent CENP-A deposition enhances USP11 recruitment to resected DNA ends that are promoted by the H3K4me2 dependent R-loop and in turn permits the recruitment of HR factors such as BRCA2–RAD51 to the resected DNA[107]. Centromeric

DNAs are prone to form secondary structures such as cruciforms that can also be found in displaced ssDNA that are associated with R-loops. Dyad symmetry enrichment was observed in centromeres of organisms that lack CENP-B, which can form a secondary structure through its dimerization[86,108]. Kasinathan et al., proposed that HJURP, which binds to a four way double-stranded DNA Holliday junction, might recognize cruciform structures in the centromere and deposit CENP-A[18]. While repair-dependent recombination mediated by secondary structures has beneficial roles in CENP-A deposition in the G_1 phase, it can cause deleterious effects during the S phase unless it is repaired properly[86].

Homology directed repair mechanisms are used in the S/G_2 phases to repair DSBs and, therefore, DSBs that are enriched in the centromere would be repaired by various pathways, influencing α-satellite array sizes and sequences differently[84]. HR can lead to the expansion or contraction of arrays due to gene conversion and unequal crossover between sister chromatids that are misaligned[105,109]. SSA leads to contraction due to the deletion of one homologous tandem repeat sequence and the non-homologous sequence between repeats[25,109]. Secondary structures in the centromere are obstacles that can prevent the duplication of the genome on time[110]. Nuclease/helicase DNA2 is a structure-specific nuclease that is mainly found in the centromere and its ability to resolve secondary structures is important for the completion of centromeric DNA replication, CENP-A deposition, and the prevention of mitotic defects[111]. When secondary structures persist, replication forks stall and eventually collapse, resulting in one-sided DSB unless the re-initiation of the stalling forks occurs[112]. The main repair response to collapsed forks is BIR

which is a mutagenic and low-fidelity repair mechanism that can lead to an extensive array size gain and substantial substitution mutations[27,113].

Common Fragile Sites (CFSs) are other difficult-to-replicate genomic regions that are prone to breaks and sensitive to RS induced by oncogene activation or replication perturbation[114,115]. CFSs are reminiscent of α-satellite arrays in many aspects because they are replicated in the late S-phase and contain scarce replication. These regions contain under-replicated loci that can be seen as gaps in mitotic chromosomes upon RS and are associated with very large genes that can form R-loops that originated from transcription-replication conflicts (TRC) and contain AT rich microsatellite repeats that are known to form DNA secondary structures[115-119]. The completion of cell division is a higher priority than DNA damage repair in mitotic cells. Therefore, DDRs that promote CDK1 mediated mitotic arrest are suppressed[120]. DSB repair, known as mitotic DNA synthesis (MiDAS), can still occur via BIR during early mitosis to remove persistent pre-mitotic errors such as the under-replicated regions mentioned and prevent chromosome mis-segregation[39,88,121-123].

2.7 DNA repair mediated mechanisms that drive rapid centromere sequence evolution

Molecular mechanisms that explain the complex organization, homogeneity, and variation in array length and sequence are mediated by homology-directed repair (HDR)[21,22]. The mechanism which is cited the most by the field is based on repeated unequal crossover events, one of the genetic outcomes of HDR[3,76]. The tandem repeats between sister chromatids that are misaligned undergo unequal crossover during HR resulting in array

expansion and homogenization[19]. Based on this process, sequences in α-satellite arrays should be random without any of the structural patterns that can arise from functional constraints on repeat sequences. Yet, HORs have multiple distinctive structures[3,21]. In addition, this model was proposed before the discovery of SSA[23], which is the predominant mechanism for the repair of repetitive regions[24]. SSA causes the deletion of one HOR unit[25] and, thus, gradually shrink the centromere over time unless there is a counteracting mechanism to compensate this loss and, thus, led to extremely large α-satellite arrays that have been observed in human populations[2,3,5,11,26]. An alternative mechanism proposed is based on BIR, which is the main repair mechanism of one-sided DSBs that are generated from fork collapses following fork stalling during replication. Tandem repeat length variation, which favors expansion, occurs via repairs of one-sided DSBs using a DNA template from an out-of-register sister chromatid[32]. The outcomes of BIR repair of one-sided DSBs occurring in centromeric arrays are replication dependent length variations (higher length variation observed in the Y chromosome compared to the X due to a different numbers of cell divisions), compensation of array loss caused by SSA because expansion is favored over contraction, and elevated substitutions associated with mutagenetic BIR[21,32,33]. BIR is best known as a telomere maintenance mechanism in alternative lengthening of telomere (ALT) positive cancers such as U2OS and MiDAS to repair CSFs in mammalian cells in difficult to replicate regions of the genome associated with high incidence of one-sided DSBs[123,124]. Although BIR can explain many characteristics of α-satellite arrays that are observed at human centromeres, this mechanism has not been experimentally validated due to technical challenges in studying centromere biology.

Chapter 3: Protocol to measure centromeric array size changes using droplet digital PCR-based quantification of higher-order repeats

Modified from an article originally published in *STAR Protocols*:

Soyeon Showman, Paul B. Talbert, Yiling Xu, and Steven Henikoff.

3.1 Abstract

Centromere length changes occurring during somatic cell divisions can be estimated by quantifying the copy numbers (CNs) of higher-order repeats (HORs), which are nested repeats of monomers that comprise centromeric arrays. Here, we present a protocol for single-cell isolation for clonal evolution followed by droplet digital PCR-based quantification. The assay measures HOR CNs across subclones to determine the frequency and degree of changes in HOR CNs. This protocol tests the underlying molecular mechanisms responsible for rapid centromere sequence evolution.

For complete details on the use and execution of this protocol, please refer to Showman et al[125].

3.2 Protocols

Before you begin

This protocol aims to quantify CNs of HORs in α-satellite arrays that can change during somatic cell divisions. The protocol below describes the specific steps for measuring D11Z1 CN in human U2OS cells, but we have also applied it to different HORs and other cell lines including human K562 cells with modification during monoclonal cell isolation. This protocol takes 10-12 h of hands-on time spread across 8 weeks to allow for clonal evolution. We recommend using cells that have undergone few sub-cultures to ensure consistent cell growth throughout clonal evolution. This protocol can be used to quantify

other repetitive elements, yet optimizations including primer efficiency, genomic DNA (gDNA) input, dilution factor, and annealing temperature are likely needed. We used the human CHM13 cell line for optimization and as a positive control[2], yet this can be customized for the tandem repetitive array of interest. This protocol does not provide primer design steps because each repetitive array needs extensive troubleshooting to determine whether the designed primers efficiently target all repeats. All primer sequences that are compatible with this protocol are from de Lima et al., (2021) and are listed in Table 1[18].

Table 1. Target primer sequences and restriction enzymes

Primer Sequence	Target (HOR or Single Gene)	Restriction Enzyme
D6Z1 F: 5' – GCGTTGAACTCACCGTCTT – 3'	HOR (Chromosome 6)	Alu I
D6Z1 R: 5' – TCCAAAGAATGCCTCCAAGG – 3'	HOR (Chromosome 6)	Alu I
TBP1 F: 5' – GATATGAGACTGTGGGTAAGT – 3'	SG (Chromosome 6)	HaeIII
TBP1 R: 5' – GATCCTTTGAACACCCTAATG – 3'	SG (Chromosome 6)	HaeIII
D11Z1 F: 5' – CTTCCTTCGAAACGGGTATATCT – 3'	HOR (Chromosome 11)	Alu I
D11Z1 R: 5' – GCTCCATCAGCAGGATTGT – 3'	HOR (Chromosome 11)	Alu I
C11orf16 F: 5' – TCCCTGACCATCTGGAAGAA – 3'	SG (Chromosome 11)	Alu I
C11orf16 R: 5' – TGATTGGCCCTAGCAGAGA – 3'	SG (Chromosome 11)	Alu I
D18Z1 F: 5' – TGGGAAACGGGATTGTCTTC – 3'	HOR (Chromosome 18)	Alu I
D18Z1 R: 5' – CTGCTCTACCAAAGGGAATGT – 3'	HOR (Chromosome 18)	Alu I
MRO F: 5' – TAGTAGGTAACACCGAGTGC – 3'	SG (Chromosome 18)	Alu I
MRO R: 5' – TCAGGGTTGTCGCAAGTA – 3'	SG (Chromosome 18)	Alu I
DXZ1 F: 5' – TGATAGCGCAGCTTTGACAC – 3'	HOR (Chromosome X)	HaeIII
DXZ1 R: 5' – TTCCAACACAGTCCTCCA – 3'	HOR (Chromosome X)	HaeIII
HPRT1 F: 5' – AAGGTGCTGGTCTCCTTTAC – 3'	SG (Chromosome X)	Alu I
HPRT1 R: 5' – GCACCAATGATTCTCTCCCT – 3'	SG (Chromosome X)	Alu I

Monoclonal cell line preparation

Timing: 4 weeks

1. Culture population of U2OS cells and collect conditioned media.
    a. Thaw a new early passage stock of frozen U2OS cells and passage the cells at least two times.
    b. Seed 1.5×10^6 U2OS cells in a T-75 flask with 15 mL of media.
    c. Incubate the cells at 37°C for 2-3 day until 80% confluence.
    d. Remove old (conditioned) media from the flask and place in a 50 mL conical tube.
    e. Wash cells in the flask once with 1X PBS.
    f. Trypsinize cells in 3 mL of 0.25% trypsin-EDTA and incubate at 37°C for 5 min.
    g. Add 7 mL of media and break up all clumps of cells into individual cells by pipetting up and down.

    Optional: Although conditioned media may be used, avoid using it from cells that are overly confluent due to the potential presence of excess cellular waste.

2. Isolate single cells from the population of U2OS cells by limiting dilution[126,127].
    a. Measure the cell number of the homogenized cells with a Vi-cell analyzer.
    b. Add 10^6 cells/mL in a 1.5 mL microcentrifuge tube. (refer to Preparation of HOR CN control)
    c. Dilute the cells to 10^5 cells/mL in a 1.5 mL microcentrifuge tube containing media and homogenize by pipetting.

d. Measure the cell number of the 10^5 cells/mL stock.

e. Dilute the 10^5 cells/mL stock solution in 1:10 (10,000 cells/mL) and 1:100 (1000 cell/mL) using 1.5 mL microcentrifuge tubes containing media.

f. Filter the conditioned media through a 0.22 µm PES filter.

g. Seed 50 single cells in a 96-well plate by adding 50 µL of the 1:100 diluted cell solution in 10.5 mL of conditioned media and mix thoroughly by pipetting.

h. Move the cell-containing media to a reservoir and dispense 100 µL of cell solution into each well of a 96-well plate except for one well on the right bottom corner of the plate (H12). This well will be used to focus the microscope while searching for single cells.

i. Add 100 µL of the 1:10 cell solution in the bottom right corner well.

j. Keep the plate undisturbed in an incubator for 7 days.

CRITICAL: It is essential to homogenize the cell solution during dilution in order to yield a high number of single cells with minimal doublets.

Optional: A hemocytometer or other cell counter may be used instead of a Vi-cell analyzer.

3. Scan the 96-well plate to find monoclonal cells on the 7^{th} day post isolation.

    a. Focus the microscope using the cells in the bottom right corner well (H12).

    b. Scan the entire plate and mark the wells containing single colonies.

c. Keep monitoring the marked wells every day at roughly the same time and discard any well containing more than a single colony or that are observed to grow at a different rate than the majority.

CRITICAL: Having two cells in a well reduces assay sensitivity due to genetic heterogeneity.

Optional: We tested a microfluidic-based single cell sorter, yet the limiting dilution was as efficient as the single cell sorter and has an easier and less expensive workflow.

Note: We observed the plate at the same time every day from day 7^{th} to 14^{th} after single cell isolation to ensure that all selected clones are monoclonal and growing at the same rate.

4. Expand the monoclonal colonies on the 14^{th} day post isolation.
    a. Wash the wells that contain monoclonal colonies twice with 1X PBS.
    b. Trypsinize the monoclonal colonies in 100 µL of 0.25% trypsin-EDTA and incubate at 37°C for 5 min.
    c. Add 100 µL of media and break up all cell clumps into individual cells by pipetting up and down.
    d. Transfer 200 µL of the cell-containing media to a single well in a 12-well plate.
    e. Add 800 µL of media to each well containing cells in a 12-well plate.

Note: We noticed that some colonies are not easily detached during trypsinization from a 96-well plate. To minimize cell loss, make sure to check cells under a microscope to determine whether they are fully detached.

5. Once the monoclonal colonies in the 12-well plate are confluent, collect the cells.
    a. Wash the cells once with 1X PBS.
    b. Trypsinize in 500 µL of 0.25% trypsin-EDTA (enough to cover the cells) and incubate at 37°C for 5 min.
    c. Add 500 µL of media (≥ volume of trypsin added) and break up all cell clumps into individual cells by pipetting up and down.

<u>Cryopreservation of monoclonal cell lines</u>

6. Select a few monoclonal cell lines from above and expand the cells by placing each individual monoclonal cell line in a separate T-25 flask.
7. Add 5 mL of media and incubate at 37°C until the cells are confluent.
8. Freeze the cells at $1.5X10^6$ cells per cryovial in 1 mL of media containing 5% DMSO.
9. Store the cells at -80°C.

    CRITICAL: The purpose of this step is to preserve monoclonal colonies that are genetically homogeneous. Pre-existing genetic heterogeneity will lead to an overestimation in array change. The frozen stocks will be used to initiate clonal evolution.

Preparation of HOR CN control

10. Centrifuge the remaining parent cells after single-cell isolation at 300 x g for 5 min.
11. Remove the supernatant.
12. Store the parent cell pellet at -20°*C* until the ddPCR run is needed.

Preparation of positive control

13. Thaw a CHM13 cryostock containing $2X10^6$ cells and expand the cells to 10^7 cells.
14. Add 10^6 cells in each 1.5 mL microcentrifuge tube.
15. Centrifuge the cells at 200 x g for 5 min.
16. Remove the supernatant.
17. Store the cell pellet at -20°*C until the ddPCR run is needed*.

Preparation of 20X primer-pair working stock

18. Prepare a 20X primer-pair working stock for the targeted HOR and reference single gene (refer to Materials and equipment setup below).
19. Store the 20X working stock at -20°*C*.

Materials and equipment setup

This protocol uses the QX200 Droplet Digital PCR system which includes an automated droplet generator, PX1 PCR plate sealer, thermal cycler with 96–deep well reaction module, and QX200 droplet reader which comes with QuantaSoft Software (version 1.7). QuantaSoft is compatible with Windows 10 operating systems. The Droplet Digital PCR Applications Guide is available from the Bio-Rad website: https://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_6407.pdf

20X primer-pair working stock

Reagent	Final concentration	Amount
100 µM forward primer	2 µM	4 µL
100 µM reverse primer	2 µM	4 µL
ddH2O	n/a	192 µL
Total	n/a	200 µL

Store 20X primer-pair working stock at -20°C.

Droplet Digital PCR (ddPCR) reaction mix

Reagent	Final concentration	Amount
20X Primer pair stock (100 µM)	1X	1 µL
Alu I	2 units/mL	0.2 µL
2x ddPCR supermix for EvaGreen	1X	10 µL
DNA template	n/a	1 µL
ddH2O	n/a	7.8 µL
Total	n/a	20 µL

CRITICAL: The fresh master reaction mix without the DNA template should be made just before the run. The gDNA of the samples will be added individually.

CRITICAL: Vortex all reagents and spin down before preparing the reaction. When preparing the ddPCR reaction mixture add 1.25 x the total required volume because the viscosity of 2X supermix may lead to a pipetting error (see Figure 2. Column 6). Prepare a reaction in multiples of 8 because the automatic droplet generator takes 8 samples (1 column of a 96-well plate) and generates droplets simultaneously.

Step-by-step-method details

<u>Isolation of subclones from a monoclonal cell line for clonal evolution</u>

1. Isolate subclones as described in the step 1-5 of the Monoclonal cell line preparation section above using the monoclonal cell line frozen stock.

   Troubleshooting 1

2. Place the cells in 1.5 mL microcentrifuge tubes.
3. Centrifuge at 300 x g for 5 min.
4. Remove the supernatant.
5. Store the cell pellets of subclones at -20°*C until the ddPCR run is needed*.

Genomic DNA extraction from subclones and controls

Timing: 30 min. Time depends on number of samples.

6. Thaw frozen sub-clone cell pellets, parent cells, and ddPCR positive control that are stored at -20°*C*.
7. Resuspend pellets in 200 µL of 1X PBS.
8. Using the Qiagen blood and tissue kit, add 20 µL of proteinase K.
9. Add 200 µL Buffer AL and mix thoroughly by vortexing.
10. Add 200 µL of ice-cold 100% ethanol and mix thoroughly.
11. Place the mixture into a DNeasy mini spin column and centrifuge at >6000 x g for 1 min.
12. Discard the flowthrough.
13. Add 500 µL Buffer AW1 and centrifuge at >6000 x g for 1 min.

14. Discard the flowthrough.
15. Add 500 µL Buffer AW2 and centrifuge at >20,000 x g for 1 min.
16. Discard the flowthrough.
17. Centrifuge at >20,000 x g for 3 min.
18. Discard the flowthrough.
19. Transfer the column to a new 1.5 mL microcentrifuge tube.
20. Add 200 µL of Buffer AE to the center of the column and incubate at room temperature for 1 min.
21. Insert columns into tubes and centrifuge at >6000 x g for 1 min.
22. Discard the column and collect the gDNA.

Note: We tested a direct PCR lysis reagent and NucleoSpin tissue XS assay to reduce the number of cells needed, but the results using these assays were not as consistent as the results from the Qiagen blood and tissue kit.

Pause point: gDNA can be stored at -20°*C*.

Genomic DNA quantification and dilution

Timing: at least 30 min. Required time depends on number of samples.

23. Measure the concentrations of the gDNA extracts from the cell pellets using a DeNovix dsDNA high sensitivity assay.
    a. Equilibrate all components to room temperature before use.
    b. Vortex and spin down all components.
    c. Prepare a working solution for the samples and two standards. Each standard and unknown sample requires 190 µL of working solution. Add 1.9 µL of 100X dye in the 188.1 µL buffer per measurement. Scale up as needed to make enough volume to use.
    d. Vortex to mix thoroughly.
    e. Aliquot 190 µL of the working solution in a thin, UV-transparent 0.5 mL tube for each sample and standard.
    f. Add 10 µL of standard and unknown sample to the assay tubes.
    g. Vortex tubes to mix.
    h. Incubate the samples and standards at room temperature for 5 min.
    i. Generate a standard curve by DeNovix DS-11 FX using two standards.
    j. Measure the concentration of the samples.

    Optional: any assay that can quantify 100pg-250ng range fluorometrically including Qubit dsDNA assays can be used.

24. Vortex thoroughly and dilute samples to 2 ng/µL in Buffer AE using Rainin low retention tips.

25. Measure the concentration of 2 ng/μL stock using a DeNovix dsDNA ultra-high sensitivity assay.
    a. Equilibrate all components to room temperature before use.
    b. Vortex and spin down all components.
    c. Prepare a working solution for the samples and two standards. Each standard and unknown sample requires 200 μL of working solution. Add 0.5 μL of the 400X dye and 2 μL of the 100X enhancer in 197.5 μL buffer per measurement. Scale up as needed to make enough volume to use.
    d. Vortex to mix thoroughly.
    e. Aliquot 200 μL of the working solution in a thin, UV-transparent 0.5 mL tube for each sample and standard.
    f. Add 10 μL of standard and unknown sample to the assay tubes.
    g. Vortex tubes to mix.
    h. Incubate samples and standards at room temperature for 5 min.
    i. Generate a standard curve by DeNovix DS-11 FX using two standards.
    j. Measure the concentration of the samples.
26. Repeat steps 24-25 until 2 ng/μL (< ±10%) stocks are made from all samples.
27. Vortex thoroughly and dilute 2 ng/μL stock to 1:20 in Buffer AE using Rainin low retention tips.

CRITICAL: The CN of the reference single gene is used to normalize the CN of the HOR of the same subclone. To minimize subsampling error, it is important to make the 2

ng/µL stock as accurately as possible and mix thoroughly before preparing the diluted gDNA stock.

Pause point: Diluted gDNA can be stored at -20°C.

ddPCR reaction setup

Timing: 1 h (96 samples)

CRITICAL: We measure the HOR and reference single gene (SG) CNs in separate reactions using the EvaGreen assay due to constraints caused by the differences in ddPCR dynamic range between the HOR and reference SG when using the same amount of gDNA input. Therefore, we use 2 ng of gDNA for the reference SG measurement and a 1:20 dilution of 2 ng gDNA stock for HOR measurement. This is the optimal condition to quantify D11Z1 CN in U2OS cells. The optimal gDNA concentration for each different SG and dilution factor for each different HOR should be determined based on the CN of target HOR and SG in the cell of interest to fit within the dynamic range of the ddPCR.

CRITICAL: EvaGreen dye binds to both double-stranded DNA and nonspecifically single-stranded DNA. Therefore, concentrations of gDNA input and primer need to be optimized to ensure accurate CN measurement.

CRITICAL: To minimize subsampling error, use low retention tubes and tips for all steps and pre-wet tips to avoid bubbles in the tips.

CRITICAL: Avoid introducing air bubbles to the droplet generator cartridge. This will impact droplet generation and reduce the assay efficiency. Therefore, add 10% excess volume (22 µL) to prevent air bubble formation (see Figure 2. Column 4).

28. Thaw the diluted gDNA stocks of subclones, parental cells, and ddPCR control.
29. Vortex thoroughly to homogenize gDNA in solution.
30. Place 2X supermix and 20X primer pair stocks at room temperature to thaw.
31. Following the ddPCR reaction mix table in the Materials and equipment section above, prepare two ddPCR master reaction mixes without gDNA (the gDNA will be added in step 35) according to the number of samples needed, in triplicate (see Figure 2. Column 5) and including the two controls.

Note: One master mix will include the HOR target primers and the other will include the single gene reference primers**.**

Note: Refer to Figure 1 for a suggested plate layout and Figure 2 for an example of the master mix volume calculation.

CRITICAL: Incubate the ddPCR reaction with the restriction enzyme for at least 30 min to ensure every CN of the HOR and SG are isolated and partitioned into droplets.

32. Label 1.5 mL low retention microcentrifuge tube that match the number of samples that are being measured.
33. Vortex thoroughly and aliquot 69 µL of reaction master mix to each tube.
34. Vortex thoroughly to mix the gDNA.
35. Add 3.6 µL of either 2 ng/µL (SG), 1:20 dilution of the 2 ng/µL stock (HOR), or ddH_2O to the corresponding tubes containing the target primer-pair.
36. Vortex thoroughly to mix the reaction.
37. Incubate the reaction in the dark for 30 min.
38. Vortex tubes thoroughly to homogenize the reactions.
39. Aliquot 22 µL from each tube to each corresponding well in a 96-well ddPCR plate in triplicate (see Figure 1).
40. Turn on the PCR plate sealer which is set to seal at 180°C for 5 sec.
41. Place a pierceable foil seal on the top of the plate (the redline should be on the top of the plate) and seal the plate.
42. Vortex the plate at 2000 rpm for 1 min.
43. Centrifuge the plate at 1000 x g for 1 min.

CRITICAL: Remove air bubbles and ensure the ddPCR reaction mix is at the bottom of the wells. If there are any bubbles after centrifugation, repeat step 42.

Droplet generation

Timing: ~45 min (for 96 samples). It will take less time for fewer samples. It takes an average of 3.5 min per column.

44. Set up the configuration on the automated droplet generator (AutoDG) as follows:
    a. Select oil type as EvaGreen.
    b. Specify the columns that contain the samples.
45. Replace the oil and consumables as follows (the green lights on the instrument decks or touchscreen will turn on if the consumables are placed correctly).
    a. Ensure that the oil volume is sufficient and the type of droplet generation oil is EvaGreen.
    b. Load droplet generator cartridges onto the cartridge blocks on the back row (the green gaskets should be on the right side).
    c. Insert an empty pipet tip waste bin.
    d. Remove the lids and place pipet tips on the pipet tip blocks on the middle row.
    e. Place a 96-well ice cold chill block on the droplet plate holder (front right).
    f. Place a new ddPCR 96-well plate in the 96-well chill block.
46. Insert the sealed ddPCR reaction plate into the sample plate holder (front left).
47. Start the AutoDG when all requirements are satisfied.

48. Once the droplet generation is complete, place a pierceable foil seal onto the 96-well plate containing the droplets in the chill block (the red line should be on the top of the plate).
49. Gently remove the plate and seal at 180°C for 5 sec.

CRITICAL: The droplets are very fragile. We recommend processing the PCR amplification step as soon as the droplet generation is complete. Do not vortex or spin down the plate. Handle the plate very carefully until it is loaded on the thermal cycler.

CRITICAL: Improper sealing of the plate leads to oil evaporation during thermal cycling and can compromise droplet data quality.

Note: See the AutoDG instrument manual for details: https://www.bio-rad.com/webroot/web/pdf/lsr/literature/10043138.pdf

Note: A QX200 droplet generator can be used instead of the AutoDG. See the instrument manual for details: https://www.bio-rad.com/webroot/web/pdf/lsr/literature/10031907.pdf

Note: Droplets will have an opaque layer at the top of each well if the droplets are properly generated.

Note: The chill block needs to be stored at -20°C upside down for more than 2 h before use. The instrument is not sensitive enough to detect errors occurring in this block such as a misplaced sample plate or a failure to place a new plate. A mistake in this step can lead to the failure of the entire run.

Note: The completion time for droplet generation can vary between runs. We recommend monitoring the instrument performance until the first column is processed before leaving the machine.

Thermal cycling (PCR)

Timing: 2.5 h

50. Transfer the sealed droplet plate to a thermal cycler and start the PCR cycling using the program as below (Table 2). The volume of the thermal cycling reaction is 40 µL. The ramp rate should be 2°C/s for all cycles.

Table 2. PCR cycling conditions

Steps	Temperature	Time	Cycles
Enzyme activation	95°C	10 min	1
Denaturation	96°C	30 sec	40 cycles
Annealing/extension	56°C	60 sec	
Signal stabilization	4°C	5 min	1
	90°C	5 min	1
Hold	4°C	Forever	

CRITICAL: The annealing temperature must be optimized for every set of primers using a temperature gradient (65°C-55°C).

Note: We noticed that incubating the plate at 4°C overnight after PCR cycling increases the number of accepted droplets.

Pause point: The plate can be stored at 4°C overnight after completion of the PCR step until droplet reading.

Droplet reading

Timing: 2h (for 96 samples). Droplet reading takes about 10 min for each column.

51. Remove the metal retainer and insert the sealed PCR-amplified reaction plate into the QX200 droplet reader and secure the retainer.
52. Open the Quantasoft software and prime the QX200 droplet reader.
53. Set up a new plate layout according to the experimental samples.
54. Designate the parameters as follows:
    a. Define “Supermix” as “EvaGreen”
    b. Set “Experiment Type” as “Direct Quantification”
    c. Assign “Assay Type” as “Single Target per Channel”
    d. Under “Target Info”, select signal Ch1 as “EvaGreen” and signal ch2 as “None”
    e. Configure the names of the samples.

55. Start the ddPCR reading.
56. Once the reading is complete, remove the empty plate from the QX200 droplet reader and dispose the waste in accordance with institutional, state, and local regulations.

    Note: ddPCR droplet reading will not proceed if the droplet reading oil is insufficient for the reaction or the waste container is full.

Expected outcomes

Successful completion of the monoclonal isolation step should result in greater than 25 monoclonal cells, with an average of 4 doublets, when approximately 0.5 cells/well was added to the 96-well plate. A successful ddPCR run should result in >16000 accepted droplets and a separation of two peaks representing the amplitudes of the negative (background) and the positive (target). The CN of D11Z1 in the positive control (CHM13) should be ~3500 copies/µL with minimal background in the negative control.

Quantification and statistical analysis

Overview

The purpose of this section is to quantify the absolute number of copies/µL for both HOR and SG and normalize the HOR CN by SG CN to estimate an average of HOR CN per centromeric array. The HOR CN per array across the subclones will be compared to the value of parental monoclonal cells that the subclones were isolated from. The frequency of CN change across the subclones is determined using an ANOVA test followed by Tukey's Honestly Significant Difference test. The two steps that are needed are as follows:

- Use Quantasoft to estimate the CN.
- Quantify HOR CN per array by normalizing with the SG CN.

The number of accepted droplets and negative droplets are used to calculate # copies/µL. The number of positive droplets influences the technical error of the measurement.

Data analysis

Timing: 0.5h

57. After completion of the droplet reading, open Quantasoft and select the run.
58. Select the well that contains the NTC and ensure the well is free of contamination (see Figure 3 and **Troubleshooting 2**)
59. Determine the amplitude threshold to separate the positive and negative droplets for the purpose of quantification as follows:
    a. Select all experimental wells targeting HORs.

b. Click the 1-D amplitude plot view.
c. Manually place a horizontal gate to separate the clusters of positive and negative droplets.
d. Repeat steps a-c with all wells targeting SG.
e. Export the results.csv file from Quantasoft.

Note: For a more conservative approach for defining positive droplets, set the horizontal gate close to the main cluster of positive droplets.

60. To further increase data quality, open the results.csv file and exclude any well that meets the following criteria.
    a. < 10000 accepted droplets.
    b. A distinctively different average amplitude from the wells that share the same target.
    c. Number of positive droplets are < 100.
    d. Number of negative droplets are < 10.
61. Multiply the HOR CN x 20 to account for the dilution factor difference between the HOR and SG gDNA input (1:20).
62. Normalize the HOR CN with the SG CN by pairing the first HOR and SG replicates and dividing the HOR CN by the SG CN.
63. Repeat step 62 for any other technical replicates.
64. Conduct the ANOVA test followed by Tukey's HSD test between each subclone CN and the parental CN value in R.

Limitations

Due to single-cell isolation requirement, this protocol requires four weeks to collect the subclones for measurement. Therefore, the protocol is limited to cell lines that can grow after single cell isolation. A lower number of cell divisions may increase the sensitivity of the CN change detection due to sharing higher genetic homogeneity in monoclonal populations. However, $5x10^5$ cells were required for gDNA extraction to yield consistent results with less technical error. Since we used previously published and tested primers[125], no primer design step is included in this protocol.

Troubleshooting

Problem 1:

A high number of doublets are observed after monoclonal isolation in step 1.

Potential solution:

Use a 40 µm cell strainer mesh instead of breaking up clumps by pipetting. Add fewer than 50 cells/96-well and isolate the monoclonal cells in two 96-well plates to collect enough subclones.

Problem 2:

Positive droplets observed in the NTC in step 58 (see Figure 3A).

Potential solution:

Due to the exceptionally high CN of HORs and sensitivity of the ddPCR, any cross-contamination between samples during ddPCR preparation can cause signal amplification. We recommend cleaning the pipettes and workbench before beginning the assay and use filter tips to prevent any contamination.

Problem 3:

Read failure or presence of extremely high copies reported in Quantasoft (see Figure 3D).

Potential solution:

Inaccurate gDNA quantification can lead to an excess amount of gDNA input. This may result in the ddPCR creating very few or even no negative droplets, resulting in read failure or an extremely high reported CN. We recommend re-quantifying the input or titrating the gDNA and running a test plate to determine the optimal gDNA input to fit within the ddPCR dynamic range.

Problem 4:

The CN in the positive control is lower than the expected value.

Potential solution:

CN measurement can be low when the restriction enzyme digestion is incomplete, resulting in partial isolation of target copies. This can be caused by an insufficient amount of enzyme, an unsatisfactory enzyme, or insufficient incubation time. We recommend using a fresh vial of enzyme and incubate at least 30 min.

Problem 5:

Inconsistent CN between replicates.

Potential solution:

High subsampling error can contribute to inconsistent CN between replicates. One of the major sources of high subsampling error is a low number of positive droplets. The coefficient of variation is 6% (~14.6 copies/µL) and 10% (5.8 copies/µL) when the positive droplets are below 250 and 100 respectively with an assumption of 20000 accepted droplets. We recommend discarding measurements that are below 100 positive droplets.

3.3 Figures

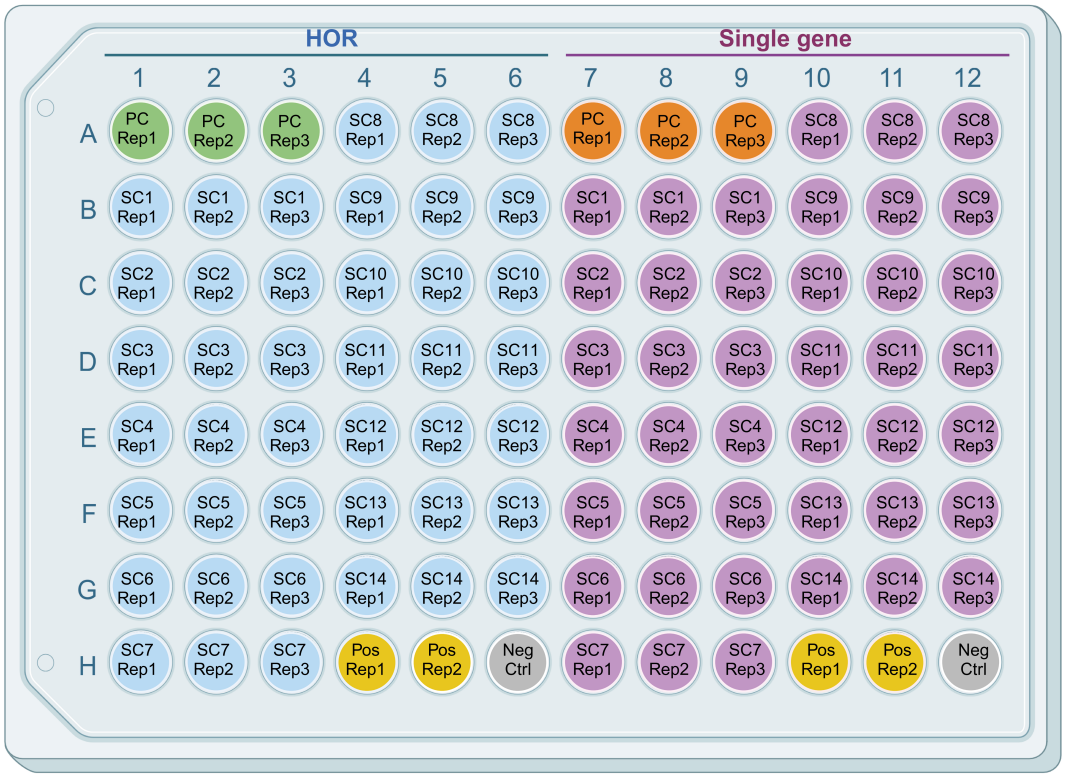


Figure 3.1. Example plate layout for quantifying copy numbers of higher order repeats (HOR) and reference single gene (SG) across subclones in a 96-well plate. In columns 1-6 the HOR measurements are highlighted blue and in columns 7-12 the reference SG are highlighted purple. The parental cells (PC) that provide the baseline HOR and SG CN are highlighted green and orange respectively. The assay positive control is highlighted yellow. All measurements are in triplicate except the assay positive control (yellow) and the no-template control (grey).

Reagent	Final Concentration	Amount	X 1.1 (Bubbles)	X 3 (Triplicate)	X 1.1 (Viscosity)
20X Primer Pair Stock (100μM)	1X	1 μL	1.1 μL	3.3 μL	3.6 μL
Alu I	2 units/mL	0.2 μL	0.22 μL	0.7 μL	0.7 μL
2X ddPCR Supermix for EvaGreen	1X	10 μL	11 μL	33 μL	36.3 μL
DNA Template	n/a	1 μL	1.1 μL	3.3 μL	**3.6 μL**
ddH_2O	n/a	7.8 μL	8.6 μL	25.7 μL	28.3 μL
Total	**n/a**	**20 μL**	**22 μL**	**66 μL**	**72.6 μL**
		Target aliquot volume per sample (triplicate)		**69 μL**	

Figure 3.2. An example volume calculation for the reaction master mix of the HOR in a subclone without the gDNA template in triplicate reflecting the source of errors. Additional master mix should be prepared to prevent bubble formation (1.1x amount for single reaction) and to account for supermix viscosity (additional 1.1x of triplicate volume). The gDNA template (red) will be added separately. The target volume for each sample in triplicate is at least 69 μL (blue). See also Figure 1 and Table 1.

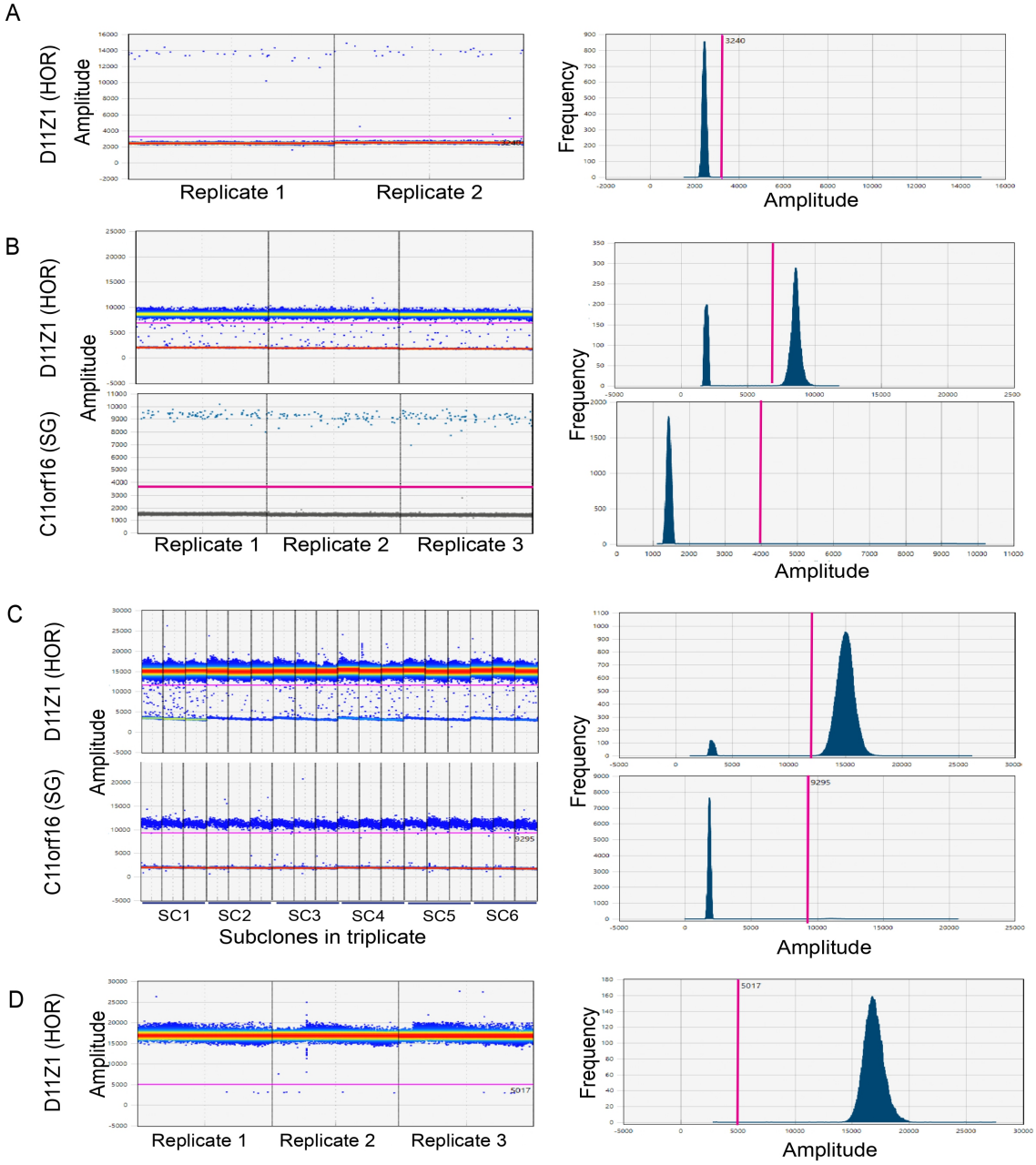

A
D11Z1 (HOR)
Amplitude
Replicate 1
Replicate 2
Frequency
3240
Amplitude
B
D11Z1 (HOR)
C11orf16 (SG)
Amplitude
Replicate 1
Replicate 2
Replicate 3
Frequency
Amplitude
C
D11Z1 (HOR)
C11orf16 (SG)
Amplitude
9295
SC1
SC2
SC3
SC4
SC5
SC6
Subclones in triplicate
Frequency
9295
Amplitude
D
D11Z1 (HOR)
Amplitude
5017
Replicate 1
Replicate 2
Replicate 3
Frequency
5017
Amplitude

Figure 3.3. Examples of ddPCR analysis using Quantasoft. (A) Example of horizontal and vertical gating to separate the negative droplets from the no-template control targeting D11Z1 higher order repeats (HOR) in two technical replicates. The pink line represents the amplitude threshold that is used to distinguish between negative and positive droplets. The blue dots above the pink line are positive droplets. (B) Gating example for HOR and reference single gene (SG) in the positive control CHM13. (C) Gating example for HOR and reference SG across six subclones of U2OS in triplicate. (D) Example of quantification failure due to excess gDNA input.

Chapter 4: Expansion of human centromeric arrays in cells undergoing break-induced replication

Modified from an article originally published in *Cell Reports*:

Soyeon Showman, Paul Talbert, Yiling Xu, Richard O. Adeyemi, Steven Henikoff.

4.1 Abstract

Human centromeres are located within α-satellite arrays and evolve rapidly, which can lead to individual variation in array lengths. Proposed mechanisms for such alterations in lengths are unequal cross-over between sister chromatids, gene conversion, and break-induced replication. However, the underlying molecular mechanisms responsible for the massive, complex, and homogeneous organization of centromeric arrays have not been experimentally validated. Here, we use droplet digital PCR assays to demonstrate that centromeric arrays can expand and contract within ~20 somatic cell divisions of an ALT-positive cell line. We find that the frequency of array variation among single-cell-derived subclones ranges from a minimum of ~7% to a maximum of ~100%. Further clonal evolution revealed that centromere expansion is favored over contraction. We find that the homologous recombination protein RAD52 and the helicase PIF1 are required for extensive array change, suggesting that centromere sequence evolution can occur via break-induced replication.

4.2 Introduction

Centromeres are chromosomal regions where kinetochore assembly and microtubule attachments occur to ensure faithful genetic transmission of chromosomes to daughter cells during mitosis and meiosis[128,129]. Active centromeres are epigenetically identified by histone H3 variant CENPA[130] and in most seed plants and animals are composed of mega-base length arrays of tandem repeats known as satellites that can phase CENPA

nucleosome positions[26,131-133]. Centromere function is essential across all eukaryotes, yet centromere sequences evolve rapidly, a phenomenon known as the centromere paradox[134]. Comparative centromere sequence analysis between two complete human hydatidiform moles (CHMs) that have fully assembled centromere sequences reveals that only ~63-71% of the sequences can be aligned between the two haplotypes, which highlights the rapid evolution of centromere sequences even within a single species[135].

Human centromeres are located at α-satellite arrays (α-Sat) comprised of blocks of 171 bp head-to-tail tandemly organized monomers that can differ by 50~80%[76], but are organized in highly homogeneous higher order repeats (HORs), which themselves have a nested structure[3,136] in which the most recent and homogeneous HOR that forms the active centromere is surrounded by older more divergent HORs flanked by divergent monomers[3,136]. Based on the layered expansion model of centromeric array evolution, the active α-Sat HOR originates from newly emerged small repeats and expands into a mega-base-sized array within the active centromere while pushing the pre-existing diverged α-Sat HORs to the periphery of the active centromere[3,76]. The copy number (CN) of the active HOR, which indicates the array size of the centromere, varies substantially among individuals (up to ~80-fold)[135]. In addition, the HOR CN between cancer cells and their normal tissue counterparts significantly differ, which reveals that the array sizes can change in the lifetime of an organism[137]. Despite the extreme degree of inter- and intra-individual polymorphism in HOR CNs, the molecular mechanisms that underlie array expansion and contraction, the rate of variation, and consequences of variations are not fully understood.

A widely cited model used to explain the expansion and contraction of satellite arrays involves unequal crossover and gene conversion between sister chromatids during homologous recombination to repair DNA double strand breaks (DSB)[19]. This model (Figure S1A), if correct, would predict stochastic expansion and contraction resulting in randomly mutated monomer sequences without any specific structure due to functional constraints[138]. Tandem repeats that are repaired by Single Strand Annealing (SSA), one of several DSB repair pathways, will cause a deletion of satellite repeats[23]. This loss of repeats will eventually shrink centromeric arrays over time unless there is a mechanism that counteracts the loss[26].

An alternative model (Figure S1B) is based on break-Induced replication (BIR)[139] which is a one-sided DSB repair mechanism that can replicate hundreds of kilobases in budding yeast[27]. BIR has been implicated in oncogene-induced DNA replication[113], replication stress-induced DNA repair synthesis in mitosis (MiDAS)[123,140], and in significantly elongating the size of the telomere in Alternative Lengthening of Telomere (ALT) positive human cancers[141]. Centromeres are enriched with DNA breaks[142] that may be caused by replication fork collapse[110] due to the presence of replication barriers such as the constitutive centromere associated network (CCAN) and non-B form DNA secondary structures[86,143-147]. Following the 5'-to-3' resection of a one-sided DSB caused by a collapsed replication fork, the 3' single-stranded DNA is available for strand invasion, which triggers BIR[27]. Because of the tandemly repetitive structure of satellites, BIR can create deletions, duplications, or neither, depending on the location of re-initiation of a collapsed fork[26]. These outcomes can cause the α-Sat monomer turnover and HOR structure frequently observed in the centromere[138,139,148]. Previous studies at the rDNA repeat arrays in budding yeast have shown

that BIR favors out-of-register re-initiation of broken forks leading to array expansion[32]. This expansion bias can counteract the erosion caused by SSA. Frequent dissociation of Pol δ from the template DNA and reduced efficiency of mismatch repair during BIR can explain an elevated nucleotide substitution mutation rate in the centromere[27,149].

These alternative models have not been experimentally validated because of the homogeneity in satellite sequences, their complex organization, and the extremely large size of the centromere. These features of satellite centromeres have hampered centromere experimental biology for decades[26] until the recent advances in telomere-to-telomere (T2T) assembly[3].

With the benefit of fully assembled centromere sequences[2], we measured the CN variation in the chromosome 11 centromeric HOR D11Z1 at intervals of ~20 somatic cell divisions across subclones of the K562 and U2OS cell lines. We found that the D11Z1 CNs vary among subclones of U2OS with a change frequency from ~7% of subclones to ~100%. Using this basal rate of change that we identified, we set out to test by mutation the involvement of the homologous recombination protein RAD52 and helicase PIF1[150,151]. Our data indicates that both RAD52 and PIF1 are required to cause a large change in CN during somatic cell divisions, suggesting that these changes occur via BIR. Our findings provide insight into the mutational mechanism that underlies rapid centromere evolution, while offering a tool to further investigate the consequences of centromeric array variation, which influences the occurrence of genetic disorders, and facilitates speciation[128,134,152].

4.3 Results and Discussion

4.3.1 Comparison of ddPCR assays for HOR copy number measurement

The molecular mechanisms that contribute to extensive array length polymorphism during mitosis in mammalian cells are unclear. Recent studies have demonstrated that centromeric array sizes can be experimentally estimated using a droplet digital PCR (ddPCR)-based method[137]. With the recent T2T CHM13 genome assembly[2], primers can now be designed to target a single, unique amplicon for each HOR based on polymorphisms that differ between monomeric units (Table S1). Because the α-Sat is composed of repeats of HORs, if the HOR CN per array is known, then the size of an array can be estimated by multiplying the size of the HOR unit by the CN (Figure 1A). For example, if the experimental D11Z1 HOR CN is 3600, then by multiplying the unit size, which is 855 bp, the size of the entire D11Z1 array is ~3.07 Mb. The ddPCR is a reference-free quantification method with a 2-fold greater sensitivity than qPCR. This method allows us to quantify the CNs of HORs within a chromosome-specific array by partitioning every copy of the HOR within an α-Sat array that is isolated by restriction enzyme digestion to >18,000 droplets (Figure 1B). The droplets are counted by the machine and used to calculate the CN. The HOR CN is normalized using the CN of a single-copy gene located on the same chromosome as the HOR being measured in a parallel reaction to ensure that the HOR CN reflects a single chromosome array. By normalizing to a single-copy gene on the same chromosome, the CN per array will be the average CN of the homologous chromosomes.

Despite promising applications to centromere biology, the single-copy ddPCR-based assay is associated with an intrinsic error rate between biological replicates of ±10%[137], probably because the parallel reactions must be carried out at different DNA concentrations. To mitigate this issue, we developed a 5S rDNA probe-based assay that reduces sub-sampling error by measuring both target and reference CNs in the same reaction, using 5S rDNA repeats as the reference gene (Figure S2A). To validate this assay, we measured the D6Z1, D11Z1, and D18Z1 CNs in CHM13 cells using the two methods and compared them with the values that are derived from the CHM13 assembly (Figure 1C)[137]. While the single-copy assay produced values close to those derived from CHM13 assembly, the 5S assay produced values nearly identical to the assembly values with less technical error at the cost of reduced dynamic detection range of the ddPCR.

4.3.2 Centromeric array CN can change within ~20 somatic cell divisions

A previous study has reported that centromeric array CNs vary substantially between cancers and their counterpart normal tissues, which indicates that centromeric array length alteration can occur in somatic cells[137]. In addition, the pediatric cancers medulloblastoma and acute lymphoblastic leukemia tend to show directionality among the five chromosome specific arrays measured, such as all gain or all loss in HOR CNs, indicating a more coordinated alternation pattern[137]. These coordinated patterns are especially useful to increase the sensitivity of the ddPCR method, since the HOR CNs estimated by ddPCR are an average over homologous chromosomes. This could cause the change in individual

chromosome arrays to be masked when the direction of HOR CN change varies between events within the pool of different homologous chromosomes. Therefore, we reasoned that pediatric cancers may be an ideal system because of the coordinated alteration patterns in HORs witnessed in the previous study. We chose the pediatric osteosarcoma cell line U2OS because its telomere maintenance mechanism is known to utilize BIR[141], which we hypothesized to be the primary mechanism of CN change in centromeres.

We first assessed whether the HOR copy in an array can change during ~20 somatic cell divisions in the U2OS cell line and determined the rate of change that can be measured using the ddPCR-based method. We measured the CNs of D6Z1, D11Z1, D18Z1, and DXZ1 centromeric arrays in single-cell-derived subclones of the U2OS cell line that had undergone ~20 somatic cell divisions, which we named Group1, along with the parental cells that were frozen down right after single cell isolation so that we could identify any subsequent changes that occurred after isolation (Figure 2A). The HOR CN is normalized by the single-gene CN to determine an average HOR CN per array or allele. To identify the subclones that had changed HOR CN significantly since the time of single cell isolation from the parental cells, Tukey's Honestly Significant Difference (HSD) test was conducted between parental measurements and subclone values following ANOVA test. Three out of seven subclones of Group1 showed an expansion in the D11Z1 CN, two subclones showed a decrease in D18Z1 CN, and one subclone showed an increase in D18Z1 CN (Figure 2B-C, Table S2). While the frequency of change is the same between D11Z1 and D18Z1, the magnitude of the change was greater in D11Z1 than D18Z1 (30% vs 22% maximum). We repeated the measurements using the 5S assay and obtained essentially identical results, which validated the changes we observed

(Figure S2B, Table S2). Although the karyotype of U2OS cells is unstable in long-term culture, the agreement between the unlinked 5S and linked single copy gene assays also indicated that the ploidy of chromosome 11 did not change over the course of our experiment. Therefore, the magnitude of CN changes occurring in U2OS cells is above the technical error threshold and can be confidently measured using the single-copy assay. Whereas D11Z1 and D18Z1 CNs changed, there were no CN changes in the D6Z1 and DXZ1 subclones (Figure 2D-E, Table S2).

Next, we were curious as to whether the CN changes observed in U2OS cells were broadly characteristic of cancer genomes or were an intrinsic property of the U2OS cell line. Therefore, we also measured the D11Z1 CNs across single-cell-derived subclones of the K562 cell line that had also undergone ~20 somatic cell divisions. Using the single-copy assay, none of the K562 Group1 subclones changed D11Z1 CN (Figure 2F, Table S2), in contrast to the high frequency and magnitude of change we observed in U2OS cells. However, using the 5S assay, the frequency of K562 clones that changed CN was ~33% (5 out of 15), and all showed expansion, similar to U2OS clones (Figure S2C, Table S2). The maximum CN alteration observed in K562 was a ~12.4% increase (SC10), which is close to the 10% technical error rate of the single-copy assay, in contrast to the ~30% maximum increase in U2OS CN (Figure 2G), likely explaining the failure to detect changes using the single-copy assay in K562 cells. Therefore, we used the single-gene assay in the U2OS cell background for follow-up experiments since it is a more conservative method and the magnitude of change in U2OS subclones far exceeds the technical error threshold.

We conclude that the centromeric array can expand and contract in mitotic cells within ~20 cell divisions in both U2OS and K562 cells, but the magnitude of change is far less in K562. One difference between these cell lines is that U2OS cells undergo BIR-mediated ALT[141], which is associated with a mutation in the *ATRX* gene that results in a short-lived, truncated protein[153]. ATRX can form a complex with cohesin and MeCP2[154], and knockout of ATRX causes a defect in cohesion at telomeres, where loss of cohesin between sister chromatids facilitates non-allelic telomere interactions[155]. ATRX depletion likewise causes cohesion defects at centromeres[156], and we hypothesize that this facilitates non-allelic, out-of-register BIR in tandem centromere arrays, resulting in greater changes in D11Z1 CN in U2OS cells than in K562 cells. Similarly, disruption of cohesion between sisters at the 35S rDNA locus in budding yeast results in the amplification of rDNA through out-of-register replication[157].

4.3.3 Expansion occurs more frequently than contraction in D11Z1

The unequal crossing-over model predicts erosion of the centromere over time because any broken replication forks that are repaired by SSA will lead to a deletion of tandem repeats such as HORs[19]. This will inevitably shrink the array unless there is a selective pressure that counteracts the loss so that mega-base array lengths that are seen in most human centromeres are maintained. We had hypothesized that out-of-register re-initiation of replication behind the fork would occur more frequently because the DNA behind the fork is more accessible compared to the positively supercoiled DNA in front of the fork, which

would favor expansion (Figure S1B)[26]. Therefore, we wondered whether centromeric arrays in Group1 that had expanded would continue to expand or would contract.

To this end, we isolated single cell subclones from the two subclones (SC3 and SC4) that increased the D11Z1 CN from the U2OS Group1 (Figure 3A) and allowed them to expand another ~20 somatic cell divisions (Group2). We then measured the D11Z1 CN across Group2 samples and compared them to the corresponding parental HOR CNs. The frequency of CN change in SC3 Group2 was ~42%. At the most extreme CN change, SC3 increased HOR CN ~30% compared to the parental cells of Group1 (Figure 2B) and its subsequent subclone, SC 3.11, gained an additional ~43% in CN resulting in a total expansion of ~86% from parental cells to the SC3.11 cells (Figure 3B, Table S2). Unexpectedly, half of the Group2 subclones of SC4 decreased the D11Z1 CNs close to the parental cell value of Group1 (Figure 3C, Table S2). This extreme drop was only observed in these subclones, which were derived from a parental cell that had among the highest starting CN. This led us to examine whether the starting parental CN might influence the direction of change in their subclones. Therefore, we selected SC3.3, which retained a high D11Z1 CN similar to its parent SC3, and SC4.4, which showed a decrease in CN compared to its parent SC4. Single cells were isolated from SC3.3 and SC4.4 (Group3), underwent ~20 somatic cell divisions, and the D11Z1 CNs were measured (Figure 3A). The frequency of CN change for Group3 of SC3.3 was ~7% with only one subclone contracting (Figure 3D, Table S2). Interestingly, the magnitude of the decrease observed in the SC3.3.6 subclone, whose SC3 parent had a similarly high CN as SC4, was similar to the decrease observed in Group2 of the SC4 subclones (Figure 3A). Finally, all Group3 subclones of SC4.4, which had

previously contracted, increased CN (Figure 3E, Table S2). Among all 55 subclones from U2OS Groups 1-3, ~35% showed expansion and ~13% contraction in D11Z1 CN. This matches our hypothesis that expansion is favored over contraction. While contraction can occur during somatic cell divisions, this only occurred when the subclones were isolated from parent cells with high CN. Therefore, it is tempting to speculate that there might be a homeostasis mechanism that sets an upper limit to centromere CN, analogous to the mechanism that constrains rDNA copy number[32].

4.3.4 Centromeric array CN alteration requires both RAD52 and PIF1

BIR-mediated satellite expansion/contraction is a compelling model that can explain unique characteristics of the αSat such as extreme length, complex repeat structure, sequence turnover, and high substitution mutations[139]. During the broken replication fork repair process in yeast, BIR can copy more than 100kb with up to a 1000x higher base substitution mutation rate compared to S-phase DNA synthesis[27]. In addition, out of register re-initiation of replication during BIR repair can cause deletion or addition of multiple copies of repeats resulting in monomer turnover[148]. However, neither the BIR mechanism nor a dependence on the proteins that are required for BIR have been experimentally validated during centromere sequence synthesis.

Thus, we sought to test BIR as the molecular mechanism underlying rapid centromere evolution[26,139] based on the frequency of D11Z1 array change that we established in our U2OS assays. BIR can occur via the RAD52-dependent pathway, which is well-known for ALT

telomere maintenance and MiDAS[141,158]. RAD52-dependent BIR has previously been suggested to mediate centromere expansions in U2OS cells[107]. Mammalian RAD52, which has strand annealing activity, facilitates strand invasion by forming a displacement loop which, in turn, promotes initiation of DNA replication after fork collapse in BIR[140]. PIF1, which is an evolutionarily conserved 5'-to-3' helicase, is indispensable during BIR for initiation of Pol δ DNA synthesis in budding yeast[159] (Figure 4A). Both RAD52 and PIF1 depletion suppressed BIR-mediated repair of DSBs that were induced by endonuclease I-SceI in U2OS cells[140,151]. In addition, PIF1 is important for BIR-mediated repair of collapsed forks induced by replication stress, resulting in much longer tracts of DNA synthesis than at endonuclease-generated DSBs[151]

As RAD52 and PIF1 are essential for BIR but not for cell viability we could use knockdowns of these proteins to ask whether BIR is required for the CN changes we observed in U2OS cells. Accordingly, we used the CRISPR-Cas9 system to generate knockdown (KD) cell lines, one with disrupted *RAD52* and two with disrupted *PIF1*, along with a non-target control that maintains undisrupted *RAD52* and *PIF1* genes in U2OS cells. To generate the *RAD52*^{KD} cell line, we used two single-guide RNAs (sgRNAs) that target exons 3 and 5, which are important for RAD52 oligomerization. For *PIF1*^{KD}, we used two sgRNAs that target exons 2 and 6, which are important for helicase activity (Table S3). We validated one *RAD52*^{KD} clone and two *PIF1*^{KD} clones using immunoblotting (Figure 4B) and confirmed the comparable growth rates between U2OS non-target control and *RAD52*^{KD} or *PIF1*^{KD} (Figure S3A). We next isolated single cells from *RAD52*^{KD} C1, *PIF1*^{KD} C2, and the non-target control and cultured them through ~20 somatic cell cycles and measured the D11Z1 CN across the subclones. While

the frequency of D11Z1 CN change was ~87% (13 out of 15) in the non-target control (Figure 4C, Table S2), the D11Z1 CN did not change across the *RAD52*^{KD} (Figure 4D, Table S2) and *PIF1*^{KD} C2 clones (Figure 4E, Table S2). These clear results strongly support our hypothesis that BIR is critical for the array size changes that we observed. We repeated the experiment with another knockdown clone, *PIF1*^{KD} C1. The frequency of D11Z1 array change was >80% in the non-target control (Figure S3B, Table S2), yet no subclone changed D11Z1 CN in the *PIF1*^{KD} C1 clone (Figure S3C, Table S2). Because both *PIF1*^{KD}*s* gave the same result, we can rule out random clonal heterogeneity as a cause of the lack of CN change. Together, these findings demonstrate that RAD52 and PIF1 are required for extensive array size change in U2OS cells, leading us to conclude that BIR best explains centromere array size alterations in cancer and likely over evolutionary time scales[139].

In summary, we have shown that the human α-Sat CN can expand and contract within ~20 somatic cell divisions with a range from ~7% to 100% change in frequency increasing by up to ~86% in CN. These CN alterations favor expansion over contraction and require RAD52 and PIF1, suggesting that BIR underlies centromere sequence evolution in somatic cells. Better understanding of the mechanisms responsible for rapid centromere sequence evolution provides opportunities to study the consequences of divergence in sequence and length and also identifies the players that are involved in this process. Understanding CN change may support the development of therapeutic strategies for cancer, where centromere dysfunction is frequently observed[160].

Limitations of the study

Whereas the ddPCR-based assay is an advanced method that allows us to study centromere biology, it has limitations. First, we used an unstable, ALT-positive cell line U2OS to provide a sensitized background for detecting CN changes, because ddPCR did not detect CN changes between trophoblast stem cells and their corresponding differentiated cells[137]. Therefore, our results may not be generalizable across all human centromeres in all cell types. Second, we used primers that were designed based on the CHM13 assembly because there is no assembly available for U2OS cells. While this might have limited the absolute HOR CN quantification in U2OS cells, it was sufficient to estimate relative CN changes within 20 cell divisions. Third, the CNs reported represent the average CN per chromosome in the single-cell-derived population, but we are unable to determine when these changes occurred or whether they occurred on one or more homologous chromosomes within the population. Though sufficiently robust to establish CN changes over subsequent clonal generations, the average CN of a highly heterogeneous population could result in false negatives. Fourth, the maximum dynamic range of the ddPCR precludes determination of the upper limits of an array size due to increased technical error when the genomic DNA input is lower than the required amount. These false negatives may cause the array change frequency to be underestimated.

4.4 Method details

Human cell lines

All human cell lines were maintained at 37°C and 5% CO2 in T-75 flasks. K562 cells (Female, ATCC) were cultured in suspension in Iscove's Modified Dulbecco's Medium (IMDM, Thermo Fisher) with 10% heat inactivated fetal bovine serum (FBS, Cytiva). U2OS cells (Female, ATCC) were cultured in Dulbecco's Modified Eagle Medium (DMEM, Thermo Fisher) with 10% FBS, Glutamax, 100 units/mL penicillin, and 100 µg/mL streptomycin. CHM13hTERT cells (hydatidiform mole, homozygous diploid with two X chromosomes, Magee-Womens Hospital) were cultured in basal medium with Amnio Max C100 1X (Thermo Fisher) and the Amnio Max C100 supplement (Thermo Fisher). HEK293T cells (originating from a female embryo, ATCC) were cultured in DMEM with GlutaMAX and 100 U/mL antibiotic-antimycotic (Thermo Fisher).

Single cell isolation

All parental cells were diluted to place 0.5 cells/well into 96-well plates. First, single cells (Group0) were isolated from the U2OS and K562 population cells and grown until 100% confluence in a 12-well plate (~500,000 cells), which is estimated to be ~20 cell divisions. Subsequently subclones (Group1) were isolated into 96-well plates from a clone in Group0 and underwent ~20 somatic cell divisions. Group2 subclones were isolated from either SC3 and SC4 of Group1 and underwent ~20 cell divisions. Group3 subclones were isolated from either SC3.3 or SC4.4 of Group2 and underwent ~20 somatic cell divisions.

DNA extraction, quantification, and dilution

Genomic DNA of subclones from a 12-well plate along with 2 million parental cells were extracted using DNeasy Blood & Tissue Kit (QIAGEN) following the manufacturer's instructions. Genomic DNA samples were quantified using the dsDNA ultra-high sensitivity fluorescent assay (DeNovix) and diluted to 2 ng/μL (single gene CN measurement). Two ng/μL samples were diluted 1:20 to ~ 0.1 ng/μL (HOR and 5S CN measurements). Extracted gDNAs were kept at -20°C.

Centromeric α-satellite repeats measurement by ddPCR

All primer sequences are listed in the Key Resources Table. The four different chromosome-specific HOR array primers (D6Z1, D11Z1, D18Z1, DXZ1) and single gene primers (TBP1, C11orf16, MRO, HPRT1) were used for single-copy assays. For the 5S assay, the same HOR primer sets from the single-copy assay were used for HOR amplification along with a 5S primer set. Two separate probes of different color were used to target the HOR and 5S amplicons. All HOR copy numbers were measured by ddPCR following the manufacturer's protocol (Bio-Rad). For the single-copy (EvaGreen) assay, each reaction contained 10 μL of 2X ddPCR EvaGreen Supermix, 0.2 μL of restriction enzyme (Alu I or HaeIII), 1 μL of 2 μM primer mix, 1 μL of 0.1 ng DNA (for HORs) or 2 ng DNA (for single copy genes) and 7.8 μL of nuclease-free water. For the 5S probe assay, each reaction contained 10 μL of 2X ddPCR Supermix for Probes (No dUTP), 0.2 μL of restriction enzyme (Alu I), 1 μL of 20X HEX target primer/probe mix (900 nM /250 nM), 1 μL of 20X FAM target primer/probe mix (900 nM /250 nM), 1 μL of 0.1 ng DNA, and 6.8 μL of nuclease-free water. The reactions were incubated at

room temperature for 30 min, emulsified with either EvaGreen or a probe droplet generator oil using an automated droplet generator (Bio- Rad), and then transferred to a 96-well plate. The plate was heat-sealed with foil (Bio-Rad) and then a thermocycling reaction was performed using the following temperature profile, where a 2°C /sec ramp rate was applied to all steps: The EvaGreen assay used a 10 min enzyme activation step at 95°C 40 cycles containing a 30 sec denaturation at 96°C and a 60 sec annealing/extension at 56°C, followed by sequential 5 min signal stabilization at 4°C and 90°C and a hold at 4°C. The 5S probe assay used a 10 min enzyme activation step at 95°C, 40 cycles containing a 30 sec denaturation at 94°C and a 60 sec annealing/extension with 2°C/sec ramp rate at 56°C, followed by 10 min enzyme deactivation at 98°C and held at 4°C. Upon completion of PCR, the 96-well plate was transferred to a QX200 droplet reader (Bio-Rad). For the single-gene assay, QuantaSoft software calculated either the HOR or single gene copies/μL. For the 5S assay, the ratio of HOR CN to 5S was calculated automatically by QuantaSoft.

Generation of CRISPR-Cas9 knockdown cells

Two sgRNA oligonucleotide probes targeting different sites in human PIF1 and RAD52 or non-target were cloned into lentiCRISPRv2 puro (Addgene). Plasmids that contain each sgRNA were transfected to HEK293T cells using Lenti-X packaging single shots (Takara) for viral packaging. The Lentivirus was harvested at 48 and 72 h after transfection, combined, and centrifuged. The supernatants were concentrated using a Lenti-X concentrator (Takara) according to the manufacturer's instructions. Viral titers were calculated using Lenti-X

GoStix Plus (Takara) according to the manufacturer's instructions. U2OS cells were transduced with a lentivirus containing polybrene and selected using 1 µg/mL of puromycin for 3 days followed by single clone isolation. Knockdown efficiency of a protein in single cell-derived clones was measured by western blotting.

Western blotting

Cells that were harvested from a confluent 6-well plate were lysed in RIPA buffer (25 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP-40) supplemented with cOmplete protease inhibitors (Roche) and incubated on ice. Cells were sonicated for 10 sec at 30% amplitude twice and the supernatant was retained after centrifugation. Proteins were quantified using a Pierce BCA protein assay kit. A SDS buffer containing 5% beta- mercaptoethanol (Bio-Rad) was added to the samples. Samples were heated at 95°C and electrophoresed on a 4-12% of Tris-Glycine gel. The gel was transferred to a nitrocellulose membrane. The membrane was blocked with Superblock blocking buffer (Thermo Fisher) and probed for RAD52 (1:100, Santa Cruz), PIF1(1: 100, Santa Cruz), and histone H3 (1: 1000, Cell Signaling Technology). After secondary antibody incubation (1: 20,000, IRDYE 800 donkey anti- mouse IgG, IRDYE 680 goat anti-rabbit IgG), the membrane was imaged using Odyssey imager.

Quantification and statistical analysis

All the statistical analysis details of experiments are included in the figure legend. For the single-gene assay, the HOR CN was normalized using the single gene copy number as follows (HOR copies per μL/single gene copies per μL)×20 (dilution factor). For the 5S assay, the HOR copy number was normalized using the 5S copy numbers multiplied by 5S CN per chromosome as follows. (HOR copies per μL/5S copies per μL)× (5S copies per μL/single gene copies per μL). The D11Z1 copy numbers of subclones were normalized to the mean of parental values. ANOVA tests were conducted among HOR copy numbers of subclones and then significance in HOR copy number change was determined by a Tukey HSD test that compared a pair of HOR copy numbers of a subclone and the value of its parental cells. The P-value is indicated as *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001. All statistical analyses and graphs were performed within the RStudio which is an integrated development environment for R.

4.5 Figures

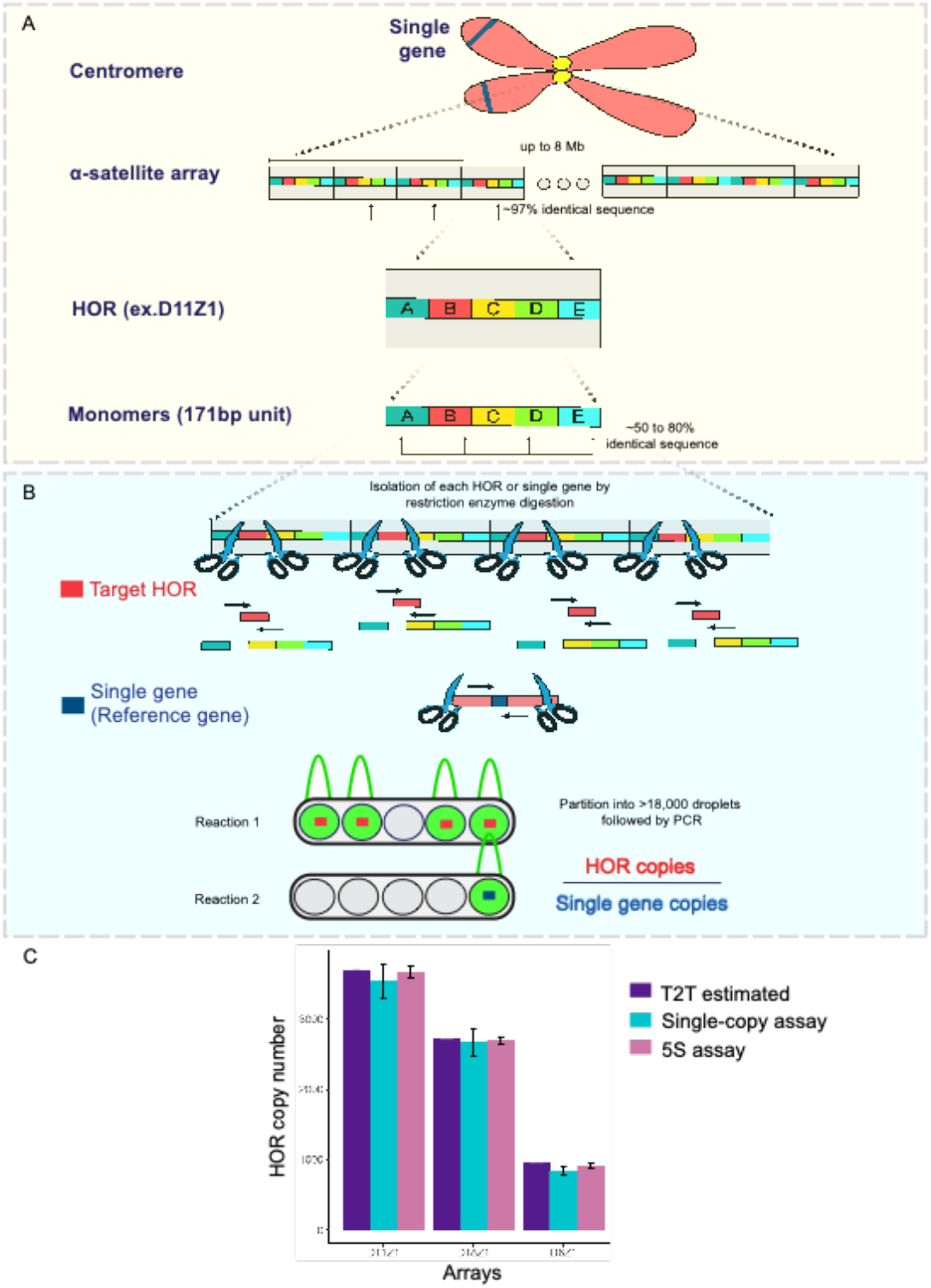

A
Single gene
Centromere
α-satellite array
up to 8 Mb
~97% identical sequence
HOR (ex.D11Z1)
A B C D E
Monomers (171bp unit)
A B C D E
~50 to 80% identical sequence
B
Isolation of each HOR or single gene by restriction enzyme digestion
Target HOR
Single gene (Reference gene)
Reaction 1
Reaction 2
Partition into >18,000 droplets followed by PCR
HOR copies
Single gene copies
C
HOR copy number
Arrays
T2T estimated
Single-copy assay
5S assay

Figure 4.1. Organization of human centromeres and copy number quantification using ddPCR-based assays.

(A) Schematic of the human centromere. Higher-order-repeats (HOR, grey box) comprise tandemly oriented 170-171 bp monomers (colored boxes). Specific HOR copy number (CN) can be quantified based on sequence identities between HORs and polymorphisms present in monomers. (B) Schematic of single-copy assay workflow. Each HOR (red box) or single-copy reference gene on the same chromosome (dark blue box) in a subclone is isolated by restriction enzyme digestion, partitioned into >18,000 droplets, and simultaneously amplified using HOR-specific or single-gene primers (black arrows) in separate reactions. The droplets that contain targets (green peaks) are counted by signal amplitude and the CN is calculated. The HOR CN per array is determined by normalization with single-gene copies (e.g. HOR copies/single-gene copies). (C) Histogram showing HOR CNs of D11Z1, D18Z1, and D6Z1 in the CHM13 cell line either measured by the single-copy assay or the 5S assay. Values represent mean $\pm$ SD of three independent measurements. For the 5S assay, the CNs of the HOR and 5S were measured and the HOR CN per 5S CN were determined. Next, the 5S and a single gene located on the same chromosome were measured to calculate the 5S CN per chromosome. Finally, this number is multiplied by the HOR CN per 5S CN to calculate the HOR CN per chromosome.

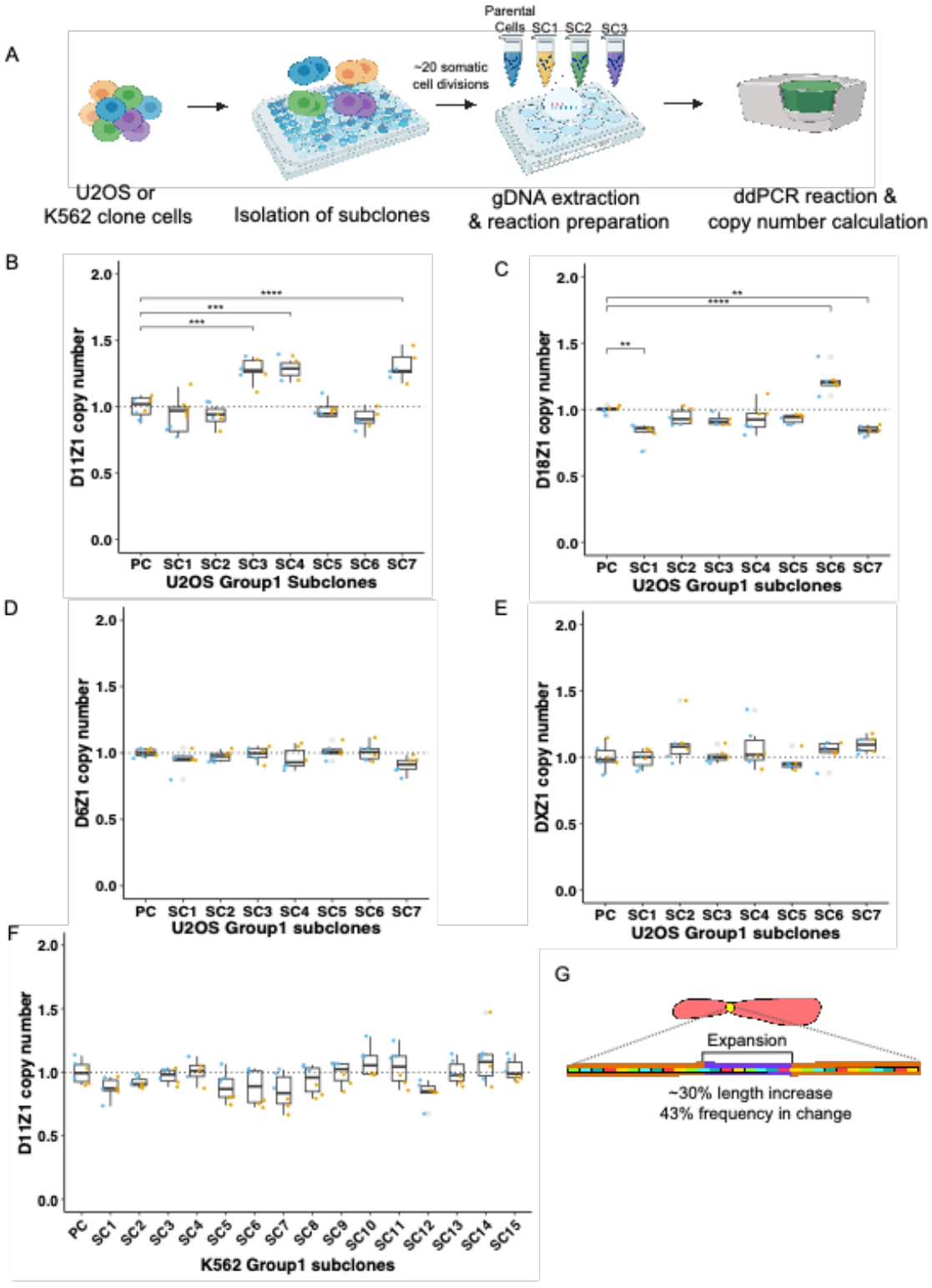

A
Parental Cells
SC1
SC2
SC3
~20 somatic cell divisions
U2OS or K562 clone cells
Isolation of subclones
gDNA extraction & reaction preparation
ddPCR reaction & copy number calculation
B
D11Z1 copy number
U2OS Group1 Subclones
C
D18Z1 copy number
U2OS Group1 subclones
D
D6Z1 copy number
U2OS Group1 subclones
E
DXZ1 copy number
U2OS Group1 subclones
F
D11Z1 copy number
K562 Group1 subclones
G
Expansion
~30% length increase
43% frequency in change

Figure 4.2. Centromere arrays can expand and contract within ~20 somatic cell divisions.

(A) Experimental scheme to quantify HOR CN within somatic cell divisions using single-copy ddPCR-based assay (image created with BioRender.com). (B-E) Box-whisker plots showing the D11Z1, D18Z1, D6Z1, and DXZ1 CNs in U2OS Group1 subclones. In these and subsequent box-whisker plots, each dot indicates a single PCR reaction, which is normalized by the mean of the parental cell (PC) HOR CN (dotted line). Colors indicate technical replicates. Asterisks indicate degree of significance in CN changes between parental cells and subclone pairs determined by Tukey's HSD test (n=8, Tukey's HSD, $P<0.05$). (F) Box-whisker plot showing the D11Z1 CN in K562 Group1 subclones. (n=16, Tukey's HSD, $P>0.05$). (G) Cartoon summary of CN changes in U2OS Group1 subclones.

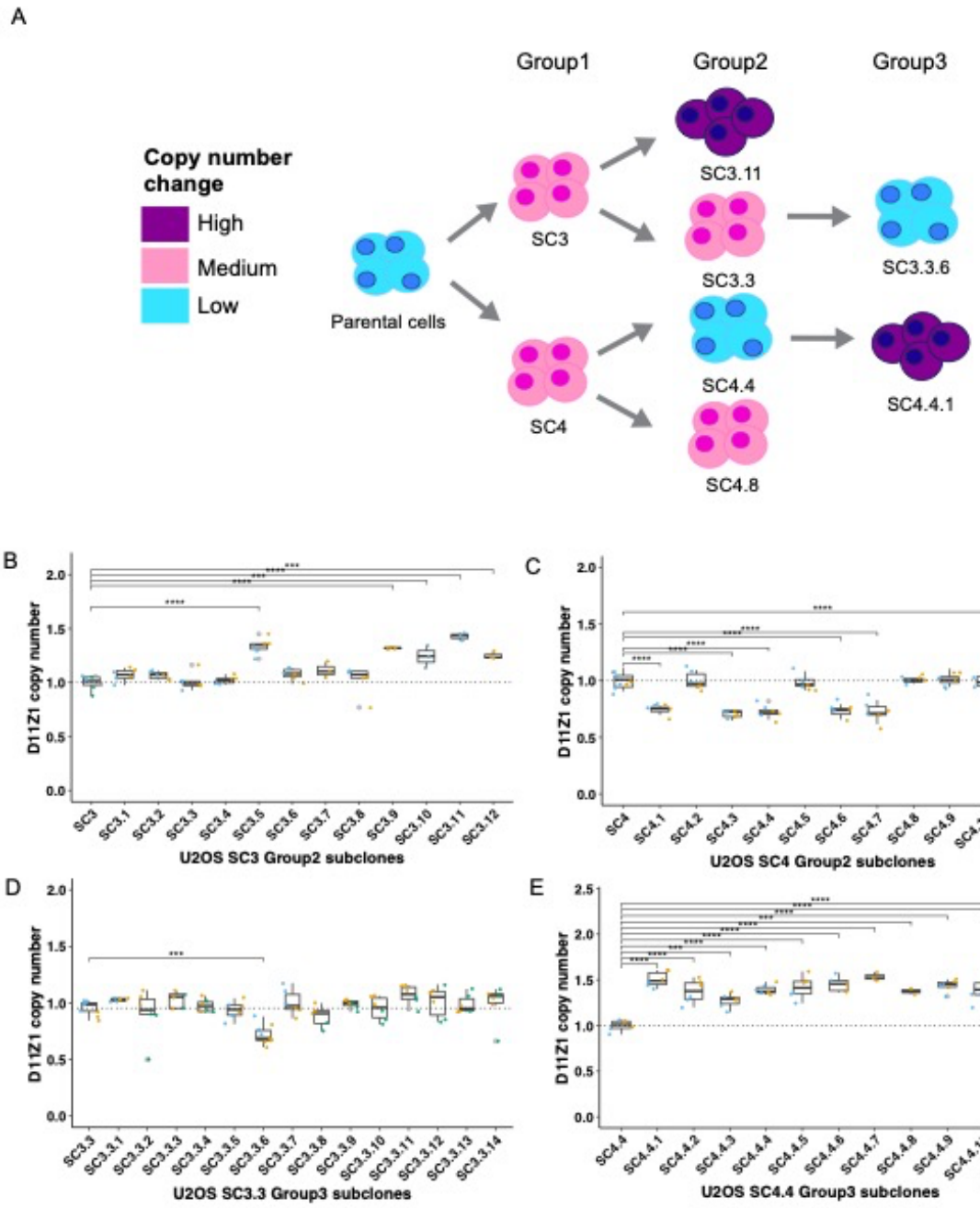

A
Group1
Group2
Group3
Copy number change
High
Medium
Low
Parental cells
SC3
SC4
SC3.11
SC3.3
SC4.4
SC4.8
SC3.3.6
SC4.4.1
B
D11Z1 copy number
U2OS SC3 Group2 subclones
C
U2OS SC4 Group2 subclones
D
U2OS SC3.3 Group3 subclones
E
U2OS SC4.4 Group3 subclones

Figure 4.3. Expansion of centromere arrays is favored over contraction.

(A) Schematic of single cell isolation and D11Z1 CN changes over time. Relative magnitudes of D11Z1 CN changes are indicated by colors. (B-E) Box-whisker plots showing D11Z1 CNs in SC3 Group2 (n=13, Tukey's HSD, $P<0.05$), SC4 Group2 (n=12, Tukey's HSD, $P<0.05$), SC3.3 Group3 (n=15, Tukey's HSD, $P<0.05$), and SC4.4 Group3 (n=12, Tukey's HSD, $P<0.05$) subclones. Individual subclones are identified as follows: parental cell name followed by a period and subclone number (e.g. SC3.3).

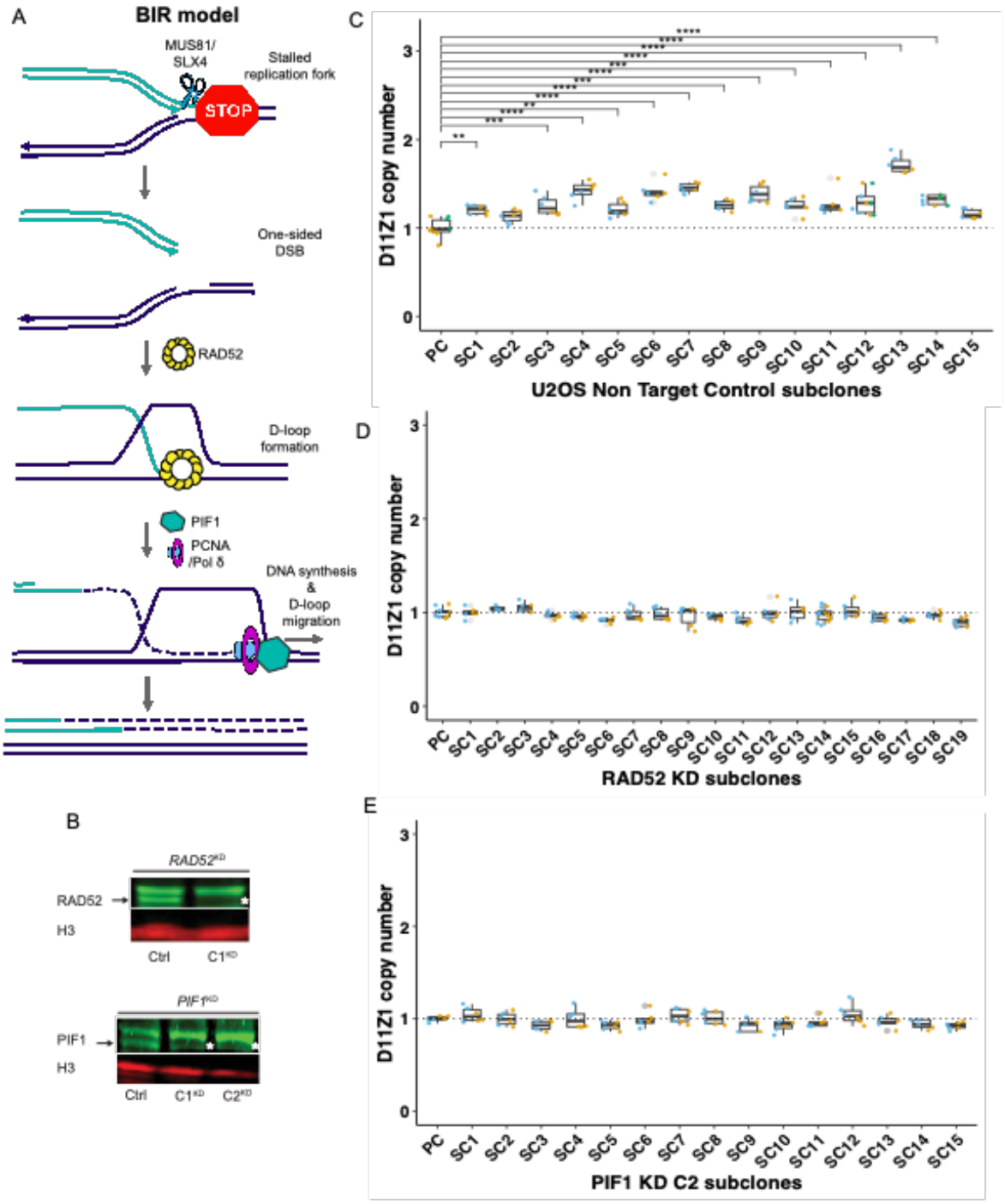

A
BIR model
MUS81/
SLX4
Stalled
replication fork
STOP
One-sided
DSB
RAD52
D-loop
formation
PIF1
PCNA
/Pol δ
DNA synthesis
&
D-loop
migration
B
RAD52KD
RAD52
H3
Ctrl
C1KD
PIF1KD
PIF1
H3
Ctrl
C1KD
C2KD
C
D11Z1 copy number
U2OS Non Target Control subclones
D
D11Z1 copy number
RAD52 KD subclones
E
D11Z1 copy number
PIF1 KD C2 subclones

Figure 4.4. RAD52 and PIF1 are required for D11Z1 CN changes.

(A) Overview of the break-induced replication (BIR) model. (B) Expression level of RAD52 (top) and PIF1 (bottom) detected by western blot analysis. Arrows indicate WT bands present in the non-target control. Asterisks indicate KD. The band above the asterisks or arrows are non-specific bands. (C-E) Box-whisker plot showing the D11Z1 CN in either NTC (n=16, Tukey's HSD, $P<0.05$), $RAD52^{KD}$ C1 (n=20, Tukey's HSD, $P>0.05$), or $PIF1^{KD}$ C2 clones (n=16, Tukey's HSD, $P>0.05$).

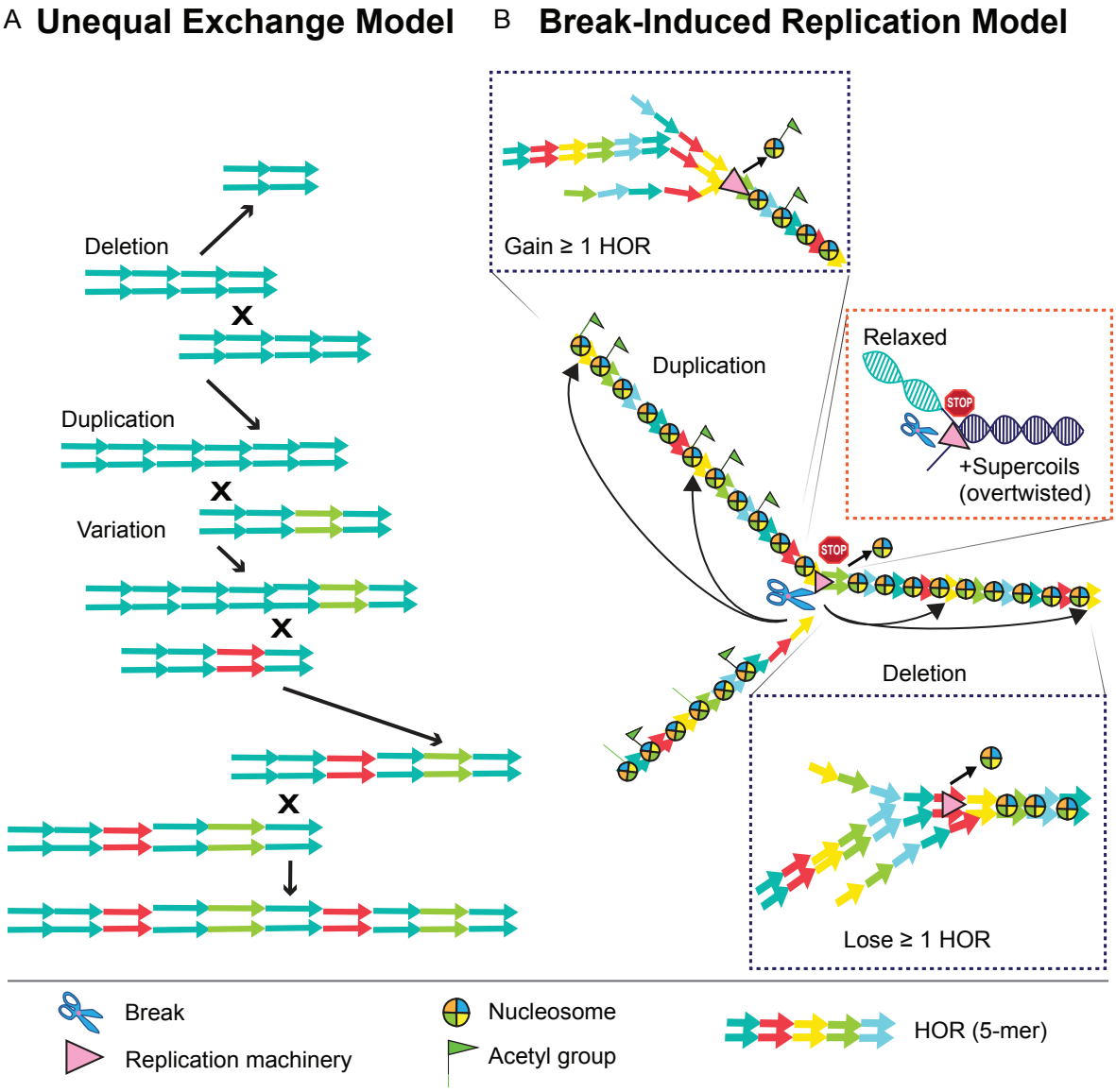
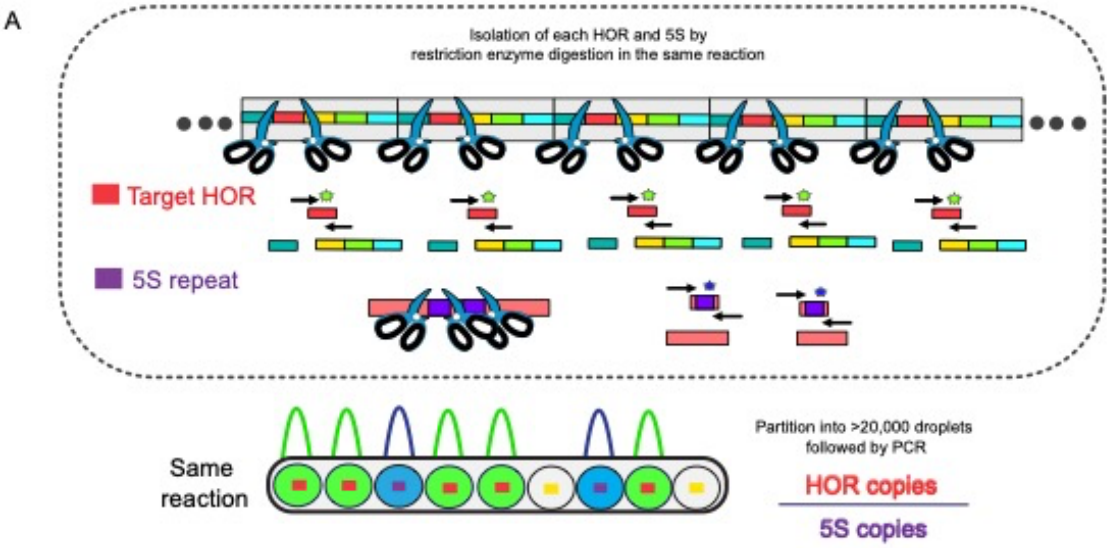
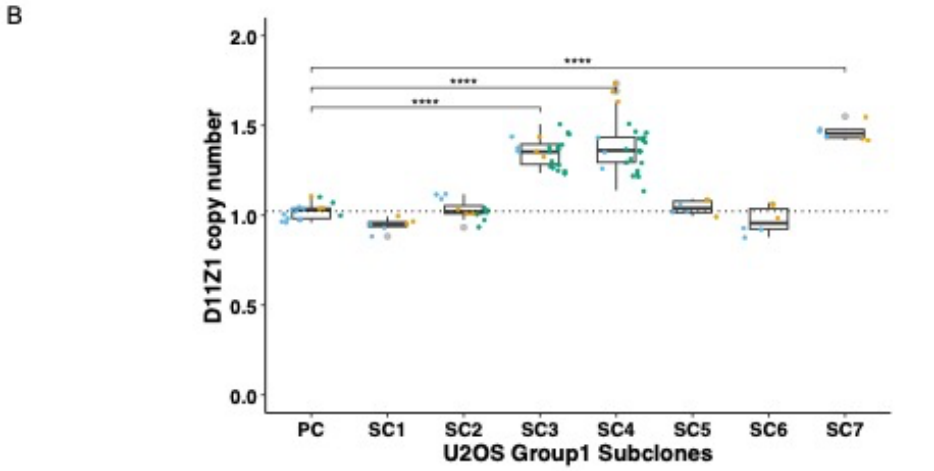

A Unequal Exchange Model
B Break-Induced Replication Model
Deletion
Duplication
Variation
Gain ≥ 1 HOR
Duplication
Relaxed
STOP
+Supercoils
(overtwisted)
STOP
Deletion
Lose ≥ 1 HOR
Break
Replication machinery
Nucleosome
Acetyl group
HOR (5-mer)

Figure S4.1. Schematic of two possible models that can lead to higher-order-repeats (HOR) copy number variation, relating to Introduction.

(A) In the unequal exchange model, tandem repeats such as alpha satellites can change their copy numbers by unequal exchange recombination between out-of-register paired sister chromatids resulting in either deletion or duplication of monomers. HORs can be generated when out-of-register unequal exchange occurs repeatedly between monomers that have variations. (B) In the BIR model, HORs can lead to either duplication or deletion of copies depending on the location of out-of-register re-initiation of replication with respect to the collapsed replication fork during the one-ended double strand break repair (blue dotted boxes). Duplication might occur more frequently than deletion because the chromatin behind the fork is more accessible to strand invasion owing to the new acetylated histones and/or the relaxed torsional state, in contrast to the overtwisted DNA ahead of the fork (orange dotted box).


A
Isolation of each HOR and 5S by
restriction enzyme digestion in the same reaction
Target HOR
5S repeat
Same
reaction
Partition into >20,000 droplets
followed by PCR
HOR copies
5S copies


B
D11Z1 copy number
2.0
1.5
1.0
0.5
0.0
****
****
****
PC
SC1
SC2
SC3
SC4
SC5
SC6
SC7
U2OS Group1 Subclones

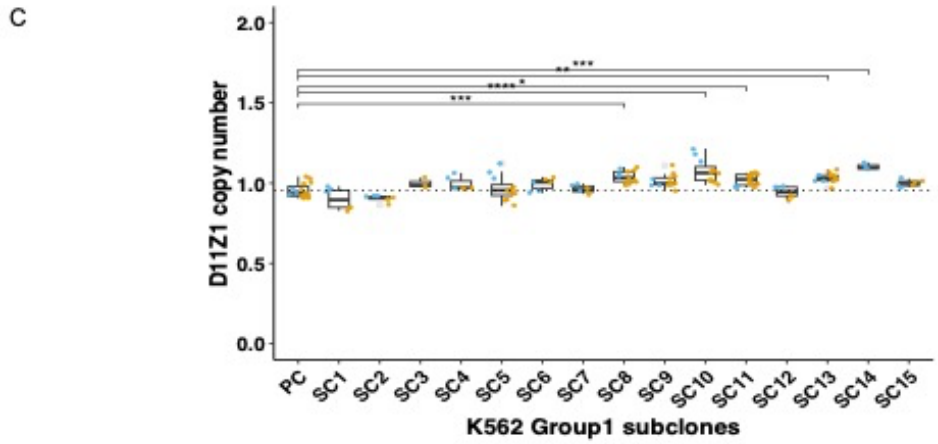
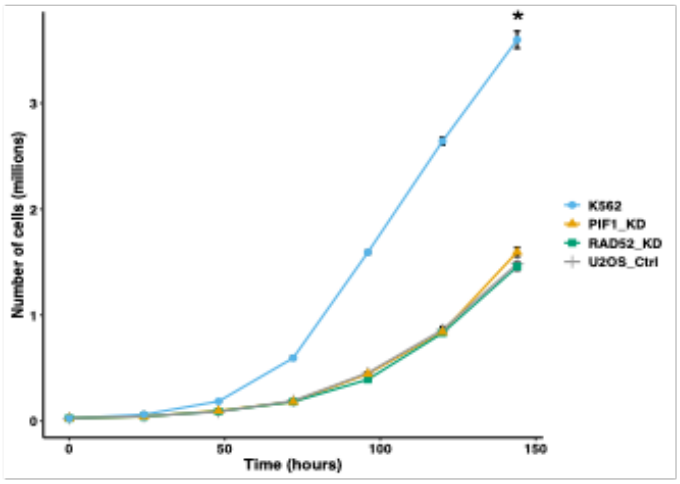

C
D11Z1 copy number
2.0
1.5
1.0
0.5
0.0
***
****
***
PC
SC1
SC2
SC3
SC4
SC5
SC6
SC7
SC8
SC9
SC10
SC11
SC12
SC13
SC14
SC15
K562 Group1 subclones

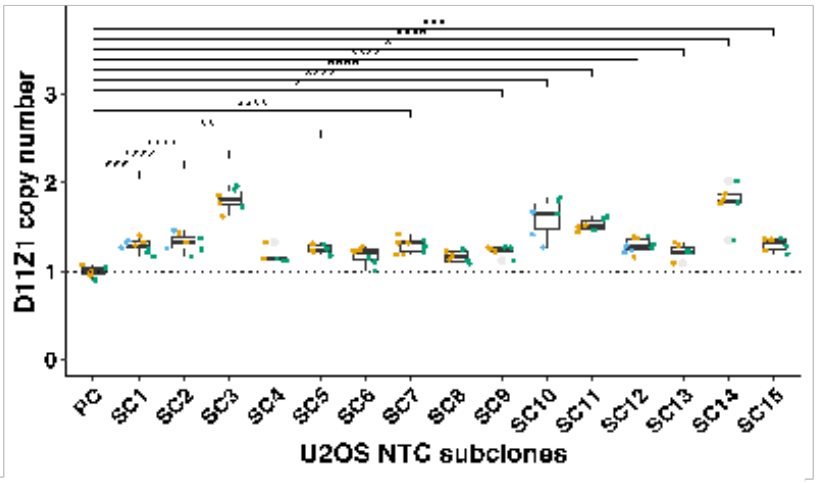
Figure S4.2. D11Z1 copy numbers using 5S assay, relating to Figure 2.

(A) Schematic of 5S assay workflow. Each HOR (red square) and 5S array (purple square) from the genome that are isolated by restriction enzyme digestion are partitioned into over 20,000 droplets. Both HOR and 5S targets are simultaneously bound with the corresponding probes and amplified in the same reaction. The droplets that containing target are measured for the two different signal amplitudes and the HOR CN per 5S CN is calculated by Bio-Rad QuantaSoft with 95% confidence intervals. (B) Box-whisker plots showing the D11Z1 CN in U2OS Group1 subclones using the 5S assay. Each dot indicates a single PCR reaction, which is normalized by the mean of the parental cell (PC) HOR CN (dotted line). Colors indicate technical replicates. Asterisks indicate degree of significance in CN changes between parental cells and subclone pairs determined by Tukey's HSD test (n=8, Tukey's HSD, $P<0.05$) (C) Box-whisker plots showing the D11Z1 CN in K562 Group1 subclones using the 5S assay. (n=16, Tukey's HSD, $P<0.05$).

A


Number of cells (millions)
Time (hours)
K562
PIF1_KD
RAD52_KD
U2OS_Ctrl
*

B


D11Z1 copy number
U2OS NTC subclones
PC
SC1
SC2
SC3
SC4
SC5
SC6
SC7
SC8
SC9
SC10
SC11
SC12
SC13
SC14
SC15

C

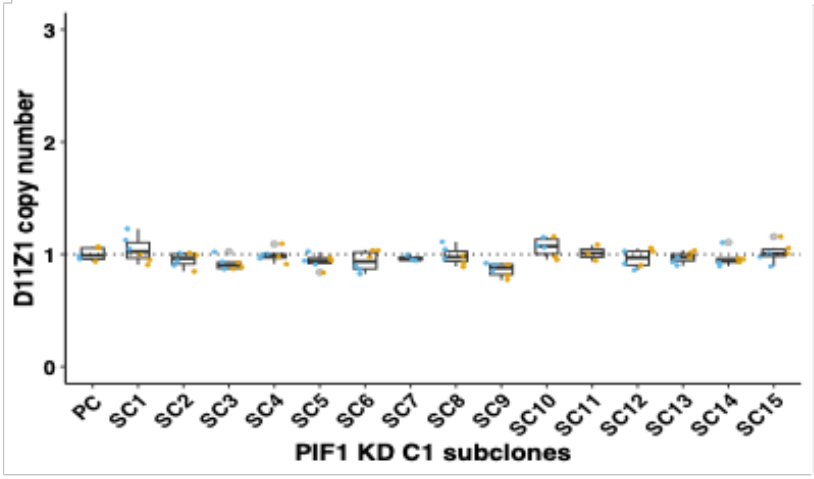

D11Z1 copy number
PIF1 KD C1 subclones
PC
SC1
SC2
SC3
SC4
SC5
SC6
SC7
SC8
SC9
SC10
SC11
SC12
SC13
SC14
SC15

Figure S4.3. No D11Z1 copy number change occurs in another PIF1 KD that has the same growth rate as U2OS control, relating to Figure 4.

(A) A line plot showing the cell numbers of U2OS control, RAD52 KD, PIF1 KD, and K562 at six time points. The colored symbols indicate the mean cell numbers of cell lines in three replicates and the bars represent ±SEM. The asterisk indicates that the K562 cell line grows differently than the other cell lines, as determined by Tukey's HSD (n=3, Tukey's HSD, $P<0.05$). (B-C) Box-whisker plot showing the D11Z1 CN in (B) NTC subclones (n=16, Tukey's HSD, $P<0.05$) and (C) PIF1^{KD} C1 subclones (n=16, Tukey's HSD, $P>0.05$).

Chapter 5: Discussion

My thesis work has expanded our understanding of a driving force in rapid centromere sequence evolution, for which the mechanisms of such a complex and dynamic process have been only hypothesized. In my doctoral research, I generated and optimized a protocol to quantify α-satellite length (HOR copy number) change that occurs within clonal evolution in a sensitized cellular system. With this protocol, I established the frequency of centromeric array change within limited somatic cell divisions and experimentally validated the BIR model for array size changes. In addition, I characterized array size change in a chromosome specific manner and discovered that the expansion of arrays occurs more frequently than contraction, yet their continuous expansion is controlled by an observed but unknown mechanism. This work has afforded many new opportunities for future research. Here, I will discuss my findings and ideas of great interest including how relevant array changes are to evolution in the centromere, what mediates these changes, how the array size is maintained, and what other types of variations occur via BIR. These studies will boost the ongoing renaissance in centromere biology.

5.1 Mechanisms of centromeric array changes in evolution

My thesis work has shown that centromeric array length changes in somatic cells occur differently between chromosome specific arrays. Chromosome specific variation manifests as different forms of centromere diversity such as sequence, length, and epigenetics due to diverging evolutionary trajectories[5]. Therefore, it is expected to observe that some arrays would have a higher frequency and/or magnitude compared to others within the limited time of clonal evolution, as we observed among subclones of the U2OS

D11Z1 and D18Z1 arrays but not the D6Z1 and DXZ1 arrays. While the frequencies of array changes between D11Z1 and D18Z1 were similar, the magnitude of D11Z1 array size change is significantly higher. Our results match the expected outcomes of evolutionary trajectory, wherein the chromosome specific array DXZ1, which has the highest concordance in sequence and length between individuals did not change copy number, whereas, D11Z1, the largest array observed among all chromosomes of the 56 haplotypes changed at the highest magnitude[2,5,6]. This suggests that the underlying molecular mechanisms that govern these changes are compatible with the forces driving human centromere evolution.

We observed that a different magnitude of change occurred at the D11Z1 array of the U2OS and K562 cell lines within somatic cell divisions. Since genetic outcomes of DSB repairs that can influence centromeric array length are different[31], the extensive array expansion observed at D11Z1 in U2OS is BIR mediated. However, BIR and/or other mechanisms might have acted on D11Z1 in K562. The major difference between the two cell lines is that U2OS has an ATRX mutation that is associated with ALT activation via BIR resulting in telomere extension[161-163]. ATRX is a multifaceted chromatin remodeler that is involved in many biological processes including the maintenance of heterochromatin such as telomeres[164] and pericentromeres[154], the resolution of DNA secondary structures[165,166], the stabilization of stalling replication forks[167,168], and the repair of DSBs[169,170] .

A recent paper has shown that ATRX-deficiency in telomerase positive cells cannot alone activate telomere synthesis via BIR[124,171] unless there are DNA-protein covalent complexes (DPCCs). These include trapped TOP1cc and/or accumulated DNA bound proteins by

PARP1 inhibitors that can cause one-sided DSBs that are a BIR substrate generated via R-loops or SSBs[112,172]. Loss of XRCC1 via IDH1 R132H mutation, which traps PARP1 on SSBs during BER, is responsible for ALT activation in ATRX-deficient cells[173,174]. BIR competes with PARP1-dependent alt-EJ to repair DNA damage lesions that are induced by the loss of BLM helicase, which resolves RS in telomeres[175-177]. Another study demonstrated that unlike PARP1, PARP2 (both members of the ADP-ribosyl transferase super family) can promote BIR-mediated telomere synthesis following BLM loss-induced fork collapse by enhancing the resection of broken DNA ends and by directly interacting with POLD3, which plays an important role in BIR[178 179].

A common observation shared by these studies is that BIR initiation only occurs when there is a significant number of replication obstacles that cause fork stalling and the mechanisms that restart these forks by efficient SSB responses are compromised, leading to one-sided DSBs that are mainly repaired by BIR. It is unclear whether a specific mechanism of ATRX or multiple mechanisms working together lead to BIR. ATRX's involvement in replication fork recovery, the resolution of secondary structures, and the repair of DSBs that compete with BIR may collectively create an unfavorable environment for BIR to win over different HR choices at the telomere. Therefore, it would be interesting to test whether ATRX KO in K562 alone would increase centromeric array size extensively via BIR, as observed in U2OS, since centromeres are already enriched with replication obstacles such as centromere binding proteins and non-B DNA structures such as R-loop and G4. In addition, most DSBs and SSBs are localized at the centromere, even in physiological conditions, unlike telomeres or other repeat regions which makes

centromere recombination hotspots[84]. R-loops are accumulated at the centromere in ATRX or DAXX deficient cells which recruits BRCA1 at centromeres[169]. BRAC1 recruitment to the centromere in response to R-loop accumulation is critical to prevent RAD52 dependent recombination[180]. ALT positive cells have a higher level of TOP1ccs compared to ALT negative cells naturally[181]. The favorable conditions for R loop formations are persistent TOP1ccs[182], replication stress, transcription-replication fork conflicts, and DNA damage. Centromeres provide a perfect environment to form R-loop-mediated one sided DSBs because centromere transcription-replication machinery will likely collide at a high level because centromeres use most of their available replication origins leading to a high frequency of converging forks. This collision and formation of R-loops will lead to replication stress and DNA damage that is exacerbated by further R-loop formation. Telomeres and centromeres are regions that are difficult to replicate but replication stress is only associated with ALT telomeres. To achieve BIR activation in non-ALT telomeres, it is necessary to induce replication stresses that have the potential to produce BIR substrates and ATRX KO. However, the intrinsic high level of DSBs and SSBs that are present at centromeres indicate that, if ATRX is not present, the centromeres possess many other substrates that allow BIR to outcompete other HR mechanisms. It would be of great interest for future studies to test whether the overexpression of RNase H would effect centromere array size alteration in U2OS using structured illumination microscopy (SIM), which offers resolution to distinguish a different array size[183]. ATR activation driven by R-loop bound RPA is important for accurate cell divisions in mitosis[146]. Therefore, it is possible that unresolved R-loops enter mitosis and are cleaved by MUS81, an

endonuclease that is critical for BIR, resulting in one-sided DSBs that trigger MiDAS at the entry of mitosis.

5.2: Characteristics of α-satellite array changes and array size maintenance

Expansion of arrays are observed more frequently than contraction in yeast rDNA repeats and our own results[21,22,83]. This expansion bias, which counteracts centromere shrinkage that could occur due to SSA, can lead to enormous arrays that are well beyond what is functionally required (50K-100K) in human centromeres[184,185]. However, this expansion does not seem to continue indefinitely. Intriguingly, we observed extensive array loss, which trimmed the array to a more manageable length, only when the subclones had evolved from parental cell lines that already had a high HOR copy number. Since DNA replication is difficult in centromeres, it would be harmful to maintain extremely large arrays that can become vulnerable to chromosome mis-segregation. rDNA arrays and telomeres possess repeat length homeostasis mechanisms to address this type of length issue[157,186,187]. Telomere length homeostasis is especially important to ensure repeat lengths are long enough to undergo controlled cell divisions, but not so long that the cells become cancerous[187]. Telomere homeostasis in stem cells and ATL positive cells that have the capacity to elongate extensive telomeres is regulated by a trimming mechanism to prevent the burden of replicative stress associated with an excessive telomere size[187]. One trimming mechanism utilized by ALT cells is the telomeric zinc finger-associated protein (TZAP) which specially binds to long telomeres when the local concentration of shelterin complex is low and rapidly reduces their length. As the local concentration of shelterin

monitors telomere copy number, the concentration of rDNA bounded upstream activating factor (UAF) senses the rDNA copy number and maintains the optimal number mediated by SIR2[188]. Tandem repeats such as α-satellites and rDNA are prone to break and easy to lose copy numbers as a result of unequal crossover during HR repair. Therefore, rDNA operates a cycling system which oscillates between contraction and expansion to maintain an optimal level of rDNA repeat copies[32]. The expansion of rDNA, which is transcriptionally controlled by Sir2 in budding yeast *S. cerevisiae*, occurs by allowing the recombination of induced forks collapsed by Fob1 at the replication fork barrier between out-of-register sister chromatids[32,157]. It would be of great interest for future studies to determine whether discrete processes exist to mediate centromere length homeostasis.

5.3: BIR at the centromere

My thesis work has experimentally validated the hypothesis that BIR can drive centromere sequence evolution such as variations in length and sequence, complex organization, and the extreme lengths observed from all human centromeres. The substitution mutation rate at centromeres is ~100 fold higher than other genomic regions[5]. BIR is a conserved mutagenic repair mechanism with a high substitution rate that is ~100 to 1000 fold higher than the canonical S-phase mutation rate[149]. Therefore, it could be exciting to measure the substitution mutation rate of arrays that have expanded via BIR in U2OS cells.

BIR is best known in telomere maintenance of ATL positive cells and MiDAS at CSFs in mammalian cells[27,123]. While BIR occurs in the G_2/M phases, both ALT telomere elongation and MiDAS, as the name indicates, occur in early mitosis, acting as a last step to rescue

stalled forks and prevent chromosome mis-segregation[27,122,123]. The spindle assembly checkpoint (SAC) is activated by DNA damage at the centromere[189]. Acute inhibition of MiDAS turns on a spindle checkpoint which delays anaphase onset and causes centromere breaks suggesting that MiDAS occurs to prevent DNA damages at centromeres before anaphase[190]. BIR mediated MiDAS at the centromere has not been experimentally tested. However, MiDAS that are labeled by a thymidine analogue EdU are often localized near centromeres, but without a centromere specific marker, leading to the speculation that centromere DNA synthesis can occur during mitosis[191]. Sister chromatid exchanges (SCE), which represent recombination events, occur at the centromeres during mitosis. The inhibition of SCEs lead to a high level of ssDNAs, intermediates of BIR, indicating the possibility that BIR is active [190,192]. Therefore, future studies to determine BIR mediated MiDAS and identify sources of BIR substrates will help us understand the effects of α-satellite sequence and size variations.

A recent paper has shown that active centromeres of vertebrates are divided into two sub-domains during mitosis where both domains can be bound by microtubules[193]. These bipartite domains are stabilized by cohesin and prone to form merotelic attachments that cause lagging chromosomes when the two sub-domains are separated. These large separations have been found in a colorectal cancer known for chromosome instability[193], highlighting the importance in organization of active centromere higher order structures which might be influenced by array size differences. Future studies investigating how bipartite domains are regulated across different chromosome-specific arrays and between

the same chromosomes of different lengths, will provide more insight into the connection between α-satellites and their 3D organization.

References

1. Eichler, E.E., Clark, R.A., and She, X. (2004). An assessment of the sequence gaps: Unfinished business in a finished human genome. Nature Reviews Genetics *5*, 345-354. 10.1038/nrg1322.
2. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2022). The complete sequence of a human genome. Science *376*, 44-53. https://doi.org/10.1126/science.abj6987.
3. 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. (2022). Complete genomic and epigenetic maps of human centromeres. Science *376*, eabl4178. https://doi.org/10.1126/science.abl4178.
4. Liao, W.W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J.K., Monlong, J., Abel, H.J., et al. (2023). A draft human pangenome reference. Nature *617*, 312-324. 10.1038/s41586-023-05896-x.
5. Logsdon, G.A., Rozanski, A.N., Ryabov, F., Potapova, T., Shepelev, V.A., Catacchio, C.R., Porubsky, D., Mao, Y., Yoo, D., Rautiainen, M., et al. (2024). The variation and evolution of complete human centromeres. Nature *629*, 136-145. 10.1038/s41586-024-07278-3.
6. Makova, K.D., Pickett, B.D., Harris, R.S., Hartley, G.A., Cechova, M., Pal, K., Nurk, S., Yoo, D., Li, Q., Hebbar, P., et al. (2024). The complete sequence and comparative analysis of ape sex chromosomes. Nature *630*, 401-411. 10.1038/s41586-024-07473-2.
7. Rhie, A., Nurk, S., Cechova, M., Hoyt, S.J., Taylor, D.J., Altemose, N., Hook, P.W., Koren, S., Rautiainen, M., Alexandrov, I.A., et al. (2023). The complete sequence of a human Y chromosome. Nature *621*, 344-354. 10.1038/s41586-023-06457-y.
8. McKinley, K.L., and Cheeseman, I.M. (2016). The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol *17*, 16-29. 10.1038/nrm.2015.5.
9. McNulty, S.M., and Sullivan, B.A. (2018). Alpha satellite DNA biology: finding function in the recesses of the genome. Chromosome Res *26*, 115-138. 10.1007/s10577-018-9582-3.
10. Miga, K.H. (2020). Centromere studies in the era of 'telomere-to-telomere' genomics. Exp Cell Res *394*, 112127. 10.1016/j.yexcr.2020.112127.
11. Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., et al. (2020). Telomere-to-telomere assembly of a complete human X chromosome. Nature *585*, 79-84. 10.1038/s41586-020-2547-7.
12. Miga, K.H., Newton, Y., Jain, M., Altemose, N., Willard, H.F., and Kent, W.J. (2014). Centromere reference models for human chromosomes X and Y satellite arrays. Genome Res *24*, 697-707. 10.1101/gr.159624.113.
13. Ahmad, S.F., Singchat, W., Jehangir, M., Panthum, T., and Srikulnath, K. (2020). Consequence of Paradigm Shift with Repeat Landscapes in Reptiles: Powerful Facilitators of Chromosomal Rearrangements for Diversity and Evolution. Genes (Basel) *11*. 10.3390/genes11070827.

14. Prakhongcheep, O., Hirai, Y., Hara, T., Srikulnath, K., Hirai, H., and Koga, A. (2013). Two types of alpha satellite DNA in distinct chromosomal locations in Azara's owl monkey. DNA Res *20*, 235-240. 10.1093/dnares/dst004.
15. Prakhongcheep, O., Thapana, W., Suntronpong, A., Singchat, W., Pattanatanang, K., Phatcharakullawarawat, R., Muangmai, N., Peyachoknagul, S., Matsubara, K., Ezaz, T., and Srikulnath, K. (2017). Lack of satellite DNA species-specific homogenization and relationship to chromosomal rearrangements in monitor lizards (Varanidae, Squamata). BMC Evolutionary Biology *17*, 193. 10.1186/s12862-017-1044-6.
16. Hartley, G., and O'Neill, R.J. (2019). Centromere Repeats: Hidden Gems of the Genome. Genes (Basel) *10*. 10.3390/genes10030223.
17. Alexandrov, I., Kazakov, A., Tumeneva, I., Shepelev, V., and Yurov, Y. (2001). Alpha-satellite DNA of primates: old and new families. Chromosoma *110*, 253-266. 10.1007/s004120100146.
18. de Lima, L.G., Howe, E., Singh, V.P., Potapova, T., Li, H., Xu, B., Castle, J., Crozier, S., Harrison, C.J., Clifford, S.C., et al. (2021). PCR amplicons identify widespread copy number variation in human centromeric arrays and instability in cancer. Cell Genom *1*. https://doi.org/10.1016/j.xgen.2021.100064.
19. Smith, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. Science *191*, 528-535. 10.1126/science.1251186.
20. Schindelhauer, D., and Schwarz, T. (2002). Evidence for a fast, intrachromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous alpha-satellite DNA array. Genome Res *12*, 1815-1826. 10.1101/gr.451502.
21. Rice, W. (2019). A Game of Thrones at Human Centromeres I. Multifarious structure necessitates a new molecular/evolutionary model. bioRxiv. https://doi.org/10.1101/731430.
22. Rice, W. (2019). A Game of Thrones at Human Centromeres II. A new molecular/evolutionary model. bioRxiv. https://doi.org/10.1101/731471.
23. Pâques, F., and Haber, J.E. (1999). Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev *63*, 349-404. 10.1128/mmbr.63.2.349-404.1999.
24. Chakraborty, U., George, C.M., Lyndaker, A.M., and Alani, E. (2016). A Delicate Balance Between Repair and Replication Factors Regulates Recombination Between Divergent DNA Sequences in Saccharomyces cerevisiae. Genetics *202*, 525-540. 10.1534/genetics.115.184093.
25. Bhargava, R., Onyango, D.O., and Stark, J.M. (2016). Regulation of Single-Strand Annealing and its Role in Genome Maintenance. Trends Genet *32*, 566-575. 10.1016/j.tig.2016.06.007.
26. Talbert, P.B., and Henikoff, S. (2022). The genetics and epigenetics of satellite centromeres. Genome Res *32*, 608-615. 10.1101/gr.275351.121.
27. Liu, L., and Malkova, A. (2022). Break-induced replication: unraveling each step. Trends Genet *38*, 752-765. 10.1016/j.tig.2022.03.011.
28. Graham, E., and Esashi, F. (2024). DNA strand breaks at centromeres: Friend or foe? Semin Cell Dev Biol *156*, 141-151. 10.1016/j.semcdb.2023.10.004.

29. Scelfo, A., Angrisani, A., Grillo, M., Barnes, B.M., Muyas, F., Sauer, C.M., Leung, C.W.B., Dumont, M., Grison, M., Mazaud, D., et al. (2024). Specialized replication mechanisms maintain genome stability at human centromeres. Molecular Cell *84*, 1003-1020.e1010. https://doi.org/10.1016/j.molcel.2024.01.018.
30. Ceccaldi, R., Rondinelli, B., and D'Andrea, A.D. (2016). Repair Pathway Choices and Consequences at the Double-Strand Break. Trends Cell Biol *26*, 52-64. 10.1016/j.tcb.2015.07.009.
31. Her, J., and Bunting, S.F. (2018). How cells ensure correct repair of DNA double-strand breaks. J Biol Chem *293*, 10502-10511. 10.1074/jbc.TM118.000371.
32. Kobayashi, T. (2014). Ribosomal RNA gene repeats, their stability and cellular senescence. Proc Jpn Acad Ser B Phys Biol Sci *90*, 119-129. 10.2183/pjab.90.119.
33. Sakofsky, C.J., Roberts, S.A., Malc, E., Mieczkowski, P.A., Resnick, M.A., Gordenin, D.A., and Malkova, A. (2014). Break-induced replication is a source of mutation clusters underlying kataegis. Cell Rep *7*, 1640-1648. 10.1016/j.celrep.2014.04.053.
34. Margolin, W. (2000). Themes and variations in prokaryotic cell division. FEMS Microbiology Reviews *24*, 531-548. 10.1111/j.1574-6976.2000.tb00554.x.
35. Matthews, H.K., Bertoli, C., and de Bruin, R.A.M. (2022). Cell cycle control in cancer. Nat Rev Mol Cell Biol *23*, 74-88. 10.1038/s41580-021-00404-3.
36. Liu, J., Peng, Y., and Wei, W. (2022). Cell cycle on the crossroad of tumorigenesis and cancer therapy. Trends Cell Biol *32*, 30-44. 10.1016/j.tcb.2021.07.001.
37. Potten, C.S., and Loeffler, M. (1990). Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. Development *110*, 1001-1020. 10.1242/dev.110.4.1001.
38. Hall, P.A., and Watt, F.M. (1989). Stem cells: the generation and maintenance of cellular diversity. Development *106*, 619-633. 10.1242/dev.106.4.619.
39. Aganezov, S., Yan, S.M., Soto, D.C., Kirsche, M., Zarate, S., Avdeyev, P., Taylor, D.J., Shafin, K., Shumate, A., Xiao, C., et al. (2022). A complete reference genome improves analysis of human genetic variation. Science *376*, eabl3533. 10.1126/science.abl3533.
40. Pennycook, B.R., and Barr, A.R. (2020). Restriction point regulation at the crossroads between quiescence and cell proliferation. FEBS letters *594*, 2046-2060.
41. Rubin, S.M., Sage, J., and Skotheim, J.M. (2020). Integrating Old and New Paradigms of G1/S Control. Mol Cell *80*, 183-192. 10.1016/j.molcel.2020.08.020.
42. den Elzen, N., and Pines, J. (2001). Cyclin A is destroyed in prometaphase and can delay chromosome alignment and anaphase. J Cell Biol *153*, 121-136. 10.1083/jcb.153.1.121.
43. Geley, S., Kramer, E., Gieffers, C., Gannon, J., Peters, J.M., and Hunt, T. (2001). Anaphase-promoting complex/cyclosome-dependent proteolysis of human cyclin A starts at the beginning of mitosis and is not subject to the spindle assembly checkpoint. J Cell Biol *153*, 137-148. 10.1083/jcb.153.1.137.
44. Malumbres, M. (2014). Cyclin-dependent kinases. Genome Biology *15*, 122. 10.1186/gb4184.

45. Bertoli, C., Skotheim, J.M., and de Bruin, R.A.M. (2013). Control of cell cycle transcription during G1 and S phases. Nature Reviews Molecular Cell Biology *14*, 518-528. 10.1038/nrm3629.
46. Simmons Kovacs, L.A., Orlando, D.A., and Haase, S.B. (2008). Transcription networks and cyclin/CDKs: the yin and yang of cell cycle oscillators. Cell Cycle *7*, 2626-2629. 10.4161/cc.7.17.6515.
47. Batty, P., and Gerlich, D.W. (2019). Mitotic Chromosome Mechanics: How Cells Segregate Their Genome. Trends Cell Biol *29*, 717-726. 10.1016/j.tcb.2019.05.007.
48. Sullivan, M., and Morgan, D.O. (2007). Finishing mitosis, one step at a time. Nature Reviews Molecular Cell Biology *8*, 894-903. 10.1038/nrm2276.
49. Taubenberger, A.V., Baum, B., and Matthews, H.K. (2020). The Mechanics of Mitotic Cell Rounding. Front Cell Dev Biol *8*, 687. 10.3389/fcell.2020.00687.
50. Venkei, Z., Przewloka, M.R., Ladak, Y., Albadri, S., Sossick, A., Juhasz, G., Novák, B., and Glover, D.M. (2012). Spatiotemporal dynamics of Spc105 regulates the assembly of the Drosophila kinetochore. Open Biol *2*, 110032. 10.1098/rsob.110032.
51. Gavet, O., and Pines, J. (2010). Activation of cyclin B1-Cdk1 synchronizes events in the nucleus and the cytoplasm at mitosis. J Cell Biol *189*, 247-259. 10.1083/jcb.200909144.
52. Musacchio, A., and Desai, A. (2017). A Molecular View of Kinetochore Assembly and Function. Biology (Basel) *6*. 10.3390/biology6010005.
53. Nagaoka, S.I., Hassold, T.J., and Hunt, P.A. (2012). Human aneuploidy: mechanisms and new insights into an age-old problem. Nat Rev Genet *13*, 493-504. 10.1038/nrg3245.
54. Willard, H.F. (1992). Centromeres-primary constrictions are primarily complicated. Human Molecular Genetics *1*, 667-668. 10.1093/hmg/1.9.667.
55. Clarke, L., and Carbon, J. (1983). Genomic substitutions of centromeres in Saccharomyces cerevisiae. Nature *305*, 23-28. 10.1038/305023a0.
56. Talbert, P.B., and Henikoff, S. (2020). What makes a centromere? Exp Cell Res *389*, 111895. 10.1016/j.yexcr.2020.111895.
57. Clarke, L., and Carbon, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature *287*, 504-509. 10.1038/287504a0.
58. Pluta, A.F., Mackay, A.M., Ainsztein, A.M., Goldberg, I.G., and Earnshaw, W.C. (1995). The Centromere: Hub of Chromosomal Activities. Science *270*, 1591-1594. doi:10.1126/science.270.5242.1591.
59. Kanesaki, Y., Imamura, S., Matsuzaki, M., and Tanaka, K. (2015). Identification of centromere regions in chromosomes of a unicellular red alga, Cyanidioschyzon merolae. FEBS Lett *589*, 1219-1224. 10.1016/j.febslet.2015.04.009.
60. Schneider, K.L., Xie, Z., Wolfgruber, T.K., and Presting, G.G. (2016). Inbreeding drives maize centromere evolution. Proceedings of the National Academy of Sciences *113*, E987-E996. doi:10.1073/pnas.1522008113.
61. Sullivan, K.F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol *127*, 581-592. 10.1083/jcb.127.3.581.

62. 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? Am J Hum Genet *52*, 1153-1163.
63. Barnhart, M.C., Kuich, P.H., 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. J Cell Biol *194*, 229-243. 10.1083/jcb.201012017.
64. Foltz, D.R., Jansen, L.E., Bailey, A.O., Yates, J.R., 3rd, 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. 10.1016/j.cell.2009.02.039.
65. Wu, J.C., and Manuelidis, L. (1980). Sequence definition and organization of a human repeated DNA. J Mol Biol *142*, 363-386. 10.1016/0022-2836(80)90277-6.
66. Willard, H.F., and Waye, J.S. (1987). Hierarchical order in chromosome-specific human alpha satellite DNA. Trends in Genetics *3*, 192-198. https://doi.org/10.1016/0168-9525(87)90232-0.
67. Willard, H.F. (1989). The genomics of long tandem arrays of satellite DNA in the human genome. Genome *31*, 737-744. 10.1139/g89-132.
68. Eichler, E.E., Clark, R.A., and She, X. (2004). An assessment of the sequence gaps: unfinished business in a finished human genome. Nat Rev Genet *5*, 345-354. 10.1038/nrg1322.
69. Miga, K.H. (2015). Completing the human genome: the progress and challenge of satellite DNA assembly. Chromosome Res *23*, 421-426. 10.1007/s10577-015-9488-2.
70. Gershman, A., Sauria, M.E.G., Guitart, X., Vollger, M.R., Hook, P.W., Hoyt, S.J., Jain, M., Shumate, A., Razaghi, R., Koren, S., et al. (2022). Epigenetic patterns in a complete human genome. Science *376*, eabj5089. 10.1126/science.abj5089.
71. Trowell, H.E., Nagy, A., Vissel, B., and Choo, K.H. (1993). Long-range analyses of the centromeric regions of human chromosomes 13, 14 and 21: identification of a narrow domain containing two key centromeric DNA elements. Hum Mol Genet *2*, 1639-1649. 10.1093/hmg/2.10.1639.
72. Jackson, M.S., Slijepcevic, P., and Ponder, B.A. (1993). The organisation of repetitive sequences in the pericentromeric region of human chromosome 10. Nucleic Acids Res *21*, 5865-5874. 10.1093/nar/21.25.5865.
73. Shepelev, V.A., Alexandrov, A.A., Yurov, Y.B., and Alexandrov, I.A. (2009). The evolutionary origin of man can be traced in the layers of defunct ancestral alpha satellites flanking the active centromeres of human chromosomes. PLoS Genet *5*, e1000641. 10.1371/journal.pgen.1000641.
74. Thakur, J., Packiaraj, J., and Henikoff, S. (2021). Sequence, Chromatin and Evolution of Satellite DNA. International Journal of Molecular Sciences *22*, 4309.
75. Logsdon, G.A., and Eichler, E.E. (2023). The Dynamic Structure and Rapid Evolution of Human Centromeric Satellite DNA. Genes *14*, 92.
76. Miga, K.H., and Alexandrov, I.A. (2021). Variation and Evolution of Human Centromeres: A Field Guide and Perspective. Annu Rev Genet *55*, 583-602. 10.1146/annurev-genet-071719-020519.

77. Alexandrov, I.A., Medvedev, L.I., Mashkova, T.D., Kisselev, L.L., Romanova, L.Y., and Yurov, Y.B. (1993). Definition of a new alpha satellite suprachromosomal family characterized by monomeric organization. Nucleic Acids Res *21*, 2209-2215. 10.1093/nar/21.9.2209.
78. 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. J Cell Biol *109*, 1963-1973. 10.1083/jcb.109.5.1963.
79. Ohzeki, J.-i., Otake, K., and Masumoto, H. (2020). Human artificial chromosome: Chromatin assembly mechanisms and CENP-B. Experimental Cell Research *389*, 111900. https://doi.org/10.1016/j.yexcr.2020.111900.
80. 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. J Cell Biol *141*, 309-319. 10.1083/jcb.141.2.309.
81. Fachinetti, D., Han, J.S., McMahon, M.A., Ly, P., Abdullah, A., Wong, A.J., and Cleveland, D.W. (2015). DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. Dev Cell *33*, 314-327. 10.1016/j.devcel.2015.03.020.
82. Okada, T., Ohzeki, J., Nakano, M., Yoda, K., Brinkley, W.R., Larionov, V., and Masumoto, H. (2007). CENP-B controls centromere formation depending on the chromatin context. Cell *131*, 1287-1300. 10.1016/j.cell.2007.10.045.
83. Rice, W. (2020). Why Do Centromeres Evolve So Fast: BIR Replication, Hypermutation, Transposition, and Molecular-Drive. Preprints. 10.20944/preprints202012.0669.v1.
84. Saayman, X., Graham, E., Nathan, W.J., Nussenzweig, A., and Esashi, F. (2023). Centromeres as universal hotspots of DNA breakage, driving RAD51-mediated recombination during quiescence. Mol Cell *83*, 523-538 e527. 10.1016/j.molcel.2023.01.004.
85. Carrasco-Salas, Y., Malapert, A., Sulthana, S., Molcrette, B., Chazot-Franguiadakis, L., Bernard, P., Chédin, F., Faivre-Moskalenko, C., and Vanoosthuyse, V. (2019). The extruded non-template strand determines the architecture of R-loops. Nucleic Acids Research *47*, 6783-6795. 10.1093/nar/gkz341.
86. Kasinathan, S., and Henikoff, S. (2018). Non-B-Form DNA Is Enriched at Centromeres. Mol Biol Evol *35*, 949-962. 10.1093/molbev/msy010.
87. Liu, Q., Yi, C., Zhang, Z., Su, H., Liu, C., Huang, Y., Li, W., Hu, X., Liu, C., Birchler, J.A., et al. (2023). Non-B-form DNA tends to form in centromeric regions and has undergone changes in polyploid oat subgenomes. Proc Natl Acad Sci U S A *120*, e2211683120. 10.1073/pnas.2211683120.
88. Bakhoum, S.F., Kabeche, L., Murnane, J.P., Zaki, B.I., and Compton, D.A. (2014). DNA-damage response during mitosis induces whole-chromosome missegregation. Cancer Discov *4*, 1281-1289. 10.1158/2159-8290.Cd-14-0403.
89. 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. 10.1038/ncb3344.

90. Trier, I., Black, E.M., Joo, Y.K., and Kabeche, L. (2023). ATR protects centromere identity by promoting DAXX association with PML nuclear bodies. Cell Rep *42*, 112495. 10.1016/j.celrep.2023.112495.
91. 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. Proc Natl Acad Sci U S A *118*. 10.1073/pnas.2015634118.
92. Talbert, P.B., and Henikoff, S. (2018). Transcribing Centromeres: Noncoding RNAs and Kinetochore Assembly. Trends Genet *34*, 587-599. 10.1016/j.tig.2018.05.001.
93. 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. The EMBO Journal *30*, 328-340. https://doi.org/10.1038/emboj.2010.329.
94. 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. Dev Cell *42*, 226-240 e226. 10.1016/j.devcel.2017.07.001.
95. Tanaka, Y., Tachiwana, H., Yoda, K., Masumoto, H., Okazaki, T., Kurumizaka, H., and Yokoyama, S. (2005). Human centromere protein B induces translational positioning of nucleosomes on alpha-satellite sequences. J Biol Chem *280*, 41609-41618. 10.1074/jbc.M509666200.
96. Keszthelyi, A., Minchell, N.E., and Baxter, J. (2016). The Causes and Consequences of Topological Stress during DNA Replication. Genes (Basel) *7*. 10.3390/genes7120134.
97. Baxter, J., and Diffley, J.F. (2008). Topoisomerase II inactivation prevents the completion of DNA replication in budding yeast. Mol Cell *30*, 790-802. 10.1016/j.molcel.2008.04.019.
98. Ahmad, S.F., Singchat, W., Jehangir, M., Suntronpong, A., Panthum, T., Malaivijitnond, S., and Srikulnath, K. (2020). Dark Matter of Primate Genomes: Satellite DNA Repeats and Their Evolutionary Dynamics. Cells *9*. 10.3390/cells9122714.
99. Chatterjee, N., and Walker, G.C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen *58*, 235-263. 10.1002/em.22087.
100. Marnef, A., and Legube, G. (2021). R-loops as Janus-faced modulators of DNA repair. Nature Cell Biology *23*, 305-313. 10.1038/s41556-021-00663-4.
101. Zhao, F., Kim, W., Kloeber, J.A., and Lou, Z. (2020). DNA end resection and its role in DNA replication and DSB repair choice in mammalian cells. Experimental & Molecular Medicine *52*, 1705-1714. 10.1038/s12276-020-00519-1.
102. Shrivastav, M., De Haro, L.P., and Nickoloff, J.A. (2008). Regulation of DNA double-strand break repair pathway choice. Cell Research *18*, 134-147. 10.1038/cr.2007.111.
103. Howard, S.M., Yanez, D.A., and Stark, J.M. (2015). DNA damage response factors from diverse pathways, including DNA crosslink repair, mediate alternative end joining. PLoS Genet *11*, e1004943. 10.1371/journal.pgen.1004943.

104. Pinto, L., Baidarjad, H., Entz-Werlé, N., and Van Dyck, E. (2021). Impact of Chromatin Dynamics and DNA Repair on Genomic Stability and Treatment Resistance in Pediatric High-Grade Gliomas. Cancers *13*, 5678.
105. Li, X., and Heyer, W.-D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. Cell Research *18*, 99-113. 10.1038/cr.2008.1.
106. Rief, N., and Löbrich, M. (2002). Efficient rejoining of radiation-induced DNA double-strand breaks in centromeric DNA of human cells. J Biol Chem *277*, 20572-20582. 10.1074/jbc.M200265200.
107. 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. 10.1038/s41586-021-04200-z.
108. Chardon, F., Japaridze, A., Witt, H., Velikovsky, 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 *82*, 1751-1767.e1758. https://doi.org/10.1016/j.molcel.2022.02.032.
109. Heyer, W.D., Ehmsen, K.T., and Liu, J. (2010). Regulation of homologous recombination in eukaryotes. Annu Rev Genet *44*, 113-139. 10.1146/annurev-genet-051710-150955.
110. Scelfo, A., and Fachinetti, D. (2023). Centromere: A Trojan horse for genome stability. DNA Repair (Amst) *130*, 103569. 10.1016/j.dnarep.2023.103569.
111. 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. The EMBO Journal *37*, e96729. https://doi.org/10.15252/embj.201796729.
112. Cannan, W.J., and Pederson, D.S. (2016). Mechanisms and Consequences of Double-Strand DNA Break Formation in Chromatin. J Cell Physiol *231*, 3-14. 10.1002/jcp.25048.
113. Costantino, L., Sotiriou, S.K., Rantala, J.K., Magin, S., Mladenov, E., Helleday, T., Haber, J.E., Iliakis, G., Kallioniemi, O.P., and Halazonetis, T.D. (2014). Break-induced replication repair of damaged forks induces genomic duplications in human cells. Science *343*, 88-91. 10.1126/science.1243211.
114. Bhowmick, R., Hickson, I.D., and Liu, Y. (2023). Completing genome replication outside of S phase. Mol Cell *83*, 3596-3607. 10.1016/j.molcel.2023.08.023.
115. Zlotorynski, E., Rahat, A., Skaug, J., Ben-Porat, N., Ozeri, E., Hershberg, R., Levi, A., Scherer, S.W., Margalit, H., and Kerem, B. (2003). Molecular basis for expression of common and rare fragile sites. Mol Cell Biol *23*, 7143-7151. 10.1128/mcb.23.20.7143-7151.2003.
116. Ozeri-Galai, E., Lebofsky, R., Rahat, A., Bester, A.C., Bensimon, A., and Kerem, B. (2011). Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. Mol Cell *43*, 122-131. 10.1016/j.molcel.2011.05.019.
117. Glover, T.W., Berger, C., Coyle, J., and Echo, B. (1984). DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. Hum Genet *67*, 136-142. 10.1007/bf00272988.

118. Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Mol Cell *44*, 966-977. 10.1016/j.molcel.2011.10.013.
119. Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W., Earle, E., Choo, K.H., and Wong, L.H. (2012). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc Natl Acad Sci U S A *109*, 1979-1984. 10.1073/pnas.1108705109.
120. Blackford, A.N., and Stucki, M. (2020). How Cells Respond to DNA Breaks in Mitosis. Trends Biochem Sci *45*, 321-331. 10.1016/j.tibs.2019.12.010.
121. Gunderson, L.L., and Tepper, J.E. (2015). Clinical radiation oncology (Elsevier Health Sciences).
122. Min, J., Wright, W.E., and Shay, J.W. (2017). Alternative Lengthening of Telomeres Mediated by Mitotic DNA Synthesis Engages Break-Induced Replication Processes. Mol Cell Biol *37*. 10.1128/MCB.00226-17.
123. Bhowmick, R., Minocherhomji, S., and Hickson, I.D. (2016). RAD52 Facilitates Mitotic DNA Synthesis Following Replication Stress. Mol Cell *64*, 1117-1126. 10.1016/j.molcel.2016.10.037.
124. Cesare, A.J., and Reddel, R.R. (2010). Alternative lengthening of telomeres: models, mechanisms and implications. Nature Reviews Genetics *11*, 319-330. 10.1038/nrg2763.
125. Showman, S., Talbert, P.B., Xu, Y., Adeyemi, R.O., and Henikoff, S. (2024). Expansion of human centromeric arrays in cells undergoing break-induced replication. Cell Reports *43*. https://doi.org/10.1016/j.celrep.2024.113851.
126. Lefkovits, I., and Waldmann, H. (1984). Limiting dilution analysis of the cells of immune system I. The clonal basis of the immune response. Immunol Today *5*, 265-268. https://doi.org/10.1016/0167-5699(84)90137-3.
127. Waldmann, H., and Lefkovits, I. (1984). Limiting dilution analysis of cells of the immune system II: What can be learnt? Immunol Today *5*, 295-298. https://doi.org/10.1016/0167-5699(84)90154-3.
128. McKinley, K.L., and Cheeseman, I.M. (2016). The molecular basis for centromere identity and function. Nature Reviews Molecular Cell Biology *17*, 16-29. 10.1038/nrm.2015.5.
129. Kixmoeller, K., Allu, P.K., and Black, B.E. (2020). The centromere comes into focus: from CENP-A nucleosomes to kinetochore connections with the spindle. Open Biol *10*, 200051. 10.1098/rsob.200051.
130. Allshire, R.C., and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nature Reviews Genetics *9*, 923-937. 10.1038/nrg2466.
131. 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. 10.1186/gb-2013-14-1-r10.
132. Birchler, J.A., Gao, Z., and Han, F. (2009). A tale of two centromeres--diversity of structure but conservation of function in plants and animals. Funct Integr Genomics *9*, 7-13. 10.1007/s10142-008-0104-9.

133. Willard, H.F. (1990). Centromeres of mammalian chromosomes. Trends in Genetics *6*, 410-416. https://doi.org/10.1016/0168-9525(90)90302-M.
134. Henikoff, S., Ahmad, K., and Malik, H.S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. Science *293*, 1098-1102. 10.1126/science.1062939.
135. Logsdon, G.A., Rozanski, A.N., Ryabov, F., Potapova, T., Shepelev, V.A., Mao, Y., Rautiainen, M., Koren, S., Nurk, S., Porubsky, D., et al. (2023). The variation and evolution of complete human centromeres. bioRxiv, 2023.2005.2030.542849. 10.1101/2023.05.30.542849.
136. Logsdon, G.A., Vollger, M.R., Hsieh, P., Mao, Y., Liskovykh, 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. 10.1038/s41586-021-03420-7.
137. de Lima, L.G., Howe, E., Singh, V.P., Potapova, T., Li, H., Xu, B., Castle, J., Crozier, S., Harrison, C.J., Clifford, S.C., et al. (2021). PCR amplicons identify widespread copy number variation in human centromeric arrays and instability in cancer. Cell Genom *1*, 100064. 10.1016/j.xgen.2021.100064.
138. Rice, W.R. (2019). A Game of Thrones at Human Centromeres I. Multifarious structure necessitates a new molecular/evolutionary model. bioRxiv, 731430. 10.1101/731430.
139. Rice, W.R. (2019). A Game of Thrones at Human Centromeres II. A new molecular/evolutionary model. bioRxiv, 731471. 10.1101/731471.
140. Sotiriou, S.K., Kamileri, I., Lugli, N., Evangelou, K., Da-Re, C., Huber, F., Padayachy, L., Tardy, S., Nicati, N.L., Barriot, S., et al. (2016). Mammalian RAD52 Functions in Break-Induced Replication Repair of Collapsed DNA Replication Forks. Mol Cell *64*, 1127-1134. 10.1016/j.molcel.2016.10.038.
141. Zhang, J.M., Yadav, T., Ouyang, J., Lan, L., and Zou, L. (2019). Alternative Lengthening of Telomeres through Two Distinct Break-Induced Replication Pathways. Cell Rep *26*, 955-968.e953. 10.1016/j.celrep.2018.12.102.
142. Saayman, X., Graham, E., Nathan, W.J., Nussenzweig, A., and Esashi, F. (2023). Centromeres as universal hotspots of DNA breakage, driving RAD51-mediated recombination during quiescence. Mol Cell *83*, 523-538.e527. 10.1016/j.molcel.2023.01.004.
143. Greenfeder, S.A., and Newlon, C.S. (1992). Replication forks pause at yeast centromeres. Mol Cell Biol *12*, 4056-4066. 10.1128/mcb.12.9.4056-4066.1992.
144. Mitra, S., Gómez-Raja, J., Larriba, G., Dubey, D.D., and Sanyal, K. (2014). Rad51-Rad52 mediated maintenance of centromeric chromatin in Candida albicans. PLoS Genet *10*, e1004344. 10.1371/journal.pgen.1004344.
145. 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. Sci Rep *5*, 13307. 10.1038/srep13307.
146. 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. 10.1126/science.aan6490.

147. 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. 10.1002/chem.201500448.
148. Rice, W. (2020). Why Do Centromeres Evolve So Fast: BIR Replication, Hypermutation, Transposition, and Molecular-Drive. Preprints *2020120669*, 25 December 2020 https://doi.org/2010.20944/preprints202012.200669.v202011. https://doi.org/10.20944/preprints202012.0669.v1.
149. Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., Chabes, A., and Malkova, A. (2011). Break-induced replication is highly inaccurate. PLoS Biol *9*, e1000594. 10.1371/journal.pbio.1000594.
150. Nogueira, A., Fernandes, M., Catarino, R., and Medeiros, R. (2019). RAD52 Functions in Homologous Recombination and Its Importance on Genomic Integrity Maintenance and Cancer Therapy. Cancers (Basel) *11*. 10.3390/cancers11111622.
151. Li, S., Wang, H., Jehi, S., Li, J., Liu, S., Wang, Z., Truong, L., Chiba, T., Wang, Z., and Wu, X. (2021). PIF1 helicase promotes break-induced replication in mammalian cells. Embo j *40*, e104509. 10.15252/embj.2020104509.
152. Saha, A.K., Mourad, M., Kaplan, M.H., Chefetz, I., Malek, S.N., Buckanovich, R., Markovitz, D.M., and Contreras-Galindo, R. (2019). The Genomic Landscape of Centromeres in Cancers. Sci Rep *9*, 11259. 10.1038/s41598-019-47757-6.
153. Li, J., Cai, Z., Vaites, L.P., Shen, N., Mitchell, D.C., Huttlin, E.L., Paulo, J.A., Harry, B.L., and Gygi, S.P. (2021). Proteome-wide mapping of short-lived proteins in human cells. Molecular Cell *81*, 4722-4735.e4725. https://doi.org/10.1016/j.molcel.2021.09.015.
154. Kernohan, K.D., Jiang, Y., Tremblay, D.C., Bonvissuto, A.C., Eubanks, J.H., Mann, M.R., and Bérubé, N.G. (2010). ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. Dev Cell *18*, 191-202. 10.1016/j.devcel.2009.12.017.
155. Lovejoy, C.A., Takai, K., Huh, M.S., Picketts, D.J., and de Lange, T. (2020). ATRX affects the repair of telomeric DSBs by promoting cohesion and a DAXX-dependent activity. PLoS Biol *18*, e3000594. 10.1371/journal.pbio.3000594.
156. Ritchie, K., Seah, C., Moulin, J., Isaac, C., Dick, F., and Bérubé, N.G. (2008). Loss of ATRX leads to chromosome cohesion and congression defects. J Cell Biol *180*, 315-324. 10.1083/jcb.200706083.
157. Kobayashi, T., and Ganley, A.R. (2005). Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. Science *309*, 1581-1584. 10.1126/science.1116102.
158. Minocherhomji, S., Ying, S., Bjerregaard, V.A., Bursomanno, S., Aleliunaite, A., Wu, W., Mankouri, H.W., Shen, H., Liu, Y., and Hickson, I.D. (2015). Replication stress activates DNA repair synthesis in mitosis. Nature *528*, 286-290. 10.1038/nature16139.
159. Wilson, M.A., Kwon, Y., Xu, Y., Chung, W.-H., Chi, P., Niu, H., Mayle, R., Chen, X., Malkova, A., Sung, P., and Ira, G. (2013). Pif1 helicase and Polδ promote recombination-coupled DNA synthesis via bubble migration. Nature *502*, 393-396. 10.1038/nature12585.

160. Mahlke, M.A., and Nechemia-Arbely, Y. (2020). Guarding the Genome: CENP-A-Chromatin in Health and Cancer. Genes (Basel) *11*. 10.3390/genes11070810.
161. Heaphy, C.M., de Wilde, R.F., Jiao, Y., Klein, A.P., Edil, B.H., Shi, C., Bettegowda, C., Rodriguez, F.J., Eberhart, C.G., Hebbar, S., et al. (2011). Altered Telomeres in Tumors with ATRX and DAXX Mutations. Science *333*, 425-425. doi:10.1126/science.1207313.
162. Lovejoy, C.A., Li, W., Reisenweber, S., Thongthip, S., Bruno, J., de Lange, T., De, S., Petrini, J.H., Sung, P.A., Jasin, M., et al. (2012). Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. PLoS Genet *8*, e1002772. 10.1371/journal.pgen.1002772.
163. Schwartzentruber, J., Korshunov, A., Liu, X.-Y., Jones, D.T.W., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.-A.K., Tönjes, M., et al. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226-231. 10.1038/nature10833.
164. Wong, L.H., McGhie, J.D., Sim, M., Anderson, M.A., Ahn, S., Hannan, R.D., George, A.J., Morgan, K.A., Mann, J.R., and Choo, K.H. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res *20*, 351-360. 10.1101/gr.101477.109.
165. Nguyen, D.T., Voon, H.P.J., Xella, B., Scott, C., Clynes, D., Babbs, C., Ayyub, H., Kerry, J., Sharpe, J.A., Sloane-Stanley, J.A., et al. (2017). The chromatin remodelling factor ATRX suppresses R-loops in transcribed telomeric repeats. EMBO Rep *18*, 914-928. 10.15252/embr.201643078.
166. Wang, Y., Yang, J., Wild, A.T., Wu, W.H., Shah, R., Danussi, C., Riggins, G.J., Kannan, K., Sulman, E.P., Chan, T.A., and Huse, J.T. (2019). G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. Nat Commun *10*, 943. 10.1038/s41467-019-08905-8.
167. Huh, M.S., Ivanochko, D., Hashem, L.E., Curtin, M., Delorme, M., Goodall, E., Yan, K., and Picketts, D.J. (2016). Stalled replication forks within heterochromatin require ATRX for protection. Cell Death Dis *7*, e2220. 10.1038/cddis.2016.121.
168. Leung, J.W., Ghosal, G., Wang, W., Shen, X., Wang, J., Li, L., and Chen, J. (2013). Alpha thalassemia/mental retardation syndrome X-linked gene product ATRX is required for proper replication restart and cellular resistance to replication stress. J Biol Chem *288*, 6342-6350. 10.1074/jbc.M112.411603.
169. Juhász, S., Elbakry, A., Mathes, A., and Löbrich, M. (2018). ATRX Promotes DNA Repair Synthesis and Sister Chromatid Exchange during Homologous Recombination. Mol Cell *71*, 11-24.e17. 10.1016/j.molcel.2018.05.014.
170. Raghunandan, M., Yeo, J.E., Walter, R., Saito, K., Harvey, A.J., Ittershagen, S., Lee, E.A., Yang, J., Hoatlin, M.E., Bielinsky, A.K., et al. (2020). Functional cross talk between the Fanconi anemia and ATRX/DAXX histone chaperone pathways promotes replication fork recovery. Hum Mol Genet *29*, 1083-1095. 10.1093/hmg/ddz250.
171. Clynes, D., Jelinska, C., Xella, B., Ayyub, H., Scott, C., Mitson, M., Taylor, S., Higgs, D.R., and Gibbons, R.J. (2015). Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. Nat Commun *6*, 7538. 10.1038/ncomms8538.

172. Muoio, D., Laspata, N., Dannenberg, R.L., Curry, C., Darkoa-Larbi, S., Hedglin, M., Uttam, S., and Fouquerel, E. (2024). PARP2 promotes Break Induced Replication-mediated telomere fragility in response to replication stress. Nature Communications *15*, 2857. 10.1038/s41467-024-47222-7.
173. Mukherjee, J., Johannessen, T.C., Ohba, S., Chow, T.T., Jones, L., Pandita, A., and Pieper, R.O. (2018). Mutant IDH1 Cooperates with ATRX Loss to Drive the Alternative Lengthening of Telomere Phenotype in Glioma. Cancer Res *78*, 2966-2977. 10.1158/0008-5472.Can-17-2269.
174. Demin, A.A., Hirota, K., Tsuda, M., Adamowicz, M., Hailstone, R., Brazina, J., Gittens, W., Kalasova, I., Shao, Z., Zha, S., et al. (2021). XRCC1 prevents toxic PARP1 trapping during DNA base excision repair. Mol Cell *81*, 3018-3030.e3015. 10.1016/j.molcel.2021.05.009.
175. Yang, Z., Takai, K.K., Lovejoy, C.A., and de Lange, T. (2020). Break-induced replication promotes fragile telomere formation. Genes Dev *34*, 1392-1405. 10.1101/gad.328575.119.
176. Doksani, Y., and de Lange, T. (2016). Telomere-Internal Double-Strand Breaks Are Repaired by Homologous Recombination and PARP1/Lig3-Dependent End-Joining. Cell Rep *17*, 1646-1656. 10.1016/j.celrep.2016.10.008.
177. Pan, X., Drosopoulos, W.C., Sethi, L., Madireddy, A., Schildkraut, C.L., and Zhang, D. (2017). FANCM, BRCA1, and BLM cooperatively resolve the replication stress at the ALT telomeres. Proceedings of the National Academy of Sciences *114*, E5940-E5949. doi:10.1073/pnas.1708065114.
178. Lüscher, B., Ahel, I., Altmeyer, M., Ashworth, A., Bai, P., Chang, P., Cohen, M., Corda, D., Dantzer, F., Daugherty, M.D., et al. (2022). ADP-ribosyltransferases, an update on function and nomenclature. Febs j *289*, 7399-7410. 10.1111/febs.16142.
179. Mocanu, C., Karanika, E., Fernandez-Casanas, M., Herbert, A., Olukoga, T., Ozgurses, M.E., and Chan, K.L. (2022). DNA replication is highly resilient and persistent under the challenge of mild replication stress. Cell Rep *39*, 110701. 10.1016/j.celrep.2022.110701.
180. Racca, C., Britton, S., Hédouin, S., Francastel, C., Calsou, P., and Larminat, F. (2021). BRCA1 prevents R-loop-associated centromeric instability. Cell Death & Disease *12*, 896. 10.1038/s41419-021-04189-3.
181. Rose, A.M., Goncalves, T., Cunniffe, S., Geiller, H.E.B., Kent, T., Shepherd, S., Ratnaweera, M., O'Sullivan, R.J., Gibbons, R.J., and Clynes, D. (2023). Induction of the alternative lengthening of telomeres pathway by trapping of proteins on DNA. Nucleic Acids Res *51*, 6509-6527. 10.1093/nar/gkad150.
182. 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. https://doi.org/10.1016/j.celrep.2019.08.041.
183. Sen Gupta, A., Seidel, C., Tsuchiya, D., McKinney, S., Yu, Z., Smith, S.E., Unruh, J.R., and Gerton, J.L. (2023). Defining a core configuration for human centromeres during mitosis. Nature Communications *14*, 7947. 10.1038/s41467-023-42980-2.

184. Yang, J.W., Pendon, C., Yang, J., Haywood, N., Chand, A., and Brown, W.R.A. (2000). Human mini-chromosomes with minimal centromeres. Human Molecular Genetics *9*, 1891-1902. 10.1093/hmg/9.12.1891.
185. Lo, A.W., Liao, G.C., Rocchi, M., and Choo, K.H. (1999). Extreme reduction of chromosome-specific alpha-satellite array is unusually common in human chromosome 21. Genome Res *9*, 895-908. 10.1101/gr.9.10.895.
186. Li, J.S., Miralles Fusté, J., Simavorian, T., Bartocci, C., Tsai, J., Karlseder, J., and Lazzerini Denchi, E. (2017). TZAP: A telomere-associated protein involved in telomere length control. Science *355*, 638-641. 10.1126/science.aah6752.
187. Rivera, T., Haggblom, C., Cosconati, S., and Karlseder, J. (2017). A balance between elongation and trimming regulates telomere stability in stem cells. Nat Struct Mol Biol *24*, 30-39. 10.1038/nsmb.3335.
188. Iida, T., and Kobayashi, T. (2019). RNA Polymerase I Activators Count and Adjust Ribosomal RNA Gene Copy Number. Molecular Cell *73*, 645-654.e613. https://doi.org/10.1016/j.molcel.2018.11.029.
189. Mikhailov, A., Cole, R.W., and Rieder, C.L. (2002). DNA damage during mitosis in human cells delays the metaphase/anaphase transition via the spindle-assembly checkpoint. Curr Biol *12*, 1797-1806. 10.1016/s0960-9822(02)01226-5.
190. Wassing, I.E., Graham, E., Saayman, X., Rampazzo, L., Ralf, C., Bassett, A., and Esashi, F. (2021). The RAD51 recombinase protects mitotic chromatin in human cells. Nature Communications *12*, 5380. 10.1038/s41467-021-25643-y.
191. Özer, Ö., Bhowmick, R., Liu, Y., and Hickson, I.D. (2018). Human cancer cells utilize mitotic DNA synthesis to resist replication stress at telomeres regardless of their telomere maintenance mechanism. Oncotarget *9*, 15836-15846. 10.18632/oncotarget.24745.
192. Ruff, P., Donnianni, R.A., Glancy, E., Oh, J., and Symington, L.S. (2016). RPA Stabilization of Single-Stranded DNA Is Critical for Break-Induced Replication. Cell Rep *17*, 3359-3368. 10.1016/j.celrep.2016.12.003.
193. Sacristan, C., Samejima, K., Ruiz, L.A., Deb, M., Lambers, M.L.A., Buckle, A., Brackley, C.A., Robertson, D., Hori, T., Webb, S., et al. (2024). Vertebrate centromeres in mitosis are functionally bipartite structures stabilized by cohesin. Cell *187*, 3006-3023 e3026. 10.1016/j.cell.2024.04.014.