

STATE-OF-THE-ART REVIEW

Breaking the paradigm: early insights from mammalian DNA *breakomes*

 Xanita Saayman and Fumiko Esashi 

Sir William Dunn School of Pathology, University of Oxford, UK

Keywords

DNA breaks; genome instability; next-generation sequencing; replication; transcription

Correspondence

 F. Esashi, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK
 Tel: +44 1865 275289
 E-mail: fumiko.esashi@path.ox.ac.uk

(Received 30 October 2020, revised 4 March 2021, accepted 29 March 2021)

doi:10.1111/febs.15849

DNA double-strand breaks (DSBs) can result from both exogenous and endogenous sources and are potentially toxic lesions to the human genome. If improperly repaired, DSBs can threaten genome integrity and contribute to premature ageing, neurodegenerative disorders and carcinogenesis. Through decades of work on genome stability, it has become evident that certain regions of the genome are inherently more prone to breakage than others, known as genome instability hotspots. Recent advancements in sequencing-based technologies now enable the profiling of genome-wide distributions of DSBs, also known as *breakomes*, to systematically map these instability hotspots. Here, we review the application of these technologies and their implications for our current understanding of the genomic regions most likely to drive genome instability. These breakomes ultimately highlight both new and established breakage hotspots including actively transcribed regions, loop boundaries and early-replicating regions of the genome. Further, these breakomes challenge the paradigm that DNA breakage primarily occurs in hard-to-replicate regions. With these advancements, we begin to gain insights into the biological mechanisms both invoking and protecting against genome instability.

Introduction

The accurate preservation and faithful transmission of genetic information is arguably the single most important task of the living organism. This is no small feat, given the considerable range and extent of DNA-damaging threats experienced by any given cell. It has been estimated that human cells undergo ~ 70 000 lesions per day, of which 25–50 of these are double-strand breaks (DSBs) [1,2]. If improperly repaired, DSBs can result in both base mutations and chromosomal rearrangements, such as inversions, amplifications, deletions and translocations. Ultimately, these events can

underlie a range of pathologies related to genome instability including carcinogenesis and premature ageing [3–5].

Double-strand breaks in the human genome can result from extrinsic sources such as ionizing or ultraviolet radiation as well as intrinsic sources such as replication, transcription and chromatin looping [6–10]. While DSBs from extrinsic sources tend to be more randomly distributed across the genome, DSBs from intrinsic sources tend to reoccur at specific genome locations. This nonrandom distribution indicates

Abbreviations

CFS, Common fragile site; CTCF, CCCTC-binding factor; DSB, Double-strand break; ERFS, Early-replicating fragile site; FRA16D, Fragile site, aphidicolin type, common, fra(16)(q23.2); FRA3B, Fragile site, aphidicolin type, common, fra(3)(p14.2); MCF7, Michigan Cancer Foundation-7 (breast cancer cell line); NHEK, Normal human epidermal keratinocytes; RAD51, DNA repair protein RAD51 homolog 1; rDNA, Ribosomal DNA; RPA, Replication protein A; SSB, Single-strand break; TAD, Topologically associated domain; TOP I, Topoisomerase I; TOP II, Topoisomerase II.

that there are regions of the genome that are unusually susceptible to DSB formation, also known as genome instability hotspots.

Historically, the mapping of genome instability hotspots has been conducted upon exposure to exogenous genotoxic stress (e.g. drug-induced replication perturbation, topoisomerase inhibitors) or genetic perturbation of DNA metabolism (e.g. mutation of the replication machinery), thereby increasing the incidence of DNA breakage and enabling the detection of DSBs. From these works, a set of genome instability hotspots have been defined, including regions of low origin density, DNA repeats, common or rare fragile sites, slow replicating zones, non-B DNA structures and telomeres [11–17], reviewed in Ref. [18]. However, given that these instability hotspots were typically characterized following exogenous stress or in mutant genotypes, it remains unclear whether these regions accurately represent genome instability hotspots in unperturbed cell populations.

With recent developments in sequencing-based technologies, we are now beginning to gain the sensitivity to quantitatively and directly detect physical DNA breaks in populations of cells. Mapping genome-wide break profiles, also known as DNA *breakomes*, identifies inherent DNA break susceptibility across the genome. Here, we review the findings following the application of these recently developed technologies profiling DNA breakomes in various experimental conditions (Table 1), and what is known of the mechanisms underlying DNA break susceptibility at the identified instability hotspots. This review focuses on the biological implications of breakome studies rather than a systematic comparison of the different technologies available, which has been reviewed extensively elsewhere [19]. Overlooking these emerging insights, we also aim to discuss how DNA breakomes challenge current paradigms of genome instability hotspots.

Genome instability hotspots

DNA breakomes conducted in the last decade have identified genome instability hotspots such as actively transcribed regions, loop boundaries, fragile sites, early-replicating regions, telomeres and centromeres (Fig. 1). Certain instability hotspots (e.g. actively transcribed regions, loop boundaries and early-replicating regions) appear to form DSBs innately, whereas others (e.g. common fragile sites (CFSs), telomeres and centromeres) become susceptible to breakage only in certain physiological settings. Here, we discuss each of these genome instability hotspots, the conditions in which they express their instability and how this

corresponds to existing paradigms of genome instability throughout the genome.

Actively transcribed regions: transcription as a source of genome instability

In recent years, it has become increasingly apparent that breakome landscapes are considerably skewed towards regions that are actively transcribed [10,20–29]. Indeed, sites of active transcription are one of the only regions to be detected in breakomes using noncancerous mammalian cells without any exogenous perturbation [25,27,30]. Further, a direct correlation between the rates of active transcription and DSB formation has also been consistently reported [22,24,26,27,31,32]. This trend remains apparent across a range of experimental conditions including replication stress and topoisomerase II (TOP II) poisoning, suggesting that transcription poses a threat to genome stability in various contexts [10,20,21,23–26,28,29]. These findings imply that genomic regions with high densities of transcriptionally active loci may become DNA instability hotspots. Alternatively, actively transcribed instability hotspots may not refer to consecutive sites in the one-dimensional genome, but rather multiple independent sites brought together to form three-dimensional compartments, such as in transcription hubs. In this sense, actively transcribed genome instability hotspots would refer to a position in nuclear space rather than a linear coordinate range within the human genome.

Increasing transcriptional activity acutely increases both mutation and recombination rates [9,33–36]. While this link between mutation rates, recombination and transcriptional activity has been recognized for several decades, the mechanistic details of transcription-related DNA fragility are only now beginning to be revealed. Several hypotheses have been proposed to explain the inherent fragility of actively transcribed regions, which can be generally classified into replication-dependent or replication-independent mechanisms.

Replication-dependent breaks

Transcription–replication collisions are rare in healthy human cells, as eukaryotes compartmentalize replication and transcription in distinct nuclear territories at distinct times [37,38]. However, collisions do occasionally occur at exceptionally long (> 800 kbp) genes, or when replication programmes are perturbed using replication inhibitors [39]. Indeed, genes that are both late-replicating and transcriptionally active have been identified as being susceptible to DNA instability, but

Table 1. Mapping genome instability hotspots genome-wide. A tabulation of recently published genome-wide DNA breakomes identifying genome instability hotspots. The method applied, biological conditions assessed and mammalian organism used are specified for each reference. On the far right, it is noted whether the respective study identified an enrichment of DNA breaks at common genome instability hotspots such as actively transcribed regions, insulating regions, early-replicating regions, fragile sites, centromeres or telomeres. BLESS, breaks labelling, enrichment of streptavidin and next-generation sequencing; BLISS, breaks labelling *in situ* and sequencing; DSBapture, double-strand DNA break capture; END-seq, end sequencing; RAFT-seq, rapid amplification of form termini sequencing; sBLISS, suspension-cell breaks labelling *in situ* and sequencing.

Reference	Method	Condition	Organism	Actively transcribed regions	Loop boundaries	Early-replicating regions	Fragile sites	(Peri)centromeres / telomeres
Crosetto <i>et al.</i> (2013) [23]	BLESS	Replication stress (APH)	<i>M. musculus</i> , <i>H. sapiens</i>	●			●	●
Baranello <i>et al.</i> (2014) [20]	SSB-seq / DSB-seq	+/- Top II poison (etoposide)	<i>H. sapiens</i> (HCT116)	●				
Yang <i>et al.</i> (2015) [24]	BLESS	+/- Anthracyclines, Top II poison (etoposide)	<i>M. musculus</i> (MSCC-CK1)	●				
Tchurikov <i>et al.</i> (2015)	RAFT-seq		<i>H. sapiens</i> (HEK293T)	●				
Lensing <i>et al.</i> (2016) [27]	DSBapture		<i>H. sapiens</i> (HeLa, U2OS, NHEK)	●	●			
Yan <i>et al.</i> (2017) [25]	BLISS	+/- Top II poison (etoposide)	<i>M. musculus</i> (primary liver cells), <i>H. sapiens</i> (U2OS, HEK293, KBM7)	●				
Canela <i>et al.</i> (2017)	END-seq	+/- Top II poison (etoposide)	<i>M. musculus</i> (activated B-cells), <i>H. sapiens</i> (activated B-cells)	●	●			
Mourad <i>et al.</i> (2018)	DSBapture, BLESS, BLISS, END-seq		<i>H. sapiens</i> (NHEK, U2OS, KBM7, MCF7)	●	●			
Tubbs <i>et al.</i> (2018)	END-seq	Low/high-dose replication stress (HU)	<i>M. musculus</i> (primary splenic B-cells), <i>H. sapiens</i> (HCT116)			●	●	
Canela <i>et al.</i> (2019) [28]	END-seq	Top II poison (etoposide)	MEFs, B-cells, HCT116	●	●			
Ballinger <i>et al.</i> (2019) [89]	DSBapture, BLISS and BLESS		<i>H. sapiens</i> (MCF7, NHEK, K562)	●	●	●		
Hazan <i>et al.</i> (2019) [30]	BLISS		<i>H. sapiens</i> MCF7, MCF10A, BJ, EndoC), <i>M. musculus</i> (embryonic, neural stem cells)	●	●			
Gothé <i>et al.</i> (2019) [26]	sBLISS	+/- Top II poison (etoposide)	<i>H. sapiens</i> (TK6, U2OS, K562)	●	●			
Promonet <i>et al.</i> (2020) [58]	i-BLESS	+/- Topoisomerase I (shRNA)	<i>H. sapiens</i> (HeLa)	●				
Chakraborty <i>et al.</i> (2020) [80]	Break-seq	+/- Replication stress (APH)	<i>H. sapiens</i> (lymphoblasts, wild-type and fragile X)	●				●

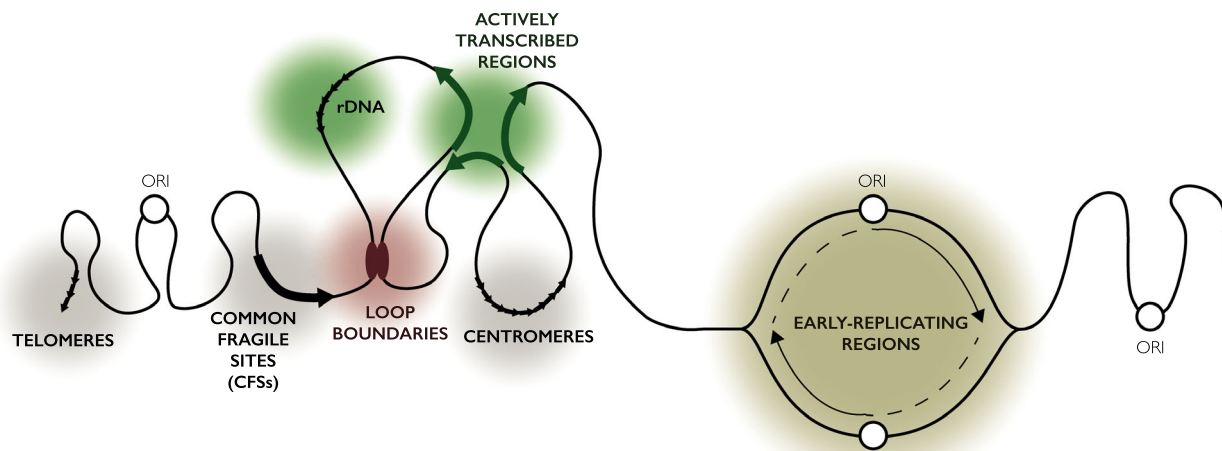


Fig. 1. Overview of genome instability hotspots. DNA breakomes profile genome-wide distributions of DNA DSBs to objectively identify genome instability hotspots in various experimental settings. In the absence of exogenous perturbation, genome instability hotspots most frequently occur at actively transcribed regions and rDNA (green), loop boundary sites between TADs (red), early-replicating regions (yellow). Upon exogenously induced replication stress, genome instability hotspots expand to include hard-to-replicate regions such as CFSs, centromeres and telomeres (grey).

only when the replication programme is perturbed [21,39–41].

Replication–transcription collisions can be induced either directly through physical obstructions of the respective machineries (Fig. 2A), or indirectly through other elements that impede replication fork progression (Fig. 2B). Direct head-on replication–transcription collisions are more prone to breakage than codirectional collisions, although both are able to interfere with replication progression in eukaryotes [42–45]. Indirect collisions can be mediated by non-canonical DNA structures formed during the normal process of replication or transcription, such as DNA supercoiling, non-B DNA structures and RNA–DNA hybrids [46–49]. In either case, stalled replication forks can be cleaved by structure-specific endonucleases to generate DSBs [8,50].

Of the elements capable of impeding fork progression through indirect transcription–replication collisions, RNA–DNA hybrids have received much interest in recent years as potential sources of genome instability. RNA–DNA hybrids, or R-loops, are formed co-transcriptionally by hybridization of nascently transcribed RNA to the template DNA strand. As a result, they are enriched at active promoters and open chromatin, but are also frequently found at insulators, enhancers and terminators [51,52]. While native RNA–DNA hybrids are transient and appear to be mostly innocuous, genotypes that stabilize these structures (i.e. RNA processing and regulatory mutants) result in an accumulation of R-loops that are commonly

associated with DNA damage [53–56]. These stabilized structures can then physically impede replication fork progression, resulting in a stalled fork or incomplete replication, but may also promote the formation of single-stranded breaks (SSBs) or even DSBs by endonucleolytic cleavage on the exposed single-stranded DNA [49,57–60]. More recently, however, the mechanism of toxicity induced by these structures has been called into question with evidence that R-loop-triggered chromatin compaction may be the true underlying cause of genome instability in situations with persistent RNA–DNA hybrid accumulation [61].

Replication-independent breaks

Transcription can also trigger DNA breaks independently of DNA replication, events that are frequently observed in postmitotic (nondividing) cell populations [62]. Replication-independent DSBs within transcriptionally active regions are proposed to be mediated through at least two mechanisms: the action of topoisomerases in transcriptional initiation and elongation (Fig. 2C), and the induction of two proximal SSBs on opposite DNA strands at RNA–DNA hybrids (Fig. 2D).

The first mechanism is catalysed by topoisomerase I (TOP I) and TOP II promoting gene expression (Fig. 2C) [63]. Topoisomerases are a class of enzymes that generate temporary DNA breaks to resolve topological constraints during replication and transcription, among many other functions [64–67]. It is widely

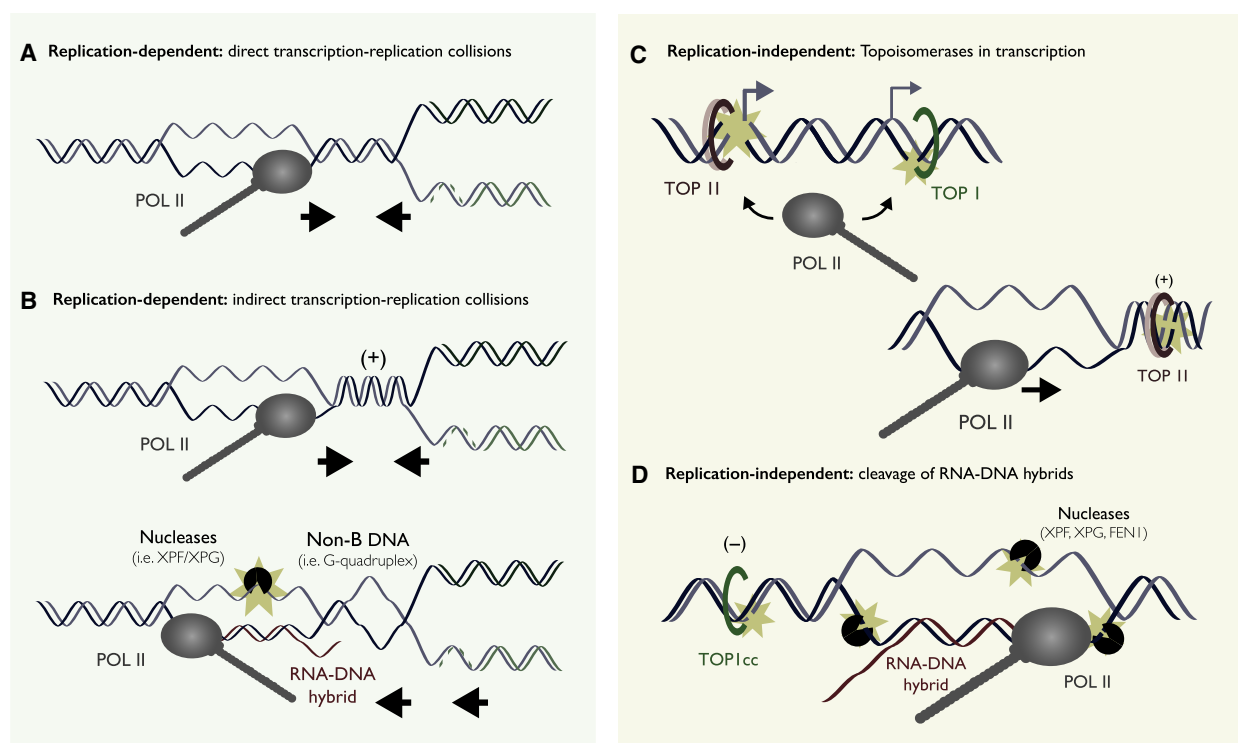


Fig. 2. Active transcription as a source of genome instability. (A) During DNA replication, transcriptionally active regions are more prone to genome instability through direct transcription–replication collisions. (B) Indirect transcription–replication collisions are mediated by elements such as RNA–DNA hybrids, non-B DNA structures and DNA supercoiling. Transcription–replication collisions are typically observed when these elements are artificially stabilized or when the replication or transcription programmes are perturbed. (C) Independent of DNA replication, TOP I and TOP II can generate transient SSBs or DSBs in transcriptional activation. Both TOP I and TOP II promote expression of highly expressed genes and enhance the recruitment of RNA polymerase II. In addition, supercoiling behind and ahead of the transcription bubble are resolved by the combined actions of TOP I and TOP II to promote transcription. (D) Independent of DNA replication, transcription can result in DSBs through two single-strand breaks (SSBs) generated on opposing DNA strands. Dual processing of topoisomerase I cleavage complexes (TOP1cc) and RNA–DNA hybrids by endonucleases (i.e. XPF, XPG, FEN1) results in proximal SSBs, forming DSBs upon strand separation. Stabilization of RNA–DNA hybrids is required for break formation.

considered that topoisomerases (e.g. TOP II β) cleave DNA to relieve topological constraints hindering RNA polymerase progression, allowing transcription elongation of long genes [68]. However, there is emerging evidence that both TOP I and TOP II can also function in transcriptional initiation itself. Although the mechanistic details behind this role are not fully understood, it has been shown that both TOP I and TOP II can promote the initial recruitment of RNA polymerase II [69]. TOP I generates DNA nicks (SSBs) to initiate RNA synthesis at many enhancer sites [70]. TOP II β , on the other hand, activates transcription by introducing DSBs at the promoters of early-response genes following a wide range of cellular stimuli including oestrogen, androgens, insulin, glucocorticoids and serum, during the process of learning and memory or upon metabolic shifting [71–77]. Remarkably, DNA cleavage by a restriction endonuclease was able to substitute for TOP II β in this capacity, suggesting that

break formation itself is sufficient for transcription activation [74]. Although breaks generated by DNA topoisomerases are normally transient and easily reversible, the conversion from reversible to irreversible breaks can be accelerated by active transcription [28]. The role of topoisomerases in transcription initiation thereby provides an alternative explanation for why actively transcribed regions exhibit higher rates of break formation.

The second mechanism involves the combined action of TOP I and endogenous endonucleases at RNA–DNA hybrids. [60,78,79] (Fig. 2D). RNA–DNA hybrids have been shown to induce genome instability by triggering replication–transcription collisions, as discussed above. However, more recent evidence suggests that R-loops may also induce genome instability independently of DNA replication. In postmitotic cells, endogenous endonucleases can induce SSBs by directly targeting RNA–DNA hybrids, or single-stranded DNA

exposed by RNA-DNA hybrids [60,78,79]. When these SSBs are in close proximity and on opposing strands, they can result in DSBs upon strand separation. Importantly, exogenous perturbations (i.e. RNA-DNA hybrid stabilization, TOP I trapping) seem to be necessary to detect these breaks.

Given that R-loop-mediated genome instability is typically characterized in cells where RNA-DNA hybrids are elevated above physiological levels, it remains unclear whether RNA-DNA hybrids are universally toxic in native conditions and how, if at all, they contribute to the break susceptibility observed at actively transcribed regions in physiological conditions. Some insight into this issue comes from two recent studies. Firstly, Promonet *et al.* profiled DSB formation genome-wide in both unperturbed wild-type cells and cells with artificially increased RNA-DNA hybrid levels. In wild-type cells, endogenous RNA-DNA hybrids were enriched at both transcription start sites and transcription termination sites, without any corresponding enrichment of DNA DSBs. Conversely, inducing RNA-DNA hybrids above physiological levels resulted in an accumulation of DNA DSBs at transcription termination sites, which the authors propose reflect an increased incidence of replication-transcription collisions. Therefore, endogenous RNA-DNA hybrids on their own did not appear to be sufficient to induce detectable DSBs [58]. Secondly, Chakraborty *et al.* [80] reported an enrichment of spontaneous DSBs at RNA-DNA-hybrid-forming sequences in cells exposed to the control solvent DMSO, but not in untreated cells. Together, these studies question the view that physiologically formed RNA-DNA hybrids induce DNA DSBs, at least at detectable levels, and also emphasize the importance of using appropriate solvent controls when profiling spontaneous DSBs.

Loop boundary regions: insulators harbour intrinsic fragility

Another genome instability hotspot identified in these breakomes is the boundary regions between topologically associated domains (TADs). TADs are considered to be the building blocks of the three-dimensional genome, with an average of ~3000 of these insulated domains in mammalian genomes [81–83]. TADs facilitate key DNA interactions (e.g. between promoters and enhancers), thereby spatially defining regulatory domains and restricting molecular activities within domains [82]. In humans, insulator proteins such as CCCTC-binding factor (CTCF) and cohesin bind the boundaries of TADs to aid compartmentalization [84–86]. At TAD boundaries, TOP II β also localizes with CTCF and cohesin to resolve any topological constraints imposed by the insulation process [87] (Fig. 3). Recently, a number of breakomes have found that TAD boundaries (i.e. CTCF binding sites) are highly susceptible to breakage in unperturbed conditions and especially when TOP II is trapped with etoposide [10,26–30,88,89].

Given that topoisomerase-induced breaks are typically reversible, the molecular mechanisms underlying the susceptibility of TAD boundaries to breakage remain elusive. Early evidence suggests that DSB formation after TOP II trapping is dependent on transcription activity at boundary regions and that depletion of the homologous recombination mediator RAD51 increases break susceptibility at insulator sites [26,28,30]. Therefore, it has been suggested that transient breaks induced by TOP II β can be turned into persistent DSBs through the act of transcription, requiring the DSB repair protein RAD51 to monitor and repair breaks (Fig. 3). Given the presence of an active repair mechanism, it remains unclear how

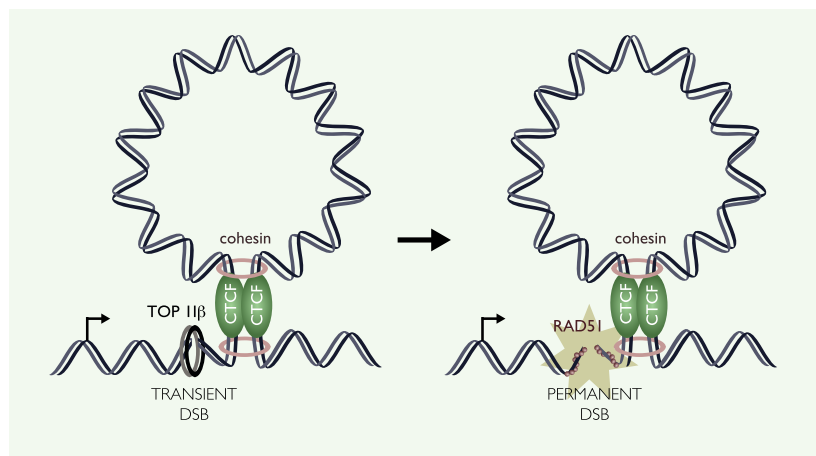


Fig. 3. Loop boundaries as sites of genome instability. Boundary regions between TADs are maintained by insulator proteins CTCF and cohesin. TOP II β generates transient DSBs to alleviate regional topological stress. TAD boundaries are highly susceptible to DNA breaks even in the absence of exogenous perturbation and especially under conditions of TOP II trapping. The conversion of transient to permanent DSBs is accelerated by nearby transcription. The single-strand DNA binding protein RAD51 protects against DSBs at loop boundaries.

persistent these breaks would be in healthy cells, although spontaneous breaks are indeed detected at loop boundaries [27,29,30]. As an alternative explanation for the DNA breakage frequently observed at loop boundary regions, TAD boundaries are enriched at sites of replication initiation [90–92]. As discussed below, replication initiation sites are inherently prone to DNA breakage, although the mechanisms underlying this fragility also remain unclear.

Early-replicating regions: a poorly understood source of break susceptibility

A surprising outcome of recently profiled breakomes is the identification of early-replicating regions as being highly susceptible to DNA breakage. Indeed, while there have been sites identified as ‘early-replicating fragile sites’ (ERFSs) since 2013, it now appears that the susceptibility to breakage extends beyond these originally defined sites to all early-replicating regions. ERFSs were defined in primary mouse cells as regions enriched with DNA damage marker γ H2A.X or bound by DNA repair proteins [RPA and Breast cancer 1 gene (BRCA1)] in response to replication stress [93]. Analogous to CFSs, discussed extensively in the following section, ERFSs are expressed as gaps in metaphase spreads, which were believed to represent DNA breakage. However, while CFSs tend to be found within late-replicating, AT-rich regions and within large isolated genes, ERFSs are largely early-replicating, GC-rich and within highly transcribed gene clusters. Supporting ERFSs as sites susceptible to breakage, more recent studies have confirmed ERFSs as break site hotspots, or identified similar sites in other organisms [40,94,95].

More strikingly, the study that has investigated the susceptibility of ERFSs to DNA breakage find that the vast majority (76%) of DSBs, while being within early-replicating regions, are outside of formally defined ERFSs [40]. This suggests that early-replicating regions as a whole may be inherently susceptible to breakage upon replicative stress regardless of their coincidence with the originally defined ERFSs. Although it is perhaps expected that breaks occur at early-replicating regions under conditions of drug-induced replication stress, there is some indication that the act of early replication itself predisposes breakage even in unperturbed conditions. For example, studies correlating breakome hotspots with epigenomic features identify early replication timing as the single greatest predictor of DSB formation in native conditions in both cancer and normal cell lines (i.e. Michigan Cancer Foundation-7 (MCF7) and Normal

human epidermal keratinocytes (NHEK), respectively) and using three independent DSB detection assays [89].

Due to the high gene density in early-replicating regions, the induction of these breaks could be explained by replication–transcription collisions that are more likely to occur in these regions. However, almost none of the break sites within early-replicating regions occur within gene bodies [40]. In addition, early replication timing was found to be more predictive of DSB formation than the classical marks of active transcription such as polymerase II β occupancy or DNase hypersensitivity, suggesting that these breaks are primarily triggered by the events uniquely associated with early replication, independently of transcription [89]. An alternative hypothesis is that a common feature of both early replication and active transcription might be imparting this shared break susceptibility (Fig. 4). Tubbs *et al.* proposed that break susceptibility at early-replicating regions is due to the presence of poly(dA:dT) tracts present in these regions. Poly(dA:dT) tracts are conserved repetitive elements in eukaryotes that have reduced nucleosome binding, making them form intrinsically open chromatin [96–98]. Poly(dA:dT) tracts have also been associated with replication initiation sites in lower eukaryotes and transcriptional activation [99–101]. Therefore, poly(dA:dT) tracts may play critical roles in both replication initiation and transcription control, at the expense of genome stability. Tubbs *et al.* propose that poly-dA strands are unable to bind the protective single-stranded DNA binding protein replication protein A (RPA) when unwinding during DNA replication. Poly-dA strands thereby act as polar replication fork barriers (RFBs), resulting in breaks at replication initiation zones and early-replicating genes. It is noteworthy that this would occur even in the absence of exogenous replication stress. In line with this notion, the DNA repair factor RAD51 similarly fails to bind the highly rigid poly-dA tracts, preventing the homology-directed repair of any breaks that do occur in these regions [102]. Nevertheless, the striking susceptibility of early-replicating regions to DNA breakage remains to be further investigated.

The observation that early replication, rather than late replication, acts as a better predictor for breakage challenges the widely accepted paradigm that late-replicating regions are more likely to be genome instability hotspots. This paradigm is conceivably based on the observation that late-replicating regions tend to form visible gaps in mitotic chromosome following exposure to replication stress [103,104] and that mutation rates are higher in late-replicating

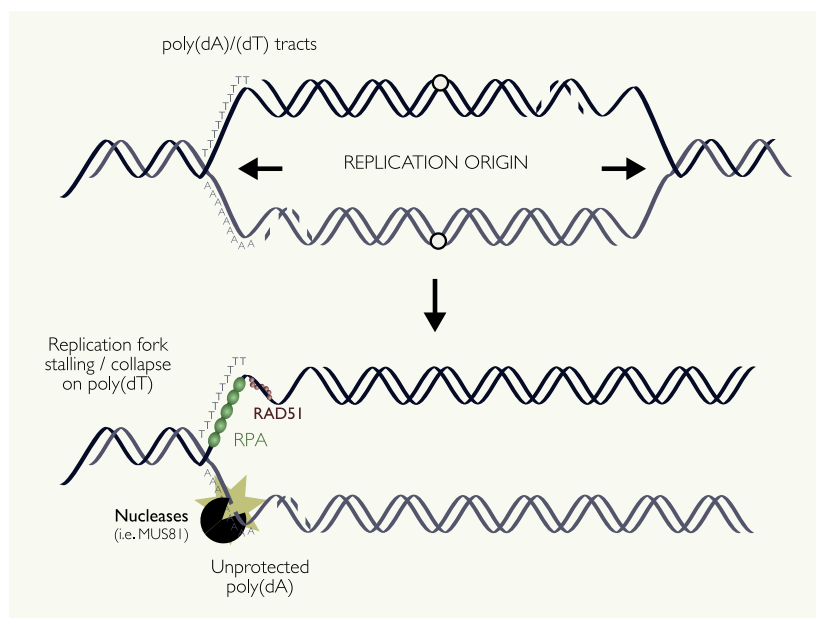


Fig. 4. Early-replicating regions as sources of genome instability. Early-replicating regions can be susceptible to breakage in the absence of any exogenous perturbation, outperforming markers of active transcription as predictors of DNA break susceptibility. The mechanisms of DNA breakage at early-replicating regions remain unknown, but it is possibly mediated by poly(dA:dT) tracts. Poly(dA:dT) is conserved repetitive elements, forming intrinsically open chromatin and acting as replication and transcription initiation sites. Poly(dA:dT) tracts are unable to bind protective single-stranded DNA binding proteins RPA and RAD51, possibly making them susceptible to DNA breakage by endonuclease activity.

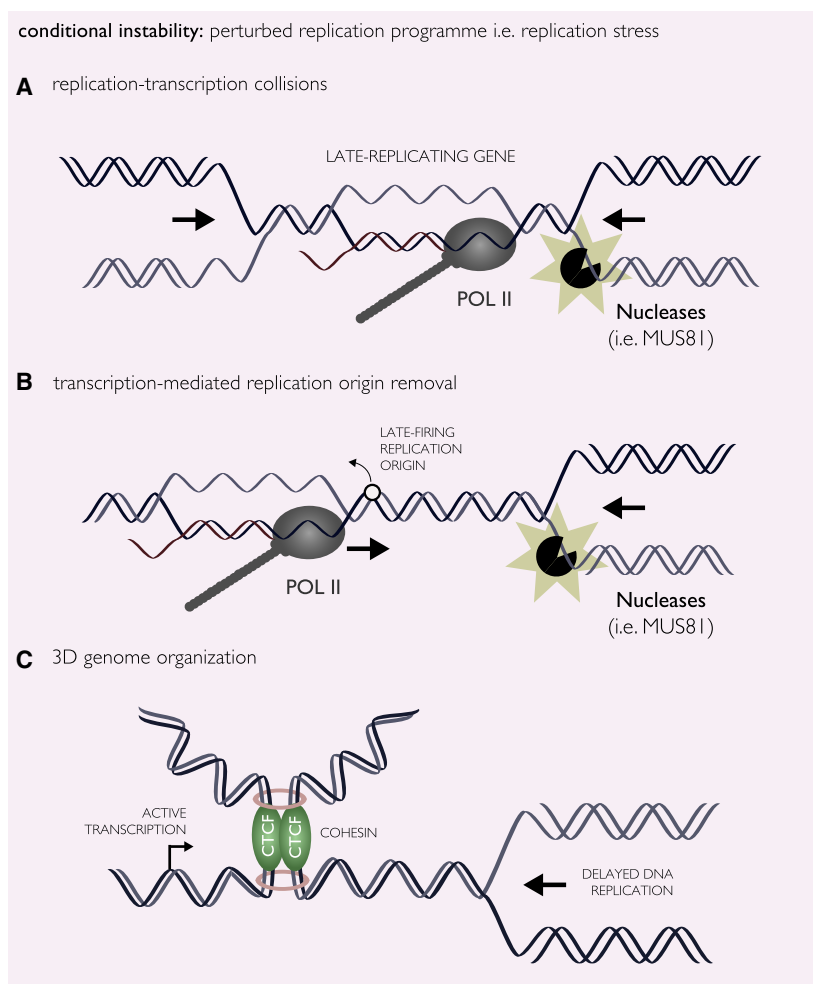
regions [105,106]. However, recently published breakome datasets provide evidence that early-replicating regions may in fact be far more susceptible to DNA breakage in the absence of exogenous perturbations such as drug-induced replication stress. The discrepancy in mutation rates can potentially be explained by considering the repair pathways and opportunities afforded to either early- or late-replicating regions. For example, early-replicating regions, while perhaps being more susceptible to breakage, would have more time to complete repair before entry into mitosis. In addition, there is evidence that high-fidelity homology-directed repair mechanisms are more frequently detectable in early-replicating regions [107]. Late-replicating regions may also experience depleted or imbalanced DNA nucleotide pools which can increase local mutation rates specifically at late-replicating regions [108]. Together, these may explain why mutation rates appear higher at late-replicating regions.

Fragile sites: not so fragile after all?

Fragile sites are genomic regions originally identified as chromosomal gaps on metaphase spreads and sites of subsequent chromosomal rearrangements, both of which are observed in response to replication stress [13,109,110]. Fragile sites are classified as either CFS (found in all human genomes) or rare fragile sites (RFS, found in only 5% of human genomes). CFSs replicate late in the cell cycle and frequently overlap with very long (> 300 kbp) genes [103,111,112].

Historically, the mechanism of CFS instability has been proposed to be based on replication–transcription collisions (Fig. 5A), as CFSs typically overlap with large transcriptional units and the formation of cytological gaps at CFSs is dependent on the transcription of these long genes [39,41]. However, this model has been challenged by more recent studies. For example, it has been demonstrated that transcription at late-replicating genes removes licensed and unfired replication origins, forcing single replication forks to replicate large stretches of DNA [104,113,114] (Fig. 5B). Such events might increase the chance of forming stalled or collapsed replication forks that can then be targeted by endonucleases for cleavage [8,50], providing an alternative explanation for gap formation. Emerging evidence also proposes an involvement of nuclear architecture in the expression of CFSs. Namely, delayed replication upon replicative stress and transcriptional activity is necessary, but not entirely sufficient, to predict gap formation at CFSs. Instead, gap formation also requires the presence of TAD boundaries at these large, late-replicating, transcribed genes [115,116] (Fig. 5C), or the attachment of stressed replication forks to the nuclear matrix [117,118]. While most of these models involve eventual DNA breakage as the source of CFS instability, it cannot be totally excluded that cytological gaps formed at CFSs actually represent chromatin compaction defects rather than physical DNA breaks. Indeed, recent evidence shows that CFSs have faulty condensin loading following replication stress, which would impede chromatin compaction during mitosis

Fig. 5. CFSs as sources of genome instability. In response to replication stress conditions, CFSs form visible gaps on metaphase spreads. Three models explaining underlying mechanisms for CFSs formation and associated genome instability are depicted. (A) The canonical model that CFS fragility is mediated through replication–transcription collisions at long genes during late DNA replication. These collisions result in endonuclease-mediated cleavage of the stalled or collapsed replication fork to initiate replication restart late in DNA replication, or even in mitosis. (B) Transcription of long genes may remove late-firing replication origins, increasing the distance by which single replication forks have to cover without stalling or collapsing. (C) A more recent model proposing the involvement of 3D genome organization in CFS instability. In this model, delayed replication, active transcription and proximity to TAD boundaries are all required to predict CFS expression in response to replication stress.



and form the chromosomal gaps visible on metaphase spreads [119].

DNA breakomes, identified by direct DSB detection, could address at least two key questions with regards to CFSs. Firstly, they can determine whether the gaps formed on metaphase chromosomes corresponds to physical DNA breakage or some other form of DNA instability, such as chromatin condensation defects. Three of the published breakome studies, all of which were performed upon exposure to low-dose replication stress, have commented on an association between DNA break sites and CFSs [23,40,80]. Specifically, Crosetto *et al.* [23] were the first to report that many CFSs were susceptible to breakage upon low-dose replication stress in human cells in 2013. In line with this observation, Tubbs *et al.* in 2018 identified clusters of breaks within the mouse orthologs of human CFS-associated genes (Fragile site, aphidicolin type, common, fra(16)(q23.2) (FRA16D) and Fragile site, aphidicolin type, common, fra(3)(p14.2) (FRA3B)), albeit that CFSs have not been formally defined in

mice. This study also identified a significant enrichment of DNA breaks within poly d(A:dT) sequences of long (> 500 kbp), late-replicating and highly transcribed genes. DNA breaks within these regions were dependent on their active transcription, analogous to human CFSs [40]. An independent study using indirect DNA break detection methods (i.e. mapping sites of chromosome translocations) has also found evidence of recurrent DSB clusters in certain CFSs upon low-dose replication stress, suggesting a subset of CFSs undergo physical DNA breakage when exposed to mild replicative stress [21]. Intriguingly, the most recent study, Chakraborty *et al.* in 2020 [80], found no correlation between the DNA break sites identified genome-wide and CFSs, although the authors speculate that this may be due to differences in solvents. Together, while initial evidence suggests that physical DNA breakage may correlate to CFS gap formation in at least some CFSs, the extent to which metaphase gaps correspond to DNA breakage remains to be fully determined.

Secondly, sequencing-based breakomes capable of detecting rare events provide opportunities to determine the conditions in which CFSs are susceptible to DNA breakage. As discussed above, CFS gap formation on metaphase spreads is typically observed in cancerous cells or in response to specific replicative stress. It is unclear whether CFS instability (i.e. DNA breakage) is also manifested in unperturbed conditions, but simply below the detection level afforded by cytology-based assays. This question ultimately addresses whether DNA breakage or instability at CFSs frequently occurs in unperturbed cells, or only when cells have transformed sufficiently to exhibit chronic replication stress. Beyond the three aforementioned studies conducted upon exposure to mild replicative stress, there is a notable absence of CFSs reported as break sites in other studies, even in replicating, highly stressed cancerous cells (Table 1) [120]. This absence ultimately poses questions as to what extent CFSs truly are a source of DNA break susceptibility in physiological settings, despite the longstanding belief that CFSs are frequent hotspots of DNA breakage.

Although chromosomal translocations at CFSs have been observed in cell lines in response to replicative stress *in vitro*, it is now increasingly recognized that CFSs are not frequently sites of gross chromosomal rearrangements in human cancers [21,121]. This contrasts to well-described oncogene-producing chromosomal translocations, for example, 9q34.1 and 22q11.2 resulting Abelson tyrosine-protein kinase 1 (ABL1)-Breakpoint cluster region protein (BCR) fusion in Philadelphia chromosome in leukaemia [122]. In addition, CFSs do not overlap with oncogenes. Rather, CFSs appear to overlap with focal deletions observed in cancer cells, including tumour suppressor genes (i.e. Fragile histidine triad protein in FRA3B and WW domain-containing oxidoreductase in FRA16D) [111,123–129]. Nonetheless, the general significance of deletions at CFSs has more recently been called into question, as only one CFS (Xp22.3) deletion was predicted to act as an early somatic driver event among 38 tumour types [130,131]. Together, these observations support the idea that DNA breakage and rearrangements at CFSs act primarily as secondary events, in response to elevated replication stress, contributing to an already unstable and evolving cancer landscape.

rDNA, telomeres and centromeres: exploring the fragility of repetitive regions

Given that breakomes have been, to date, based on short-read sequencing technologies, repetitive regions are typically excluded from analyses due to being

difficult to align with read-mapping algorithms [132]. This is despite the fact that repetitive regions may comprise up to two-thirds of the human genome and are generally considered to be significant threats to genome stability [133,134]. However, a few breakome analyses have included repetitive regions, as outlined below. From these studies, it appears that certain repetitive regions can indeed act as hotspots for genome instability, but sometimes only in perturbed conditions such as drug-induced replicative stress.

rDNA

The ribosomal RNA-encoding genes (rDNA) are composed of long (~ 45 kbp) tandem repeat arrays, which display a high level of transcription activity throughout interphase. Within rDNA arrays, a unique DNA structure that blocks replication fork progression is found at the 3' end of each transcribed genes. This structure, named RFBs, inhibits the movement of replication machinery in the direction opposite to rDNA transcription, thereby preventing transcription–replication collisions [135]. Nonetheless, with programmed fork stalling at RFBs and high transcriptional activity, rDNA repeats are frequently proposed to be a prominent source of genome instability [136]. Indeed, rDNA loci are hotspots for recombination in cancer cells [137], although whether they are a source of genome instability in physiological settings remains unclear. The notion that rDNA repeats are instability hotspots is supported by a recent report by Zhu *et al.*, in which breaks at rDNA RFBs were identified in the budding yeast *S. cerevisiae* in unperturbed conditions. In this study, these breaks were further shown to be specific to cells in S-phase, but not G1, suggesting that endogenous replication pausing at RFBs can result in spontaneous DSBs within the rDNA locus. In mouse cells, Tubbs *et al.* similarly identified spontaneous DNA DSBs at the rDNA RFBs, once again dependent on active DNA replication. Intriguingly, breakage at RFBs decreased with both low- and high-dose replication stress treatments, suggesting that these breaks are specific to unperturbed replication programmes [40]. To our best knowledge to date, however, enrichment of DNA breaks at rDNA loci has not been reported in human DNA breakomes, and it remains to be elucidated whether spontaneous DNA breakage also occurs at human rDNA repeats.

Telomeres

Telomeres are composed of very short (6-bp) repeats of GC-rich which have telomere-specific loop

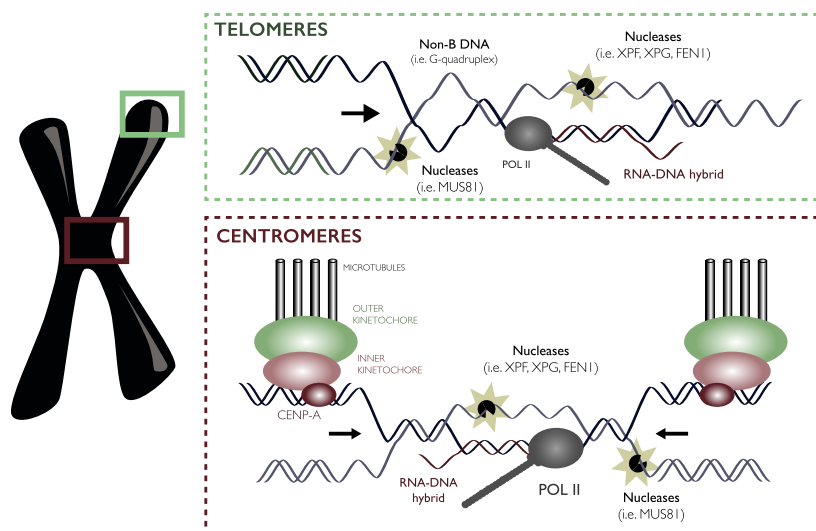


Fig. 6. Telomeres and centromeres as genome instability hotspots in response to replication stress. In response to replication stress, DNA breaks can be detected at telomere-associated repeats and alpha-satellite repeats, the main constituent of mammalian centromeres. At telomeres, RNA-DNA hybrids, non-B DNA structures (loop structures, G-quadruplexes), and transcription can impede replication fork progression. At centromeres, RNA-DNA hybrids, non-B DNA structures, transcription, and high protein occupancy can impede replication fork progression. In both scenarios, persistent RNA-DNA hybrids or failure to complete replication can provoke DNA breakage by endonucleases (i.e. XPF, XPG, FEN1, MUS81).

structures, active transcription and high RNA-DNA hybrid densities [138–141] (Fig. 6). As all of these can theoretically impede replication fork progression and result in fork stalling and collapse, there is a surprising paucity of breaks identified at or near telomeres in recent breakomes. However, it is often unclear whether telomeres are being included in these breakome analyses, as they are composed of repetitive DNA that is typically filtered out during sequencing processing. In addition, because telomeric ends can be considered as DSBs themselves, certain studies actively exclude them from analysis (e.g. Ref. [27]). One exception to this was Crosetto *et al.* 2013, where they assessed the enrichment of various human repeat sequences within break-susceptible sites in response to replication stress. They find that indeed, telomere-associated repeats are enriched within break-susceptible regions, although the enrichment is small compared to other satellite repeats such as centromeric alpha-satellite repeats [23]. As long-range sequencing technologies further develop, the fragility of these repetitive sequences will be an important question to address.

Centromeres

Centromeres are composed of relatively longer (173-bp) repeat arrays, but are also typically excluded from breakome analyses. Centromeric repeats were significantly and reproducibly detected in Crosetto *et al.* as

enriched for breaks in response to low-dose replication stress, suggesting that centromeres are prone to breakage in replication stress conditions. DNA SSBs were also notably enriched in human cancer cells at satellite repeats [29], although it was not commented on whether these included centromere-specific alpha-satellites. Chakraborty *et al.* also highlighted an apparent enrichment of DSBs at pericentromeric regions, which are similarly composed of satellite repeats as well as transposons and retrotransposons, although potential mechanisms underlying this fragility had not been examined [80]. Therefore, it seems clear that further work has to be done to assess how, if at all, (peri)centromeric regions are subjected to DNA breakage. Indeed, the repetitive nature of centromeric DNA, active transcription, RNA-DNA hybrid occupancy, late replication, high protein occupancy and substantial mechanical stress during chromosome segregation, poises centromeres as prime candidates for sites of DNA breakage during DNA replication and perhaps even during chromosome segregation in native conditions (Fig. 6) [142–147]. Aberrant DNA structures resulting from break repair are also frequently observed at centromeres in the absence of exogenous perturbation, and genome instability deriving from centromeres has recently been a topic of keen interest [148–151].

It is becoming more commonly accepted that breaks at or near pericentromeric or centromeric regions can

result in rearrangements and subsequent arm loss [151]. Direct evidence from Taylor *et al.* [152] demonstrates that CRISPR-induced breaks near centromeric regions can indeed cause chromosome arm loss. This notion is further supported by reports identifying frequent chromosome breaks around centromeric regions in tumours, suggesting that these breaks may contribute to the genome instability observed in cancer cells [153–158]. Breaks at these regions may further result in isochromosome formation, the most recurrent abnormality in certain cancers and frequently an early somatic driver event in carcinogenesis [131,159]. It should be noted, however, that centromeres do not meet the classic predictors of break susceptibility (i.e. early replication, high transcriptional activity, strong enhancer activity) according to computational modelling for DSB susceptibility [89]. Nonetheless, a class of genomic regions that show frequent breakage in carcinogenesis, but with low DSB susceptibility by modelling, are characterized by heterochromatic regions composed of repetitive DNA which the authors propose may correspond to repeat-rich regions near centromeres.

Conclusion and Future Perspectives

Genome-wide mapping of DSBs in mammalian cells has previously posed significant challenges due to technical constraints. Epigenetic markers such as γ H2A.X have frequently been used as proxies for breaks [160,161], but this mark can extend up to megabases around DSBs yet be depleted within kilobases directly surrounding the break. In addition, γ H2A.X is dependent on H2A.X densities in chromatin, which can further depend on replication timing and transcriptional activity [88,160–164]. Attempts to define the sites of DSBs have also been made through other indirect measurements, such as high-throughput genome-wide translocation sequencing or translocation-capture sequencing [31,32]. It is noteworthy that, while these translocation-based approaches capture DNA breakage events over a large window of time, translocations are typically generated by error-prone repair mechanisms, such as nonhomologous end-joining, fusing two DSB ends in close proximity [165,166]. As DNA repair pathway choice and the likelihood of translocation are dependent on the local chromatin environment [167], this may lead to a biased distribution of translocation occurrence, and thereby a biased detection of the history of DNA breakages. In contrast, techniques that directly detect physical DNA breaks offer an unbiased, comprehensive genome-wide mapping of DSB sites with high space–time resolution. These techniques are

beginning to shed new light on how DNA instability is manifested across the genome across various physiological conditions.

The DNA breakomes discussed in this review consolidate known hotspots of genome instability (e.g. actively transcribed regions), but also reveal less expected instability hotspots (e.g. boundary regions between TADs, early-replicating regions). These studies further challenge current paradigms, for example that endogenous DNA breaks are primarily through replication-based mechanisms in hard-to-replicate regions. Ultimately, DNA breakomes are beginning to objectively determine the regions of the mammalian genome most likely to initiate DNA instability in pathologies such as cancer, premature ageing and neurodegenerative disorders.

An important outstanding question is how the spatial distribution of DNA breakage actually correlates to the resulting DNA instability. As discussed above, DNA repair pathway choice is dependent on local chromatin environments, with actively transcribed regions preferring dedicated and high-fidelity repair mechanisms [167–170]. In addition, early-replicating regions may have a higher chance of being repaired before entry into mitosis. Therefore, there may be a discrepancy between DNA DSBs and the mutagenic events that ultimately influence disease onset and progression. Indeed, computational modelling reveals several instances of mismatches between predicted DSBs and structural variant densities observed in tumours [89]. It will therefore be critical to assess genome mutations and rearrangements alongside DNA breakomes to further understand how breakomes shape genome evolution, and vice versa, throughout disease progression.

As the era of next- and third-generation sequencing technologies develop, the detection and quantification of DNA breaks on a genome-wide scale will be critical to objectively profiling genome instability hotspots in various pathologies, and understanding the mechanisms protecting and repairing breaks in a spatially sensitive manner. Recent technological advancements, for example long-read sequencing approaches, may accelerate the assessment of the prevalence of DNA breaks in more enigmatic regions of the genome such as repetitive regions (e.g. rDNA, telomeres and centromeres), allowing us to understand how repetitive elements truly contribute to genome instability. In addition, the question of cell-to-cell variance in DNA break profiles remains to be resolved. Identifying such variability at single-cell level may elucidate break prevalence and distribution throughout the population, or whether certain break profiles correspond to cell cycle or transcription status.

These technologies ultimately have the potential to address key questions in the field. Understanding innate sources of vulnerability in the human genome in an unbiased fashion will inform what sorts of exogenous threats are most likely to provoke DNA instability, and thereby initiate and drive human disease. Further, breakomes conducted in specific disease states may identify conditional vulnerabilities which may then guide, for example, the development of safe and effective chemotherapeutics, with minimal off-target effects to noncancerous cells. Another important outcome of these studies is the observation that DSB profiles vary notably between cell types, and certain features are found to be predictive of DSB incidence only in certain cell lines [30,89]. This variance in cell responses calls for further investigation into why and how different cell types exhibit different instability hotspots. Given this intrinsic variance in DNA break profiles, other important questions include how DSB profiles vary between human tissues, or change throughout a single-cell cycle, stem cell differentiation, or cellular and organismal ageing. Addressing these questions will help to delineate how genome instability is invoked and protected against throughout our lifetimes and in human disease.

Acknowledgements

FE was supported by Wellcome Trust Senior Research Fellowships in Basic Biomedical Science (101009/Z/13/Z). XS receives the Oxford Cancer Centre Cancer Research UK D. Phil studentship (CRUK-OC-DPhil17-XS). We are thankful for supports from the Edward Penley Abraham Research Fund. We apologize for not being able to cite all research articles relevant to this review, due to spatial constraints.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

XS searched the literature and drafted the manuscript. FE critically reviewed and revised the manuscript. XS and FE contributed equally to edit and revise the final version.

References

- 1 Lindahl T & Barnes DE (2000) Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* **65**, 127–133.
- 2 Vilenchik MM & Knudson AG (2003) Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc Natl Acad Sci US* **100**, 12871–12876.
- 3 Hansemann D (1890) Ueber asymmetrische Zelltheilung in Epithelkrebsen und deren biologische Bedeutung. *Archiv Für Pathol Anat Physiol Für Klin Med* **119**, 299–326.
- 4 Boveri T (1914) Zur Frage der Entstehung maligner Tumoren. *Science* **40**, 857–859.
- 5 Strehler BL (1986) Genetic instability as the primary cause of human aging. *Exp Gerontol* **21**, 283–319.
- 6 Sax K (1938) Chromosome aberrations induced by X-rays. *Genetics* **23**, 494–516.
- 7 Peak MJ & Peak JG (1982) Single-strand breaks induced in bacillus subtilis DNA by ultraviolet light: action spectrum and properties. *Photochem Photobiol* **35**, 675–680.
- 8 Hanada K, Budzowska M, Davies SL, van Drunen E, Onizawa H, Beverloo HB, Maas A, Essers J, Hickson ID & Kanaar R (2007) The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. *Nat Struct Mol Biol* **14**, 1096–1104.
- 9 Thomas BJ & Rothstein R (1989) Elevated recombination rates in transcriptionally active DNA. *Cell* **56**, 619–630.
- 10 Canela A, Maman Y, Jung S, Wong N, Callen E, Day A, Kieffer-Kwon K-R, Pekowska A, Zhang H, Rao SSP *et al.* (2017) Genome organization drives chromosome fragility. *Cel* **170**, 507–521.e18.
- 11 Lengronne A & Schwob E (2002) The yeast CDK inhibitor Sic1 prevents genomic instability by promoting replication origin licensing in late G1. *Mol Cell* **9**, 1067–1078.
- 12 Sutherland GR & Richards RI (1995) Simple tandem DNA repeats and human genetic disease. *Proc Natl Acad Sci USA* **92**, 3636–3641.
- 13 Glover TW, Berger C, Coyle J & 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.
- 14 Cha RS & Kleckner N (2002) ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. *Science* **297**, 602–606.
- 15 Gorgoulis VG, Vassiliou L-VF, Karakaidos P, Zacharatos P, Kotsinas A, Liloglou T, Venere M, DiTullio RA, Kastriakis NG, Levy B *et al.* (2005) Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–913.
- 16 Lopes J, Piazza A, Bermejo R, Kriegsman B, Colosio A, Teulade-Fichou M-P, Foiani M & Nicolas A (2011) G-quadruplex-induced instability during leading-strand replication. *EMBO J* **30**, 4033–4046.

- 17 Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL & de Lange T (2009) Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138**, 90–103.
- 18 Aguilera A & García-Muse T (2013) Causes of genome instability. *Annu Rev Genet* **47**, 1–32.
- 19 Bouwman BAM & Crosetto N (2018) Endogenous DNA double-strand breaks during DNA transactions: emerging insights and methods for genome-wide profiling. *Genes* **9**, 632.
- 20 Baranello L, Kouzine F, Wojtowicz D, Cui K, Przytycka TM, Zhao K & Levens D (2014) DNA break mapping reveals topoisomerase II activity genome-wide. *Int J Mol Sci* **15**, 13111–13122.
- 21 Wei P-C, Chang AN, Kao J, Du Z, Meyers RM, Alt FW & Schwer B (2016) Long neural genes harbor recurrent DNA break clusters in neural stem/progenitor cells. *Cell* **164**, 644–655.
- 22 Schwer B, Wei P-C, Chang AN, Kao J, Du Z, Meyers RM & Alt FW (2016) Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. *Proc Natl Acad Sci USA* **113**, 2258–2263.
- 23 Crosetto N, Mitra A, Silva MJ, Bienko M, Dojer N, Wang Q, Karaca E, Chiarle R, Skrzypczak M, Ginalski K *et al.* (2013) Nucleotide-resolution DNA double-strand breaks mapping by next-generation sequencing. *Nat Methods* **10**, 361–365.
- 24 Yang F, Kemp CJ & Henikoff S (2015) Anthracyclines induce double-strand DNA breaks at active gene promoters. *Mutat Res* **773**, 9–15.
- 25 Yan WX, Mirzazadeh R, Garnerone S, Scott D, Schneider MW, Kallas T, Custodio J, Wernersson E, Li Y, Gao L *et al.* (2017) BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat Commun* **8**, 15058.
- 26 Gothe HJ, Bouwman BAM, Gusmao EG, Piccinno R, Petrosino G, Sayols S, Drechsel O, Minneker V, Josipovic N, Mizi A *et al.* (2019) Spatial chromosome folding and active transcription drive DNA fragility and formation of oncogenic MLL translocations. *Mol Cell* **75**, 267–283.e12.
- 27 Lensing SV, Marsico G, Hänsel-Hertsch R, Lam EY, Tannahill D & Balasubramanian S (2016) DSBapture: *in situ* capture and sequencing of DNA breaks. *Nat Methods* **13**, 855–857.
- 28 Canela A, Maman Y, Huang S-YN, Wutz G, Tang W, Zagnoli-Vieira G, Callen E, Wong N, Day A, Peters J-M *et al.* (2019) Topoisomerase II-induced chromosome breakage and translocation is determined by chromosome architecture and transcriptional activity. *Mol Cell* **75**, 252–266.e8.
- 29 Cao H, Salazar-García L, Gao F, Wahlestedt T, Wu C-L, Han X, Cai Y, Xu D, Wang F, Tang L *et al.* (2019) Novel approach reveals genomic landscapes of single-strand DNA breaks with nucleotide resolution in human cells. *Nat Commun* **10**, 5799.
- 30 Hazan I, Monin J, Bouwman BAM, Crosetto N & Aqeilan RI (2019) Activation of oncogenic super-enhancers is coupled with DNA repair by RAD51. *Cell Rep* **29**, 560–572.e4.
- 31 Klein IA, Resch W, Jankovic M, Oliveira T, Yamane A, Nakahashi H, Di Virgilio M, Bothmer A, Nussenzweig A, Robbiani DF *et al.* (2011) Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* **147**, 95–106.
- 32 Chiarle R, Zhang Y, Frock RL, Lewis SM, Molinie B, Ho Y-J, Myers DR, Choi VW, Compagno M, Malkin DJ *et al.* (2011) Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107–119.
- 33 Herman RK & Dworkin NB (1971) Effect of gene induction on the rate of mutagenesis by ICR-191 in *Escherichia coli*. *J Bacteriol* **106**, 543–550.
- 34 Datta A & Jinks-Robertson S (1995) Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* **268**, 1616–1619.
- 35 García-Rubio M, Huertas P, González-Barrera S & Aguilera A (2003) Recombinogenic effects of DNA-damaging agents are synergistically increased by transcription in *Saccharomyces cerevisiae*. New insights into transcription-associated recombination. *Genetics* **165**, 457–466.
- 36 Majewski J (2003) Dependence of mutational asymmetry on gene-expression levels in the human genome. *Am J Hum Genet* **73**, 688–692.
- 37 Wei X, Samarabandu J, Devdhar RS, Siegel AJ, Acharya R & Berezney R (1998) Segregation of transcription and replication sites into higher order domains. *Science* **281**, 1502–1505.
- 38 Meryet-Figuere M, Alaei-Mahabadi B, Ali MM, Mitra S, Subhash S, Pandey GK, Larsson E & Kanduri C (2014) Temporal separation of replication and transcription during S-phase progression. *Cell Cycle* **13**, 3241–3248.
- 39 Helmrich A, Ballarino M & Tora L (2011) Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell* **44**, 966–977.
- 40 Tubbs A, Sridharan S, van Wietmarschen N, Maman Y, Callen E, Stanlie A, Wu W, Wu X, Day A, Wong N *et al.* (2018) Dual Roles of Poly(dA:dT) tracts in replication initiation and fork collapse. *Cell* **174**, 1127–1142.e19.
- 41 Wilson TE, Arlt MF, Park SH, Rajendran S, Paulsen M, Ljungman M & Glover TW (2015) Large transcription units unify copy number variants and

- common fragile sites arising under replication stress. *Genome Res* **25**, 189–200.
- 42 Prado F & Aguilera A (2005) Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J* **24**, 1267–1276.
 - 43 Mirkin EV & Mirkin SM (2005) Mechanisms of transcription-replication collisions in bacteria. *Mol Cell Biol* **25**, 888–895.
 - 44 Srivatsan A, Tehranchi A, MacAlpine DM & Wang JD (2010) Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet* **6**, e1000810.
 - 45 Azvolinsky A, Giresi PG, Lieb JD & Zakian VA (2009) Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. *Mol Cell* **34**, 722–734.
 - 46 Brill SJ, DiNardo S, Voelkel-Meiman K & Sternglanz R (1987) Need for DNA topoisomerase activity as a swivel for DNA replication for transcription of ribosomal RNA. *Nature* **326**, 414–416.
 - 47 Voineagu I, Narayanan V, Lobachev KS & Mirkin SM (2008) Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc Natl Acad Sci USA* **105**, 9936–9941.
 - 48 Paeschke K, Capra JA & Zakian VA (2011) DNA replication through G-quadruplex motifs is promoted by the *S. cerevisiae* Pif1 DNA helicase. *Cell* **145**, 678–691.
 - 49 Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL & Li X (2011) R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev* **25**, 2041–2056.
 - 50 Naim V, Wilhelm T, Debatisse M & Rosselli F (2013) ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nat Cell Biol* **15**, 1008.
 - 51 Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, Xu X & Chédin F (2016) Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol Cell* **63**, 167–178.
 - 52 Crossley MP, Boczek M & Cimprich KA (2019) R-loops as cellular regulators and genomic threats. *Mol Cell* **73**, 398–411.
 - 53 Huertas P & Aguilera A (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* **12**, 711–721.
 - 54 Paulsen RD, Soni DV, Wollman R, Hahn AT, Yee M-C, Guan A, Hesley JA, Miller SC, Cromwell EF, Solow-Cordero DE *et al.* (2009) A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Mol Cell* **35**, 228–239.
 - 55 Stirling PC, Chan YA, Minaker SW, Aristizabal MJ, Barrett I, Sipahimalani P, Kobor MS & Hieter P (2012) R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev* **26**, 163–175.
 - 56 Wahba L, Amon JD, Koshland D & Vuica-Ross M (2011) RNase H and multiple RNA biogenesis factors cooperate to prevent RNA-DNA hybrids from generating genome instability. *Mol Cell* **44**, 978–988.
 - 57 Tuduri S, Crabbé L, Conti C, Tourrière H, Holtgreve-Grez H, Jauch A, Pantescio V, De Vos J, Thomas A, Theillet C *et al.* (2009) Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat Cell Biol* **11**, 1315–1324.
 - 58 Promonet A, Padioleau I, Liu Y, Sanz L, Biernacka A, Schmitz A-L, Skrzypczak M, Sarrazin A, Mettling C, Rowicka M *et al.* (2020) Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. *Nat Commun* **11**, 3940.
 - 59 Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715.
 - 60 Sollier J, Stork CT, García-Rubio ML, Paulsen RD, Aguilera A & Cimprich KA (2014) Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* **56**, 777–785.
 - 61 Rondón AG & Aguilera A (2019) What causes an RNA-DNA hybrid to compromise genome integrity? *DNA Rep* **81**, 102660.
 - 62 Iyama T & Wilson DM (2013) DNA repair mechanisms in dividing and non-dividing cells. *DNA Rep* **12**, 620–636.
 - 63 Kouzine F, Gupta A, Baranello L, Wojtowicz D, Benaissa K, Liu J, Przytycka TM & Levens D (2013) Transcription dependent dynamic supercoiling is a short-range genomic force. *Nat Struct Mol Biol* **20**, 396–403.
 - 64 Liu LF & Wang JC (1987) Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci USA* **84**, 7024–7027.
 - 65 Champoux JJ & Dulbecco R (1972) An activity from mammalian cells that untwists superhelical DNA—a possible swivel for DNA replication (polyoma-ethidium bromide-mouse-embryo cells-dye binding assay). *Proc Natl Acad Sci USA* **69**, 143–146.
 - 66 Gellert M, Mizuuchi K, O'Dea MH & Nash HA (1976) DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc Natl Acad Sci USA* **73**, 3872–3876.
 - 67 Wang JC (1971) Interaction between DNA and an *Escherichia coli* protein omega. *J Mol Biol* **55**, 523–533.

- 68 Joshi RS, Piña B & Roca J (2012) Topoisomerase II is required for the production of long Pol II gene transcripts in yeast. *Nucleic Acids Res* **40**, 7907–7915.
- 69 Sperling AS, Jeong KS, Kitada T & Grunstein M (2011) Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. *Proc Natl Acad Sci USA* **108**, 12693–12698.
- 70 Puc J, Kozbial P, Li W, Tan Y, Liu Z, Suter T, Ohgi KA, Zhang J, Aggarwal AK & Rosenfeld MG (2015) Ligand-dependent enhancer activation regulated by topoisomerase-I activity. *Cell* **160**, 367–380.
- 71 Ju B-G, Lunyak VV, Perissi V, Garcia-Bassets I, Rose DW, Glass CK & Rosenfeld MG (2006) A topoisomerase II β -mediated dsDNA break required for regulated transcription. *Science* **312**, 1798–1802.
- 72 Haffner MC, Aryee MJ, Toubaji A, Esopi DM, Albadine R, Gurel B, Isaacs WB, Bova GS, Liu W, Xu J *et al.* (2010) Androgen-induced TOP2B mediated double strand breaks and prostate cancer gene rearrangements. *Nat Genet* **42**, 668–675.
- 73 Wong RHF, Chang I, Hudak CSS, Hyun S, Kwan H-Y & Sul HS (2009) A role of DNA-PK for the metabolic gene regulation in response to insulin. *Cell* **136**, 1056–1072.
- 74 Trotter KW, King HA & Archer TK (2015) Glucocorticoid receptor transcriptional activation via the BRG1-dependent recruitment of TOP2 β and Ku70/86. *Mol Cell Biol* **35**, 2799–2817.
- 75 Bunch H, Lawney BP, Lin Y-F, Asaithamby A, Murshid A, Wang YE, Chen BPC & Calderwood SK (2015) Transcriptional elongation requires DNA break-induced signalling. *Nat Commun* **6**, 10191.
- 76 Madabhushi R, Gao F, Pfenning AR, Pan L, Yamakawa S, Seo J, Rueda R, Phan TX, Yamakawa H, Pao P-C *et al.* (2015) Activity-induced DNA breaks govern the expression of neuronal early-response genes. *Cell* **161**, 1592–1605.
- 77 Roedgaard M, Fredsoe J, Pedersen JM, Bjergbaek L & Andersen AH (2015) DNA topoisomerases are required for preinitiation complex assembly during GAL gene activation. *PLoS One* **10**, e0132739.
- 78 Sordet O, Redon CE, Guirouilh-Barbat J, Smith S, Solier S, Douarre C, Conti C, Nakamura AJ, Das BB, Nicolas E *et al.* (2009) Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. *EMBO Rep* **10**, 887–893.
- 79 Cristini A, Ricci G, Britton S, Salimbeni S, Huang S-YN, 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 Rep* **28**, 3167–3181.e6.
- 80 Chakraborty A, Jenjaroenpun P, Li J, El Hilali S, McCulley A, Haarer B, Hoffman EA, Belak A, Thorland A, Hehnly H *et al.* (2020) Replication stress induces global chromosome breakage in the fragile X Genome. *Cell Rep* **32**, 108179.
- 81 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS & Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380.
- 82 Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation center. *Nature* **485**, 381–385.
- 83 Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A & Cavalli G (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472.
- 84 Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T *et al.* (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801.
- 85 Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680.
- 86 Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, Ong C-T, Hookway TA, Guo C, Sun Y *et al.* (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295.
- 87 Uusküla-Reimand L, Hou H, Samavarchi-Tehrani P, Rudan MV, Liang M, Medina-Rivera A, Mohammed H, Schmidt D, Schwalie P, Young EJ *et al.* (2016) Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol* **17**, 182.
- 88 Mourad R, Ginalska K, Legube G & Cuvier O (2018) Predicting double-strand DNA breaks using epigenome marks or DNA at kilobase resolution. *Genome Biol* **19**, 34.
- 89 Ballinger TJ, Bouwman BAM, Mirzazadeh R, Garnerone S, Crosetto N & Semple CA (2019) Modeling double strand break susceptibility to interrogate structural variation in cancer. *Genome Biol* **20**, 28.
- 90 Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, Vera DL, Wang Y, Hansen RS, Canfield TK *et al.* (2014) Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–405.
- 91 Lombardi EP & Tarsounas M (2020) Topologically associating domain boundaries are enriched in early

- firing origins and restrict replication fork progression. *bioRxiv*. 10.21.348946. "[PREPRINT]"
- 92 Moindrot B, Audit B, Klous P, Baker A, Thermes C, de Laat W, Bouvet P, Mongelard F & Arneodo A (2012) 3D chromatin conformation correlates with replication timing and is conserved in resting cells. *Nucleic Acids Res* **40**, 9470–9481.
 - 93 Barlow J, Faryabi RB, Callen E, Wong N, Malhowski A, Chen HT, Gutierrez-Cruz G, Sun H-W, McKinnon P, Wright G *et al.* (2013) A novel class of early replicating fragile sites that contribute to genome instability in B cell lymphomas. *Cell* **152**, 620–632.
 - 94 Pentzold C, Shah SA, Hansen NR, Le Tallec B, Seguin-Orlando A, Debatisse M, Lisby M & Oestergaard VH (2018) FANCD2 binding identifies conserved fragile sites at large transcribed genes in avian cells. *Nucleic Acids Res* **46**, 1280–1294.
 - 95 Waisertreiger I, Popovich K, Block M, Anderson KR & Barlow JH (2020) Visualizing locus-specific sister chromatid exchange reveals differential patterns of replication stress-induced fragile site breakage. *Oncogene* **39**, 1260–1272.
 - 96 Dechering KJ, Konings RNH, Cuelenaere K & Leunissen JAM (1998) Distinct frequency-distributions of homopolymeric DNA tracts in different genomes. *Nucleic Acids Res* **26**, 4056–4062.
 - 97 Rhodes D (1979) Nucleosome cores reconstituted from poly (dA-dT) and the octamer of histones. *Nucleic Acids Res.* **6**, 1805–1816.
 - 98 Suter B, Schnappauf G & Thoma F (2000) Poly (dA-dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Res* **28**, 4083–4089.
 - 99 Zhu J, Carlson DL, Dubey DD, Sharma K & Huberman JA (1994) Comparison of the two major ARS elements of the *ura4* replication origin region with other ARS elements in the fission yeast, *Schizosaccharomyces pombe*. *Chromosoma* **103**, 414–422.
 - 100 Struhl K (1985) Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc Natl Acad Sci USA* **82**, 8419–8423.
 - 101 Iyer V & Struhl K (1995) Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J* **14**, 2570–2579.
 - 102 Paoletti F, El-Sagheer AH, Allard J, Brown T, Dushek O & Esashi F (2020) Molecular flexibility of DNA as a key determinant of RAD51 recruitment. *EMBO J* **39**, e103002.
 - 103 Le Beau MM, Rassool FV, Neilly ME, Espinosa Iii R, Glover TW, Smith DI & McKeithan TW (1998) Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Hum Mol Genet* **7**, 755–761.
 - 104 Letessier A, Millot GA, Koundrioukoff S, Lachagès A-M, Vogt N, Hansen RS, Malfoy B, Brison O & Debatisse M (2011) Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* **470**, 120–123.
 - 105 Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM & Sunyaev SR (2009) Human mutation rate associated with DNA replication timing. *Nat Genet* **41**, 393–395.
 - 106 Chen C-L, Rappailles A, Duquenne L, Huvet M, Guilbaud G, Farinelli L, Audit B, d'Aubenton-Carafa Y, Arneodo A, Hyrien O *et al.* (2010) Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res* **20**, 447–457.
 - 107 Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR & McCarroll SA (2012) Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am J Hum Genet* **91**, 1033–1040.
 - 108 Watt DL, Buckland RJ, Lujan SA, Kunkel TA & Chabes A (2016) Genome-wide analysis of the specificity and mechanisms of replication infidelity driven by imbalanced dNTP pools. *Nucleic Acids Res* **44**, 1669–1680.
 - 109 Durkin SG & Glover TW (2007) Chromosome fragile sites. *Annu Rev Genet* **41**, 169–192.
 - 110 Hirsch B (1991) Sister chromatid exchanges are preferentially induced at expressed and nonexpressed common fragile sites. *Hum Genet* **87**, 302–306.
 - 111 Le Tallec B, Millot GA, Blin ME, Brison O, Dutrillaux B & Debatisse M (2013) Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep* **4**, 420–428.
 - 112 Helmrich A, Stout-Weider K, Hermann K, Schrock E & Heiden T (2006) Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. *Genome Res* **16**, 1222–1230.
 - 113 Brison O, El-Hilali S, Azar D, Koundrioukoff S, Schmidt M, Nähse V, Jaszczyzyn Y, Lachages A-M, Dutrillaux B, Thermes C *et al.* (2019) Transcription-mediated organization of the replication initiation program across large genes sets common fragile sites genome-wide. *Nat Commun* **10**, 5693.
 - 114 Blin M, Le Tallec B, Nähse V, Schmidt M, Brossas C, Millot GA, Prioleau M-N & Debatisse M (2019) Transcription-dependent regulation of replication dynamics modulates genome stability. *Nat Struct Mol Biol* **26**, 58–66.
 - 115 Sarni D, Sasaki T, Irony Tur-Sinai M, Miron K, Rivera-Mulia JC, Magnuson B, Ljungman M, Gilbert DM & Kerem B (2020) 3D genome organization

- contributes to genome instability at fragile sites. *Nat Commun* **11**, 3613.
- 116 Kaiser VB & Semple CA (2018) Chromatin loop anchors are associated with genome instability in cancer and recombination hotspots in the germline. *Genome Biol* **19**, 101.
 - 117 Pentzold C, Kokal M, Pentzold S & Weise A (2020) Sites of chromosomal instability in the context of nuclear architecture and function. *Cell Mol Life Sci* **78**, 2095–2103.
 - 118 Guadarrama-Ponce R & Aranda-Anzaldo AA-O (2020) The epicenter of chromosomal fragility of Fra14A2, the mouse ortholog of human FRA3B common fragile site, is largely attached to the nuclear matrix in lymphocytes but not in other cell types that do not express such a fragility. *J Cell Biochem* **121**, 2209–2224.
 - 119 Boteva L, Nozawa R-S, Naughton C, Samejima K, Earnshaw WC & Gilbert N (2020) Common fragile sites are characterised by faulty condensin loading after replication stress. *bioRxiv*, 508713. “[PREPRINT]”.
 - 120 Hills SA & Diffley JFX (2014) DNA replication and oncogene-induced replicative stress. *Curr Biol* **24**, R435–444.
 - 121 Glover TW, Wilson TE & Arlt MF (2017) Fragile sites in cancer: more than meets the eye. *Nat Rev Cancer* **17**, 489–501.
 - 122 Shtivelman E, Lifshitz B, Gale RP & Canaani E (1985) Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* **315**, 550–554.
 - 123 Huebner K, Garrison PN, Barnes LD & Croce CM (1998) The role of the Fhit/Fra3b locus in cancer. *Ann Rev Genet* **32**, 7–31.
 - 124 McAvoy S, Ganapathiraju SC, Ducharme-Smith AL, Pritchett JR, Kosari F, Perez DS, Zhu Y, James CD & Smith DI (2007) Non-random inactivation of large common fragile site genes in different cancers. *Cytogenet Genome Res* **118**, 260–269.
 - 125 Bignell GR, Greenman CD, Davies H, Butler AP, Edkins S, Andrews JM, Buck G, Chen L, Beare D, Latimer C *et al.* (2010) Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–898.
 - 126 Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhang C-Z, Wala J, Mermel CH *et al.* (2013) Pan-cancer patterns of somatic copy number alteration. *Nat Genet* **45**, 1134–1140.
 - 127 Zimonjic DB, Druck T, Ohta M, Kastury K, Croce CM, Popescu NC & Huebner K (1997) Positions of chromosome 3p14.2 fragile sites (FRA3B) within the FHIT Gene. *Cancer Res* **57**, 1166–1170.
 - 128 Bednarek AK, Laflin KJ, Daniel RL, Liao Q, Hawkins KA & Aldaz CM (2000) WWOX, a Novel WW domain-containing protein mapping to human chromosome 16q23.3–24.1, a region frequently affected in breast cancer. *Cancer Res* **60**, 2140–2145.
 - 129 Gao G & Smith DI (2014) Very large common fragile site genes and their potential role in cancer development. *Cell Mol Life Sci* **71**, 4601–4615.
 - 130 Arlt MF, Miller DE, Beer DG & Glover TW (2002) Molecular characterization of FRAXB and comparative common fragile site instability in cancer cells. *Genes Chromosomes Cancer* **33**, 82–92.
 - 131 Gerstung M, Jolly C, Leshchiner I, Dentre SC, Gonzalez S, Rosebrock D, Mitchell TJ, Rubanova Y, Anur P, Yu K *et al.* (2020) The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128.
 - 132 Altemose N, Miga KH, Maggioni M & Willard HF (2014) Genomic characterization of large heterochromatic gaps in the human genome assembly. *PLOS Comput Biol* **10**, e1003628.
 - 133 de Koning AP, Gu W, Castoe TA, Batzer MA & Pollock DD (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLOS Genet* **7**, e1002384.
 - 134 Madireddy A & Gerhardt J (2017) Replication Through Repetitive DNA Elements and Their Role in Human Diseases. In *DNA Replication: From Old Principles to New Discoveries* (Masai H & Foiani M, eds) pp. 549–581. Springer, Singapore.
 - 135 Little RD, Platt TH & Schildkraut CL (1993) Initiation and termination of DNA replication in human rRNA genes. *Mol Cell Biol* **13**, 6600–6613.
 - 136 Salim D & Gerton JL (2019) Ribosomal DNA instability and genome adaptability. *Chromosome Res* **27**, 73–87.
 - 137 Stults DM, Killen MW, Williamson EP, Hourigan JS, Vargas HD, Arnold SM, Moscow JA & Pierce AJ (2009) Human rRNA gene clusters are recombinational hotspots in cancer. *Cancer Res* **69**, 9096–9104.
 - 138 Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H & De Lange T (1999) Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514.
 - 139 Doksan Y, Wu JY, de Lange T & Zhuang X (2013) Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. *Cell* **155**, 345–356.
 - 140 Azzalin CM, Reichenbach P, Khoraiuli L, Giulotto E & Lingner J (2007) Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**, 798–801.
 - 141 Parkinson GN, Lee MPH & Neidle S (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876–880.
 - 142 Manuelidis L (1978) Chromosomal localization of complex and simple repeated human DNAs. *Chromosoma* **66**, 23–32.

- 143 Mitchell AR, Gosden JR & Miller DA (1985) A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. *Chromosoma* **92**, 369–377.
- 144 Chan FL, Marshall OJ, Saffery R, Won Kim B, Earle E, Choo KHA & Wong LH (2012) Active transcription and essential role of RNA polymerase II at the centromere during mitosis. *Proc Natl Acad Sci USA* **109**, 1979–1984.
- 145 Kabeche L, Nguyen HD, Buisson R & Zou L (2018) A mitosis-specific and R loop–driven ATR pathway promotes faithful chromosome segregation. *Science* **359**, 108–114.
- 146 Ten Hagen KG, Gilbert DM, Willard HF & Cohen SN (1990) Replication timing of DNA sequences associated with human centromeres and telomeres. *Mol Cell Biol* **10**, 6348–6355.
- 147 Erliandri I, Fu H, Nakano M, Kim J-H, Miga KH, Liskovych M, Earnshaw WC, Masumoto H, Kouprina N, Aladjem MI *et al.* (2014) Replication of alpha-satellite DNA arrays in endogenous human centromeric regions and in human artificial chromosome. *Nucleic Acids Res* **42**, 11502–11516.
- 148 Jaco I, Canela A, Vera E & Blasco MA (2008) Centromere mitotic recombination in mammalian cells. *J Cell Biol* **181**, 885–892.
- 149 Giunta S & Funabiki H (2017) Integrity of the human centromere DNA repeats is protected by CENP-A, CENP-C, and CENP-T. *Proc Natl Acad Sci USA* **114**, 1928–1933.
- 150 Black EM & Giunta S (2018) Repetitive fragile sites: centromere satellite DNA as a source of genome instability in human diseases. *Genes* **9**, 615.
- 151 Barra V & Fachinetti D (2018) The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. *Nat Commun* **9**, 4340.
- 152 Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, Schumacher SE, Wang C, Hu H, Liu J *et al.* (2018) Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689.e3.
- 153 Hermsen MAJA, Joenje H, Arwert F, Welters MJP, Braakhuis BJM, Bagnay M, Westerveld A & Slater R (1996) Centromeric breakage as a major cause of cytogenetic abnormalities in oral squamous cell carcinoma. *Genes, Chromosomes Cancer* **15**, 1–9.
- 154 Martins C, Jin Y, Jin C, Wennerberg J, Höglund M & Mertens F (1999) Fluorescent in situ hybridisation (FISH) characterisation of pericentromeric breakpoints on chromosome 5 in head and neck squamous cell carcinomas. *Eur J Cancer* **35**, 498–501.
- 155 Wang JC, Hajianpour A & Habibian R (2009) Centromeric alpha-satellite DNA break in reciprocal translocations. *Cytogenet Genome Res* **125**, 329–333.
- 156 Martínez-A C & van Wely KHM (2011) Centromere fission, not telomere erosion, triggers chromosomal instability in human carcinomas. *Carcinogenesis* **32**, 796–803.
- 157 Martínez JG, Pérez-Escuredo J, Llorente JL, Suárez C & Hermsen MA (2012) Localization of centromeric breaks in head and neck squamous cell carcinoma. *Cancer Genet* **205**, 622–629.
- 158 Barbouti A, Stankiewicz P, Nusbaum C, Cuomo C, Cook A, Höglund M, Johansson B, Hagemeijer A, Park S-S, Mitelman F *et al.* (2004) The breakpoint region of the most common isochromosome, i(17q), in human neoplasia is characterized by a complex genomic architecture with large, palindromic, low-copy repeats. *Am J Hum Genet* **74**, 1–10.
- 159 Knutsen T, Padilla-Nash HM, Wangsa D, Barenboim-Stapleton L, Camps J, McNeil N, Difilippantonio MJ & Ried T (2010) Definitive molecular cytogenetic characterization of 15 colorectal cancer cell lines. *Genes Chromosomes Cancer* **49**, 204–223.
- 160 Iacovoni JS, Caron P, Lassadi I, Nicolas E, Massip L, Trouche D & Legube G (2010) High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. *EMBO J* **29**, 1446–1457.
- 161 Szilard RK, Jacques P-É, Laramée L, Cheng B, Galicia S, Bataille AR, Yeung M, Mendez M, Bergeron M, Robert F *et al.* (2010) Systematic identification of fragile sites via genome-wide location analysis of γ -H2AX. *Nat Struct Mol Biol* **17**, 299–305.
- 162 Rogakou EP, Boon C, Redon C & Bonner WM (1999) Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J Cell Biol* **146**, 905–916.
- 163 Seo J, Kim K, Chang D-Y, Kang H-B, Shin E-C, Kwon J & Choi JK (2014) Genome-wide reorganization of histone H2AX toward particular fragile sites on cell activation. *Nucleic Acids Res* **42**, 1016–1025.
- 164 Savic V, Yin B, Maas NL, Bredemeyer AL, Carpenter AC, Helmink BA, Yang-Iott KS, Sleekman BP & Bassing CH (2009) Formation of dynamic gamma-H2AX domains along broken DNA strands is distinctly regulated by ATM and MDC1 and dependent upon H2AX densities in chromatin. *Mol Cell* **34**, 298–310.
- 165 Simsek D & Jasin M (2010) Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat Struct Mol Biol* **17**, 410–416.
- 166 Ghezraoui H, Piganeau M, Renouf B, Renaud JB, Sallmyr A, Ruis B, Oh S, Tomkinson AE, Hendrickson EA, Giovannangeli C *et al.* (2014) Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Mol Cell* **55**, 829–842.
- 167 Aymard F, Bugler B, Schmidt CK, Guillou E, Caron P, Briois S, Iacovoni JS, Daburon V, Miller KM,

- Jackson SP *et al.* (2014) Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol* **21**, 366–374.
- 168 Clouaire T & Legube G (2015) DNA double strand break repair pathway choice: a chromatin based decision? *Nucleus* **6**, 107–113.
- 169 Kantor GJ, Barsalou LS & Hanawalt PC (1990) Selective repair of specific chromatin domains in UV-irradiated cells from xeroderma pigmentosum complementation group C. *Mutat Res* **235**, 171–180.
- 170 Hanawalt PC & Spivak G (2008) Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* **9**, 958–970.