



Review

# Common Chromosomal Fragile Sites—Conserved Failure Stories

Vasileios Voutsinos , Sebastian H. N. Munk and Vibe H. Oestergaard \* 

Department of Biology, University of Copenhagen, 2200 Copenhagen N, Denmark;  
vasileios.voutsilos@bio.ku.dk (V.V.); tzk640@alumni.ku.dk (S.H.N.M.)

\* Correspondence: vibe@bio.ku.dk; Tel.: +45-35-330444

Received: 31 October 2018; Accepted: 21 November 2018; Published: 27 November 2018

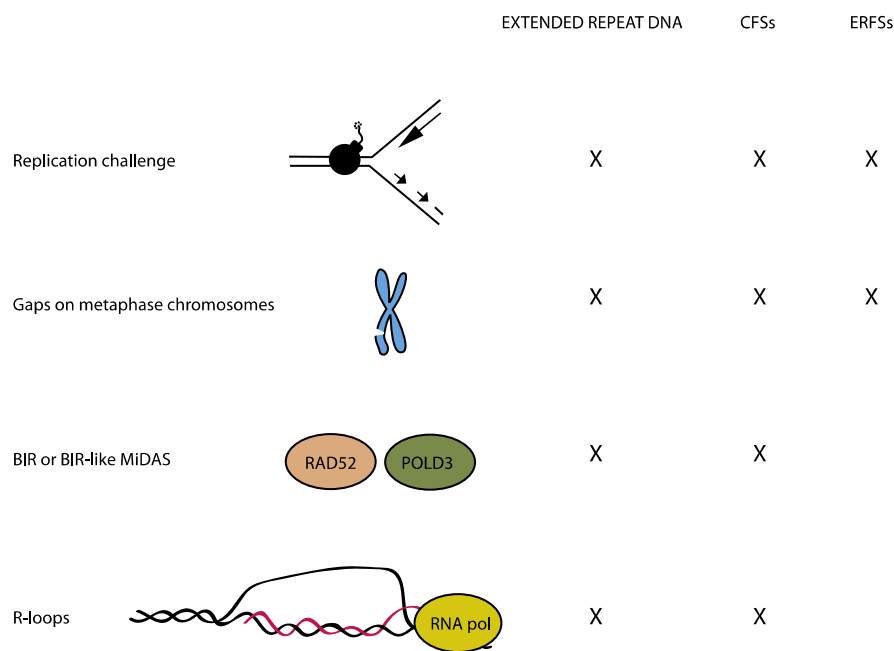


**Abstract:** In order to pass on an intact copy of the genome during cell division, complete and faithful DNA replication is crucial. Yet, certain areas of the genome are intrinsically challenging to replicate, which manifests as high local mutation propensity. Such regions include trinucleotide repeat sequences, common chromosomal fragile sites (CFSs), and early replicating fragile sites (ERFSs). Despite their genomic instability CFSs are conserved, suggesting that they have a biological function. To shed light on the potential function of CFSs, this review summarizes the similarities and differences of the regions that challenge DNA replication with main focus on CFSs. Moreover, we review the mechanisms that operate when CFSs fail to complete replication before entry into mitosis. Finally, evolutionary perspectives and potential physiological roles of CFSs are discussed with emphasis on their potential role in neurogenesis.

**Keywords:** common chromosomal fragile sites; large genes; introns; mitosis; early replicating fragile sites; neuronal diversification

## 1. Introduction

Regions that are difficult to replicate are at high risk of failing replication before cells enter mitosis, and this gives rise to microscopically visible gaps on metaphase chromosomes (Figure 1). Such regions, which are incompletely replicated in mitosis, are referred to as underreplicated. Several causes of local replication difficulties have been identified, including tandem repetitive DNA sequences, transcription activities, and timing of replication. Trinucleotide repeat expansions have been extensively characterized due to their involvement in diseases [1]. The challenge they pose to replication is intrinsic to the DNA sequence. However, studies of trinucleotide repeat expansion led to the discovery of another type of regions that are difficult to replicate due to certain transcription and replication patterns. These regions were dubbed common chromosomal fragile sites (CFSs) [2]. They are late-replicating regions with large transcribed genes and they are fragile in all individuals. In this review, the large genes at CFSs are referred to as CFS genes. More recently, another type of fragile sites was discovered. These were called early replicating fragile sites (ERFSs), because they are replicated early in S phase and moreover, they are highly transcribed regions [3].



**Figure 1.** Similarities of three types of regions causing replication problems. Schematic summarizing similarities of expanded short tandem repeat sequence, common chromosomal fragile sites (CFSSs), and early replicating fragile sites (ERFSs). See text for further details.

## 2. Short Tandem DNA Repeats

Some types of repetitive DNA sequences are comprised by short stretches of DNA that are repeated in tandem with a variable number of repeat units. More than a million short tandem DNA repeats (minisatellites and microsatellites, which includes trinucleotide repeats) are found in the human genome [4], and interestingly, many of these seem to be involved in the regulation of gene expression [5,6]. Short tandem DNA repeats have a tendency to diversify, which has been suggested as a means of genetic diversification with consequences for gene expression levels [7]. While the proposed physiological role of inherently unstable short tandem DNA repeats is relatively recent, pathological consequences of trinucleotide repeat expansions have been known for many years. Due to their involvement in diseases, trinucleotide DNA repeats have been widely studied and it is clear that this type of repeat DNA causes replication fork stalling and DNA damage [8]. Because short tandem DNA repeats cause replication problems, they are highly genetically unstable from generation to generation with extensive changes in copy numbers between individuals [9]. Therefore, short tandem DNA repeats are extensively used for forensic purposes [10]. Short tandem DNA repeats are also evolutionarily unstable as expected, and even at the somatic level, they vary in copy number [7,11].

The tendency of trinucleotide DNA repeats to increase in copy number is the underlying cause of a number of genetic diseases, collectively referred to as trinucleotide repeat expansion syndromes. These include Fragile X syndrome, Friedreich's Ataxia, Huntington Disease, and many others, most of which affect the nervous system [7,12]. Fragile X syndrome was one of the first diseases found to result from repeat expansion on chromosome X [13,14], which causes the silencing of the gene *FMR1*. Moreover, alleles carrying the disease-causing repeat expansion at the *FMR1* locus were found to replicate later during the cell cycle [15]. Early studies of Fragile X revealed a correlation between Fragile X patients and microscopically visible gaps on the X chromosome in metaphase chromosome spreads from patient lymphoblasts, thus providing a cytogenetic feature that is useful for diagnosis and identification of carriers of Fragile X. However, conclusions that are based on cytogenetic analyses were initially uncertain because only a small fraction of the X chromosomes from male patients appeared as broken.

Researchers thus explored culture conditions that would further induce gaps at the X chromosomes [16]. They found that when replication was challenged in cultured lymphoblasts, the fragility on the X chromosome became more evident in Fragile X patients. Taken together, these early studies hinted that replication difficulties at repeat expansions are a key problem. Indeed, many studies have confirmed that replication fork stalling frequently occurs at repeat DNA [1,17,18]. Short tandem repetitive DNA sequences, many of which can form secondary structures, such as hairpins and G-quadruplexes [19], may comprise obstacles to the replication machinery. At the same time, they have a tendency to form gaps on metaphase chromosomes in response to replication stress, which is why extensive trinucleotide repeat expansions give rise to so-called rare fragile sites that are occasionally associated with disease.

### 3. Common Chromosomal Fragile Sites

During the above-mentioned search for culture conditions that promoted Fragile X chromosome breakage, another important observation was made: lymphoblasts from healthy individuals also displayed recurrent chromosome gaps on metaphase chromosomes when they were cultured with low doses of the replication inhibitor aphidicolin (APH) (Figure 1) [20,21]. Regions that are susceptible to form gaps were named CFSs.

Years later, interest in CFSs was renewed when advanced array and sequencing techniques revealed that CFSs were highly mutated in a broad range of cancer genomes [22–24]. While cancer driver genes show a distinct pattern of enrichment for homozygous deletion, seminal work from Bignell et al. revealed that mutation patterns at CFSs were strikingly enriched for both hemizygous and homozygous deletions (with an average size of around 500 kb) across cancer genomes [22]. This reflects that whilst tumor suppressors are generally recessive, meaning that there is a selection for homozygous deletion during carcinogenesis, CFS deletions are bystander effects that are caused by their high mutation rates during conditions of replication stress. A caveat of this view is that haploinsufficient tumor suppressors may show enrichment for hemizygous deletions similar to fragile sites. Indeed, there are observations supporting that the fragile site gene *WWOX* is a haploinsufficient tumor suppressor [25]. Though certain CFSs seem to hold genes that may be tumor suppressors, it is still debated whether mutations at CFSs in general contribute to tumorigenesis [25].

Due to the shared features of CFSs with trinucleotide repeats (Figure 1), initial studies were focused on identifying sequence characteristics that would render CFSs prone to replication problems. This led to the finding that some CFSs are AT rich [26,27] and indeed the AT-rich regions of certain CFSs are difficult to replicate and they constitute hotspots for replication fork stalling [27–29]. However, the repertoire of CFSs is cell type-specific and varies between tissues showing that the fragility is not fully intrinsic to the DNA sequence [30–32]. Rather, CFSs coincide with large genes, the transcription of which is tightly linked to CFS breakage and their tendency to acquire deletions [33–35]. This is in agreement with the CFS characteristics that are derived from studies of cancer genomes [22–24].

### 4. Early Replicating Fragile Sites

Hydroxyurea (HU) is an inhibitor of the ribonucleotide reductase, thus HU treatment leads to the depletion of dNTPs. High doses of HU were found to result in DNA breaks at specific sites in the genome, called ERFs [3]. As the name implies, ERFs are located at early replicating regions of the genome. However, upon HU treatment followed by release into normal medium, the replication problem that is caused by HU persists throughout interphase and lead to the formation of cytogenetic gaps at ERFs on metaphase chromosomes (Figure 1). ERFs contain highly transcribed genes, the transcription of which contributes to their replication difficulties [3]. Recent work suggests that poly(dA:dT) tracts at ERFs are responsible for replication initiation, yet they can also trigger replication fork collapse [36]. This was suggested to be due to the inefficient binding of RPA to poly(dA) sequence upon unwinding by the replicative helicase, leaving this uncoated strand vulnerable to nuclease attacks [36]. In avian DT40 cells, ERF-like regions were identified based on their enrichment for the protein FANCD2, when cells were subjected to replication stress by APH treatment [33]. Similar to the

previously described ERFs, ERF-like regions in DT40 contained dense clusters of highly transcribed genes that were replicated early in the S phase [33]. FANCD2 is a protein that is involved in the replication of CFSs as described below [28].

## 5. Transcription-Replication Conflicts

The transcription machinery may hinder the progression of the replication machinery since they work on the same template. Specifically, so-called R loops (Figure 1), where nascent RNA hybridizes with the DNA template and displaces the non-template strand can cause replication fork stalling and DNA double strand breaks (DSBs) [37–39]. Two types of transcription-replication conflicts (TRCs) can occur, namely co-directional (the two processes travel in same direction) and head-on (the two processes have opposite directionality). Both types cause replication problems but head-on collisions are more deleterious [40] and the cellular response is also different [38]. Using an episomal plasmid with a strong R-loop seeding sequence, it was found that head-on collisions exacerbate R-loop formation and trigger the ATR kinase that is normally associated with the replication stress response, whereas co-directional collisions alleviate R-loop formation and activates the ATM kinase known for its role in response to DSB [38]. Whilst TRCs that are caused by R loops are likely to be involved in CFS breakage in mitosis [28,41,42], it is more uncertain whether TRCs are direct causes of replication problems at ERFs and expanded nucleotide repeats.

Transcription at ERFs is certainly contributing to their fragility [3], suggesting that TRCs directly induce replication problems at ERFs. Alternatively, transcription at ERFs indirectly influences the coinciding fragility since transcription in G1 phase shapes the replication program in the following S phase [43]. Specifically, transcription activity in G1 clears transcribed regions of origin complexes from which replication is initiated [36,44]. Thus, the transcription profile of a cell is very important for determining the timing of replication. Likewise, the duration of G1 is crucial, because if G1 is shortened, full clearance of replication origins from long intragenic regions may not be accomplished, which may trigger TRCs during the S phase [43].

R loops often form at expanded trinucleotide repeats due to secondary structure formation on one strand, leaving the other strand susceptible to annealing with nascent RNA. Such R loops at the trinucleotide repeats contribute to repeat instability [45,46]. One well-characterized effect of R loops on repeat stability involves single-strand break formation, followed by base excision repair [45]. Another recent study finds that R loops at a trinucleotide repeat triggers so-called break-induced replication (BIR), and this mechanism underlies repeat expansion [47].

## 6. Transcription Timing of Common Chromosomal Fragile Sites

Transcription activity is evidently regulated to ensure that transcription and replication are kept separate [48,49]. In *Drosophila melanogaster*, active genes are replicated predominantly in early S phase, indicating a regulation of transcription timing in higher eukaryotes [50]. Similarly, one study found a spatial separation of replication from transcription in mouse cells [49], while another study established a temporal separation of the two cellular processes [48]. In fact, by utilizing a nascent RNA capture assay to map the transcriptional events throughout the S phase, transcription activity at a given time and place was found to inversely correlate to those of replication [48]. The spatiotemporal separation between replication and transcription serves to reduce conflicts between the two machineries. In the case of CFS genes, however, it is argued that TRCs are inevitable, since their transcription can span more than one cell cycle [41]. By performing quantitative reverse transcriptase PCR (RT-qPCR) on introns of CFS genes Helmrich et al. characterized the nascent expression of those genes in relation to the cell cycle phase. Surprisingly, they found that the transcription of at least some of the CFS genes start at the G2 or M phase, continues throughout the next cell cycle and it is completed in the G1 or early S phase in the cell cycle after that [41]. Such a transcription pattern makes TRCs inevitable, since the transcription of the genes will definitely coincide with the replication of the CFSs [41]. Replication of CFSs takes place late in S phase and replication stress can further delay replication with the result

that CFSs go into G2 and M phase without being fully replicated [33,51,52]. The replication challenges are also caused by a scarcity of origins at CFSs [53]. After the induction of mild replication stress, increased rates of chromosome breakage are seen in CFS genes that are actively transcribed [41]. Interestingly, the hotspots for breakage are located at the areas that are transcribed during the S phase of the cell, which again indicates that TRCs are responsible for the CFS fragility [41]. R loops are also elevated at CFSs after APH treatment, which may correlate with increased incidence of TRCs in the CFS genomic region [41]. It is thus apparent that the transcription of CFS genes very often collides with the replication of the CFSs, which creates genomic instability.

## 7. Early Replicating Fragile Sites and Common Chromosomal Fragile Sites Comprise Underreplicated Regions in Mitosis

Collectively, it seems that transcription at ERFs and CFSs, either directly or indirectly, increases the risk of cells entering mitosis prematurely before replication has been completed at these sites [3,33,34,42]. Yet, it is unclear whether the type of replication impediments that occurs at ERFs or CFSs are intrinsically prone to escape checkpoint detection. Recent work from Lemmens et al. suggests that the replication process itself is inhibiting the activation of the mitotic kinases PLK1 and CDK1 via CHK1 [54]. However, a few ongoing replication forks may not be sufficient to repress PLK1 and CDK1. Also, work from the Cimprich lab shows that ATR activity relays ongoing replication to suppress the transcription of a set of mitosis-promoting genes in the primary RPE1 cell line [55]. This is in agreement with previous studies in cancer cell lines, which also found that ATR is key for suppressing mitosis until replication is complete [56,57]. Underreplicated genomic regions per se are not sufficient to suppress mitotic entry [58]. Rather, it seems that the active replication machinery signals via ATR and CHK1 to ensure the complete replication before mitotic onset. Consistently, both ERFs and CFS stability is dependent on ATR [3,59], but it remains unclear whether their tendency to escape checkpoint activation is linked to their transcription or other specific features.

## 8. The Consequences of Underreplicated DNA

### 8.1. FANCD2 and Anaphase Bridges

Research from the last decade has given important insight into the cell biology of CFSs by deciphering cellular consequences of replication problems at CFSs. It is now more clear how underreplicated CFSs manifest and how they are processed in mitosis and beyond. The DNA damage bypass mechanism called translesion synthesis may have a role in completing replication of CFSs [60,61]. This backup pathway probably works in G2 before mitosis [62]. If the replication problem at a CFS persists into mitosis, it can be visualized as so-called FANCD2 sister foci at the CFS on each sister chromatid [63,64]. FANCD2 is a key component of the Fanconi Anemia repair pathway, which provides cellular resistance to DNA cross-linking agents [65]. Similar to cross link-induced FANCD2 foci, FANCD2 foci at CFSs on metaphase chromosomes colocalize with the FANCD2 binding partner FANCI and they are dependent on FANCD2 monoubiquitylation by the Fanconi Anemia core complex [63,64].

FANCD2 foci in the M phase are thought to mark regions that have not been fully replicated before cells enter mitosis and chromatin immunoprecipitation sequencing (ChIP-seq) of FANCD2 in APH-treated cells consistently identifies CFSs [33,66]. In agreement with the view that gaps on metaphase chromosomes are the consequence of incomplete replication, G2-M checkpoint deficiency provokes and exacerbates replication stress-induced gaps at CFSs on metaphase chromosomes. Specifically, ATR, CHK1, and BRCA1 via their involvement in the G2-M checkpoint are important to prevent CFS breakage [59,67,68]. Notably, FANCD2 foci were observed both at broken and unbroken CFSs on chromosomes in metaphase. Upon progression into anaphase, FANCD2 foci segregate symmetrically on the sister chromosomes [63,64]. Interestingly, so-called ultrafine anaphase bridges (UFBs) often form between FANCD2 sister foci on anaphase chromosomes. UFBs are threads of DAPI-negative DNA interconnecting sister centromeres during normal anaphase [69], whereas



replication stress specifically induces UFBs at CFSs [63,64]. The proteins PICH and BLM coat all types of UFBs [69], whereas FANCD2 defines the UFBs from CFSs and most often localizes to the termini of UFBs [63,64]. The recruitment of BLM to replication stress-induced UFBs is compromised in FANCA and FANCC deficient cell lines, whereas PICH recruitment is not affected by FANCD2 deficiency [64,70].

The recruitment mechanism as well as the exact role of FANCD2 at CFSs in mitosis is somewhat obscure. FANCD2 does not seem to be required for recruitment of endonucleases (described below) to underreplicated CFSs in mitosis [70]. However, recent work from Madireddy et al. shows that FANCD2 is required for replication through an AT-rich region in the core of the most fragile region of FRA16D, which houses the *WWOX* gene [28]. Moreover, it was found that FANCD2 deficiency results in a marked increase in RNA-DNA hybrid formation at FRA16D and that the overexpression of RNaseH1 (an enzyme that is capable of degrading RNA in RNA-DNA hybrids) restores replication progression through FRA16D, collectively suggesting that FANCD2 works to facilitate replication through RNA-DNA hybrids at CFSs [28]. FANCD2 has a well-characterized role in inter-strand cross-link repair, which depends on its monoubiquitylation. In contrast, the role of FANCD2 in the replication of CFSs is not fully dependent on its monoubiquitylation [28]. Interestingly, BRCA2/FANCD1, which normally works downstream of FANCD2 monoubiquitylation, has a role in the replication of FRA16D that is similar to FANCD2 [28].

### *8.2. Endonucleases Promote Common Chromosomal Fragile Sites Breakage and DNA Synthesis in Mitosis*

A number of structure selective endonucleases, including MUS81 and its binding partners EME1 and XPF-ERCC1, as well as the nuclease scaffold SLX4, localize to CFSs in mitosis where they are squeezed in between sister FANCD2 foci [70,71]. Here, the nuclease activity of MUS81 is promoting DNA synthesis in mitosis (referred to as MiDAS [72]) [73]. MUS81 activity is thought to create DNA strand breaks that are used to initiate DNA synthesis in a process that requires RAD52 but not RAD51 [72]. Similar to the previously characterized DNA synthesis pathway named BIR [74,75], DNA synthesis in mitosis depends on the non-catalytic subunit of polymerase delta called POLD3 (Figure 1) [73]. The cleavage activity of MUS81 is stimulated by the helicase RECQL5 and it was suggested that RECQL5 removes RAD51 from stalled forks, making them accessible to MUS81. Interestingly, this function of RECQL5 is dependent on Ser727 phosphorylation (in human RECQL5) by CDK1 [76]. CDK-mediated phosphorylation also promotes the complex formation of SLX4-SLX1 with MUS81-EME1 at mitotic onset [77]. Thus, the microscopically visible gaps on human metaphase chromosomes are actively generated by a process that requires MUS81 [70,71], and some of the gaps probably represent newly synthesized DNA at the mitotic chromosomes [70,71,73]. Interestingly, the recruitment of SLX4 to mitotic chromosomes depends on TopBP1 in avian DT40 cells [78]. Consistently, TopBP1 is found at APH-induced gaps on metaphase chromosomes and certain UFBs. TopBP1 also facilitates DNA synthesis in mitosis [78,79]. TopBP1 interacts with SLX4 in a manner that depends on Thr1260 (in human SLX4). This residue lies within a conserved motif, where the homologous motif in yeast SLX4 was shown to undergo CDK1-dependent phosphorylation, which stimulated the interaction between the yeast TopBP1 homolog Dpb11 and SLX4 specifically in mitosis [80].

Preliminary work that was published on BioRxiv describes a process that may work as an alternative to MiDAS for processing of underreplicated regions in mitosis [81]. This work suggests that the E3 ubiquitin ligase TRAIIP mediates the unloading of the replicative helicase when cells enter mitosis and this unloading promotes error-prone repair of underreplicated regions by either RAD52- or Polθ-mediated annealing [81]. This model is consistent with the frequent deletions and sister chromatid exchanges that were observed at CFSs [22,82].

### *8.3. Consequences of Underreplicated Regions in G1 Daughter Cells*

Underreplicated regions that persist in anaphase can lead to non-disjunction manifesting as either UFBs or chromatin bridges. UFBs may still be resolved in anaphase or telophase but they can leave a scar marked by large 53BP1 foci in subsequent G1 [83]. These are named 53BP1 nuclear bodies

(53BP1 NBs) [83,84]. The 53BP1 NBs in G1 colocalize with  $\gamma$ H2AX suggesting that they mark DNA damage. When cells enter S phase, 53BP1 NBs dissolve. This may reflect that they require homologous recombination-mediated repair rather than non-homologous endjoining (NHEJ), which is the favored pathway for repair in G1 [85]. The presence of 53BP1 NBs in G1 cells leads to extension of the G1 phase or entry in to a quiescent state in a manner that depends on p53 [86–88]. 53BP1 NBs also colocalize with so-called OPT-domains that contain the transcription factors Oct-1 and PTF, as well as the RNA helicase DDX1 [84], but transcription is repressed in 53BP1 NBs [84]. Importantly, other types of DNA stress that do not manifest as DNA bridges contribute substantially to 53BP1 NBs [78]. Thus, TopBP1 foci that persist throughout mitosis colocalize with 53BP1 NBs in G1, showing that TopBP1 constitutes a useful marker for DNA structures that precipitate DNA damage in daughter cells [62,78]. Deficiency in mitotic DNA processing also leads to missegregation, which in turn can cause binucleation [78] or the formation of micronuclei [63,64]. Micronuclei can trigger the highly mutagenic event chromothripsis [89].

## 9. Conserved Features of Common Chromosomal Fragile Sites Genes

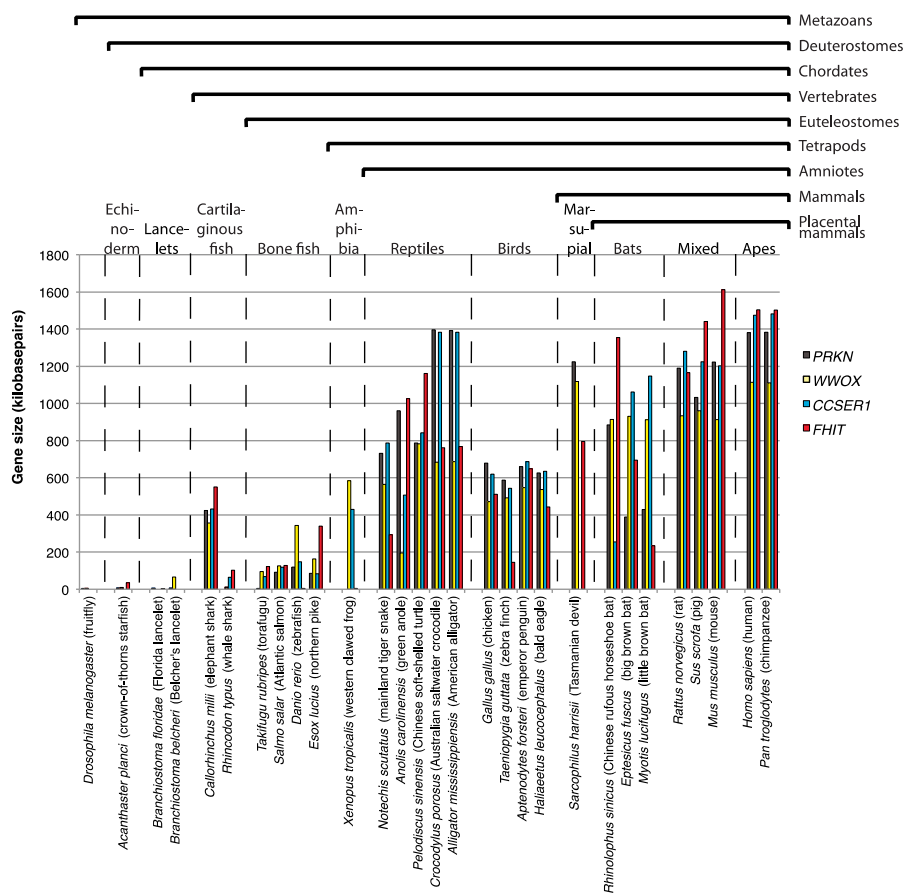
Despite being highly unstable genomic regions, increasing evidence suggests that CFSs may hold conserved biological functions. Molecular mapping revealed that CFSs are conserved between mammals and birds and coincide with orthologous genes [33] and comparison of CFS fragility levels have shown that CFS fragility is conserved between mouse and human in homologous regions in lymphocytes [35]. Genes hosting CFSs are extremely large and they consist mainly of very long intronic sequences, which have persisted during evolution despite rendering the genes susceptible to genomic instability [33]. Analysis of gene lengths in 203 vertebrates revealed that large gene bodies of extremely long genes are conserved between most vertebrates [33]. Furthermore, while a global tendency of intron size reduction is observed in bird genomes, large introns in all CFS genes investigated (*PRKN*, *MARCOD2*, *GRID1*, *CCSER1*, *SETBP1*, *IMMP2L*, and *DACH1*) have to some extent resisted size reduction [33], supporting that large introns in CFS genes may hold a conserved biological function.

The role of introns in creating diverse and complex transcriptomes through alternative splicing is well established. In addition, a recent study by Bonnet et al. proposes a role of introns in preventing R-loop accumulation and the associated genomic instability [90]. Genome-wide mapping of R loops and DNA-damage histone markers in budding yeast and human cell lines showed that among highly transcribed genes, genes without introns are more prone to form R loops and accumulate DNA damage than intron-containing genes. In budding yeast, intron deletion from highly transcribed genes led to de novo formation of R loops at the loci. Intriguingly, spliceosome assembly on intron-containing precursor messenger RNA (pre-mRNA) suppressed R-loop formation while insertion of a group I self-splicing intron failed to do so, suggesting that the recruitment of splicing factors to small introns prevent the formation of R loops rather than splicing per se. In contrast to this proposed role of small introns in budding yeast, it is evident that the extremely long introns in CFS genes promote conflicts between transcription and replication processes, including R loops rather than protect against them [28,41], which may be due to differences in the protein composition of small versus long introns in nascent pre-mRNA.

Altogether, while this recent study proposes a role of small introns in maintaining genome integrity of transcribed loci, the role of long introns in CFS genes seems to be associated with the coinciding fragility. In line with this, the extreme length of fragile site genes may generate extensive topological problems because negative supercoils build up in the DNA behind RNA pol II [91]. This phenomenon is thought to underlie R-loop formation at long genes, which occurs in response to the depletion of Top1, the topoisomerase responsible for relief of topological stress in DNA [92].

The fragility of CFSs may carry a functional significance [93]. As previously suggested, DNA breaks at CFSs may serve to activate the DNA damage checkpoint and induce apoptosis or senescence, thus constituting a barrier to transformation, as proposed for oncogene-induced replication stress [93,94].

To understand at which point during the evolution the vast size of CFS genes arose, we performed a simple analysis of CFS orthologue length in their phylogenetic context (Figure 2). We here focused on four well-studied CFS genes, *PRKN*, *CCSER1*, *WVOX*, and *FHIT*, where two of these, *PRKN* and *CCSER1*, were confirmed as conserved CFSs by their fragility in chicken DT40 cells [33]. When total gene length is plotted in selected grouped animals it is clear that the CFS genes under investigation started their expansion in an early vertebrate ancestor (Figure 2). Curiously, one of the represented cartilaginous fish, the elephant shark has larger CFS genes than any of the bone fish represented. Unfortunately, the *PRKN*, *CCSER1*, *WVOX*, and *FHIT* are poorly annotated in amphibian but based on the sizes of *WVOX* and *CCSER1*, which are represented it seems that they gain extra length in amphibians and certainly in reptiles, where the crocodile and alligator genomes present with *PRKN* and *CCSER1* genes that exceed 1.3 megabases in length. Birds have undergone a reduction in total genome size during evolution, which reflects an adaption to flight [95]. Thus, in general introns as well as intergenic regions are reduced in size in birds. However, our recent work showed that *PRKN* and *CCSER1* to some degree have escaped the reduction in intron sizes, suggesting that the extreme size of the genes may provide fitness to the organisms [33]. Mammals in general seem to have retained the extreme size of CFS genes although there are some exceptions in bats. Similar to birds, bat genomes have been reduced during evolution, which similar to birds is thought to be an adaption to flight. Thus, bats have the most compact of mammalian genomes. The tendency for reduced size is also reflected in the slightly smaller size of the CFS genes in bats, although it may not be proportional to the overall degree of genome reduction. Finally, we note that some of the inconsistencies within groups are likely due to annotation errors.



**Figure 2.** Bar chart showing gene sizes of the CFS genes *PRKN*, *WVOX*, *CCSER1*, and *FHIT* in various metazoan species with focus on vertebrates. Latin name and common name is given below the bars. Phylogenetic classification is indicated above the chart. Gene sizes are retrieved from the National Center for Biotechnology Information (NCBI) genes database.



## 10. Neuronal Genetic Diversity Caused by Common Chromosomal Fragile Sites

Since CFSs seem to be detrimental, or at best neutral for the cells, why are they retained throughout evolution? A proposed explanation suggests that CFSs may have a role in producing the immense diversity [96,97] that has been observed in the brain cells [98,99]. CFSs are particularly fragile when subjected to replication stress and this can lead to genomic structural variations [34,100,101], which in turn can be part of the underlying cause for the diversity of the neurons. In agreement with this hypothesis, single cell sequencing of the human frontal cortex has shown that 13–41% of its neurons contain at least one mega-base scale copy number variation [101]. Moreover, the condition of replication stress for structural variations to arise is met as neural stem cells undergo massive rapid proliferation during neurogenesis generating replication stress [102]. This is also evident by the fact that DNA repair factors that are implicated in replication stress-associated damage are critical for proper neurodevelopment. One example of this is the ATR-Seckel syndrome that is a result of ATR deficiency. The ATR-Seckel syndrome is largely characterized by defects in the nervous system [103]. TopBP1 is another DNA repair factor deficiency of which results in the ablation of the structure of the cortex in a p53 dependent manner [104]. It has also been shown that neural proliferation continues in adult brains as well, particularly in the subventricular zone and subgranular zone [105,106]. In fact, approximately 700 new neurons per day are generated in the hippocampus. This corresponds to 1.75% annual turnover of the neurons in the renewing fraction [107]. Such proliferation implies that replication stress could still occur in the adult brain and its consequences could be aggravated given that the availability of dormant replication origins on the genome diminish in neural stem progenitor cells (NSPCs) as compared to embryonic stem cells (ESCs) [108].

In a recent study by Wei et al. recurrent DSB clusters (RDCs) occurring in NSPCs of mice were identified using high throughput genome-wide translocation sequencing (HTGTS) [96]. Using stringent criteria they found 27 RDCs. These were mostly located on transcriptionally active, large (>100 kb), late-replicating genes, which are all characteristics that are associated with CFSs and six out of the 27 RDCs identified in the study correspond to known CFSs in the human genome [96]. Importantly, most of those RDCs could only be detected after induction of mild replication stress with APH. Remarkably, the RDCs are all associated with genes that are implicated in neural processes and disorders [96]. Wei et al. propose that DSBs may at least partially explain the CNVs that were observed in the human brain. Thus, CFSs may contribute to the genetic diversity of the neurons.

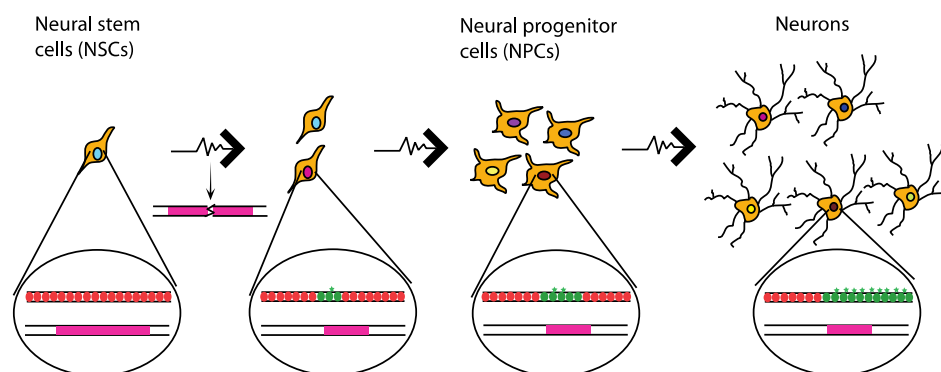
## 11. A potential Link between Common Chromosomal Fragile Sites and LINE1-Dependent Neuronal Genetic Diversification

CFSs might also have a supportive role in retrotransposon LINE1-dependent diversification of the neuronal genomes [109,110]. LINE1 retrotransposons have been shown to influence neuronal cell differentiation/maturation in cell culture and cause neuronal mosaicism in the brain of mice [111]. Furthermore, some LINE1 retrotransposons can be expressed as a result of the Wnt signaling pathway in the same way that the proneurogenic factor NEUROD1 is expressed, indicating a possible role of LINE1 in neurogenesis [112]. Although, LINE1 seems to have an important role in neurogenesis, its integration process can induce genomic instability [113]. It was shown that LINE1 expression could cause a large number of DSBs, as indicated by increased  $\gamma$ -H2AX foci formation. The number of DSBs is greater than the predicted number of successful integrations of LINE1, indicating that the LINE1 integration process is largely inefficient [114]. According to a recent study, replication stress that results from expression of long neural genes renders glioblastoma cancer stem-like cells radiation resistant [115]. The group proposes that this is because of the constitutive activation of the DNA damage response, as a result of the replication stress at the long neural genes. When active, the DNA damage response can in turn facilitate DSB repair that is caused by radiation, thus conferring resistance to radiation [115]. Following this line of logic, it is conceivable that CFSs hosting long neural genes and the replication stress that they generate when transcribed act as a minor challenge that keeps the neural cell prepared for the major physiological challenge that is imposed by LINE1 retro

transposition. CFS rearrangements may distinctly promote neuronal diversification through both direct and indirect mechanisms.

## 12. Potential Neuronal Epigenetic Diversity Caused by Common Chromosomal Fragile Sites

Genetic diversity is not the only potential explanation of the mosaicism observed in the brain. Such diversity could also stem from epigenetic diversification of the neurons. In particular, replication stress in primary fibroblasts can cause the accumulation of the histone variant macroH2A1.2 at CFSs in a FACT-dependent manner [116]. MacroH2A1.2 facilitates homologous recombination-mediated DNA repair by recruiting the recombination factor BRCA1 [116] and it has been associated with H3K27me<sub>3</sub>-containing chromatin and senescence-associated heterochromatin [117,118]. Similarly, Papadopoulou et al. found replication stress-induced changes of gene expression due to epigenetic changes in a region of the genome that forms G4 quadruplex structures, which complicates replication of the region [119]. In fact, expression levels of the BU-1 gene, which contains a G4 quadruplex structure, are reduced when replication stress is induced by APH or HU treatment. The reduction in expression is interpreted as a consequence of the lack of coordination of replication and histone deposition processes, due to the delayed replication that is caused by the G4 structure and aggravated by the aforementioned drug treatments [119]. It is therefore tempting to speculate that the gradual accumulation of epigenetic changes in CFSs owed to replication stress [116] can cause alterations in the expression of the CFS genes, which are largely associated with neural related functions [96,120] and therefore lead to cell diversification of the neurons (Figure 3).



**Figure 3.** Hypothetical model showing diversification of neurons as a result of replication stress at CFSs. As neural stem cells (NSCs) proliferate, replication stress (spiked arrow) can result in loss of coordination of the histone deposition and replication process. As a consequence, histones that carry modifications (red circles) are replaced by newly synthesized unmodified histones (green circles). Replication stress at CFSs also causes accumulation of the histone variant macroH2A1.2 (green star). At the genetic level structural variations, like deletions, can arise due to replication stress-induced breaks. The variance at the genetic and the epigenetic level that is caused by replication stress throughout the differentiation of neurons can lead to variance of the expression of the genes located at CFSs. Each individual cell will then have a different expression profile (as indicated by the different colours of cell nuclei), which will eventually lead to a diverse population of neurons that is necessary to perform the complex functions of the brain.

## 13. Conclusions and Perspectives

CFSs are regions prone to TRCs, replication stress and breakage as well as sources of increased genomic instability. Nevertheless, they, along with their characteristics, seem to be conserved among vertebrates. The large size of CFS genes somehow has resisted shortening throughout the evolutionary timeline, despite the fact that it contributes to their instability. Their conservation seems even more surprising when considering that their large size is owed to large introns and not to the coding regions of the genes. This suggests that CFSs might have a physiological role in the organisms. So far there are

indications pointing to a possible role of CFS-associated replication stress in neuronal diversification, however further investigation is needed to establish whether that is true or not. Future studies focusing on the role of the large introns of CFS genes both on cellular and organismal level will be necessary. Identifying the potential physiological role that replication stress might have in an organism is key to understand why CFSs are conserved and it will pave the way for a better understanding of cancer, brain development, and neurodegenerative diseases.

**Author Contributions:** Conceptualization, V.H.O. and V.V.; writing—original draft preparation, V.H.O, V.V. and S.H.N.M.; writing—review and editing, V.H.O, V.V. and S.H.N.M.; supervision, V.H.O.; funding acquisition, V.H.O.

**Funding:** This work was supported by the Villum Foundation.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Kim, J.C.; Mirkin, S.M. The balancing act of DNA repeat expansions. *Curr. Opin. Genet. Dev.* **2013**, *23*, 280–288. [[CrossRef](#)] [[PubMed](#)]
- Glover, T.W.; Wilson, T.E.; Arlt, M.F. Fragile sites in cancer: More than meets the eye. *Nat. Rev. Cancer* **2017**, *17*, 489–501. [[CrossRef](#)] [[PubMed](#)]
- Barlow, J.H.; Faryabi, R.B.; Callen, E.; Wong, N.; Malhowski, A.; Chen, H.T.; Gutierrez-Cruz, G.; Sun, H.W.; McKinnon, P.; Wright, G.; et al. Identification of early replicating fragile sites that contribute to genome instability. *Cell* **2013**, *152*, 620–632. [[CrossRef](#)] [[PubMed](#)]
- Subramanian, S.; Mishra, R.K.; Singh, L. Genome-wide analysis of microsatellite repeats in humans: Their abundance and density in specific genomic regions. *Genome Biol.* **2003**, *4*, R13. [[CrossRef](#)] [[PubMed](#)]
- Gymrek, M.; Willems, T.; Guilmatre, A.; Zeng, H.; Markus, B.; Georgiev, S.; Daly, M.J.; Price, A.L.; Pritchard, J.K.; Sharp, A.J.; et al. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat. Genet.* **2016**, *48*, 22–29. [[CrossRef](#)] [[PubMed](#)]
- Quilez, J.; Guilmatre, A.; Garg, P.; Highnam, G.; Gymrek, M.; Erlich, Y.; Joshi, R.S.; Mittelman, D.; Sharp, A.J. Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans. *Nucleic Acids Res.* **2016**, *44*, 3750–3762. [[CrossRef](#)] [[PubMed](#)]
- Hannan, A.J. Tandem repeats mediating genetic plasticity in health and disease. *Nat. Rev. Genet.* **2018**, *19*, 286–298. [[CrossRef](#)] [[PubMed](#)]
- Samadashwily, G.M.; Raca, G.; Mirkin, S.M. Trinucleotide repeats affect DNA replication in vivo. *Nat. Genet.* **1997**, *17*, 298–304. [[CrossRef](#)] [[PubMed](#)]
- Jeffreys, A.J.; Royle, N.J.; Wilson, V.; Wong, Z. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* **1988**, *332*, 278–281. [[CrossRef](#)] [[PubMed](#)]
- Thompson, R.; Zoppis, S.; McCord, B. An overview of DNA typing methods for human identification: Past, present, and future. *Methods Mol. Biol.* **2012**, *830*, 3–16. [[CrossRef](#)]
- McMurray, C.T. Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet.* **2010**, *11*, 786–799. [[CrossRef](#)] [[PubMed](#)]
- Orr, H.T.; Zoghbi, H.Y. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* **2007**, *30*, 575–621. [[CrossRef](#)] [[PubMed](#)]
- Kremer, E.J.; Pritchard, M.; Lynch, M.; Yu, S.; Holman, K.; Baker, E.; Warren, S.T.; Schlessinger, D.; Sutherland, G.R.; Richards, R.I. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)<sub>n</sub>. *Science* **1991**, *252*, 1711–1714. [[CrossRef](#)] [[PubMed](#)]
- Verkerk, A.J.; Pieretti, M.; Sutcliffe, J.S.; Fu, Y.H.; Kuhl, D.P.; Pizzuti, A.; Reiner, O.; Richards, S.; Victoria, M.F.; Zhang, F.P.; et al. Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **1991**, *65*, 905–914. [[CrossRef](#)]
- Hansen, R.S.; Canfield, T.K.; Lamb, M.M.; Gartler, S.M.; Laird, C.D. Association of fragile X syndrome with delayed replication of the *FMR1* gene. *Cell* **1993**, *73*, 1403–1409. [[CrossRef](#)]
- Sutherland, G.R. Fragile sites on human chromosomes: Demonstration of their dependence on the type of tissue culture medium. *Science* **1977**, *197*, 265–266. [[CrossRef](#)]
- Pelletier, R.; Krasilnikova, M.M.; Samadashwily, G.M.; Lahue, R.; Mirkin, S.M. Replication and expansion of trinucleotide repeats in yeast. *Mol. Cell. Biol.* **2003**, *23*, 1349–1357. [[CrossRef](#)] [[PubMed](#)]

18. Voineagu, I.; Surka, C.F.; Shishkin, A.A.; Krasilnikova, M.M.; Mirkin, S.M. Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. *Nat. Struct. Mol. Biol.* **2009**, *16*, 226–228. [[CrossRef](#)]
19. Mirkin, E.V.; Mirkin, S.M. Replication fork stalling at natural impediments. *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 13–35. [[CrossRef](#)]
20. Glover, T.W.; Berger, C.; Coyle, J.; Echo, B. DNA polymerase  $\alpha$  inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum. Genet.* **1984**, *67*, 136–142. [[CrossRef](#)]
21. Hecht, F.; Glover, T.W. Cancer chromosome breakpoints and common fragile sites induced by aphidicolin. *Cancer Genet. Cytogenet.* **1984**, *13*, 185–188. [[CrossRef](#)]
22. Bignell, G.R.; Greenman, C.D.; Davies, H.; Butler, A.P.; Edkins, S.; Andrews, J.M.; Buck, G.; Chen, L.; Beare, D.; Latimer, C.; et al. Signatures of mutation and selection in the cancer genome. *Nature* **2010**, *463*, 893–898. [[CrossRef](#)]
23. Beroukhim, R.; Mermel, C.H.; Porter, D.; Wei, G.; Raychaudhuri, S.; Donovan, J.; Barretina, J.; Boehm, J.S.; Dobson, J.; Urashima, M.; et al. The landscape of somatic copy-number alteration across human cancers. *Nature* **2010**, *463*, 899–905. [[CrossRef](#)] [[PubMed](#)]
24. Dereli-Oz, A.; Versini, G.; Halazonetis, T.D. Studies of genomic copy number changes in human cancers reveal signatures of DNA replication stress. *Mol. Oncol.* **2011**, *5*, 308–314. [[CrossRef](#)]
25. Karras, J.R.; Schrock, M.S.; Batar, B.; Huebner, K. Fragile genes that are frequently altered in cancer: Players not passengers. *Cytogenet. Genome Res.* **2016**, *150*, 208–216. [[CrossRef](#)] [[PubMed](#)]
26. Zlotorynski, E.; Rahat, A.; Skaug, J.; Ben-Porat, N.; Ozeri, E.; Hershberg, R.; Levi, A.; Scherer, S.W.; Margalit, H.; Kerem, B. Molecular basis for expression of common and rare fragile sites. *Mol. Cell. Biol.* **2003**, *23*, 7143–7151. [[CrossRef](#)] [[PubMed](#)]
27. Ozeri-Galai, E.; Lebofsky, R.; Rahat, A.; Bester, A.C.; Bensimon, A.; Kerem, B. Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. *Mol. Cell* **2011**, *43*, 122–131. [[CrossRef](#)] [[PubMed](#)]
28. Madireddy, A.; Kosiyatrakul, S.T.; Boisvert, R.A.; Herrera-Moyano, E.; Garcia-Rubio, M.L.; Gerhardt, J.; Vuono, E.A.; Owen, N.; Yan, Z.; Olson, S.; et al. FANCD2 facilitates replication through common fragile sites. *Mol. Cell* **2016**, *64*, 388–404. [[CrossRef](#)]
29. Zhang, H.; Freudenreich, C.H. An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in *S. cerevisiae*. *Mol. Cell* **2007**, *27*, 367–379. [[CrossRef](#)]
30. Le Tallec, B.; Millot, G.A.; Blin, M.E.; Brison, O.; Dutrillaux, B.; Debatisse, M. Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep.* **2013**, *4*, 420–428. [[CrossRef](#)]
31. Le Tallec, B.; Koundrioukoff, S.; Wilhelm, T.; Letessier, A.; Brison, O.; Debatisse, M. Updating the mechanisms of common fragile site instability: How to reconcile the different views? *Cell. Mol. Life Sci.* **2014**, *71*, 4489–4494. [[CrossRef](#)] [[PubMed](#)]
32. Le Tallec, B.; Dutrillaux, B.; Lachages, A.M.; Millot, G.A.; Brison, O.; Debatisse, M. Molecular profiling of common fragile sites in human fibroblasts. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1421–1423. [[CrossRef](#)] [[PubMed](#)]
33. Pentzold, C.; Shah, S.A.; Hansen, N.R.; Le Tallec, B.; Seguin-Orlando, A.; Debatisse, M.; Lisby, M.; Oestergaard, V.H. FANCD2 binding identifies conserved fragile sites at large transcribed genes in avian cells. *Nucleic Acids Res.* **2018**, *46*, 1280–1294. [[CrossRef](#)] [[PubMed](#)]
34. Wilson, T.E.; Arlt, M.F.; Park, S.H.; Rajendran, S.; Paulsen, M.; Ljungman, M.; Glover, T.W. Large transcription units unify copy number variants and common fragile sites arising under replication stress. *Genome Res.* **2015**, *25*, 189–200. [[CrossRef](#)] [[PubMed](#)]
35. Helmrich, A.; Stout-Weider, K.; Hermann, K.; Schrock, E.; Heiden, T. Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. *Genome Res.* **2006**, *16*, 1222–1230. [[CrossRef](#)]
36. Tubbs, A.; Sridharan, S.; van Wietmarschen, N.; Maman, Y.; Callen, E.; Stanlie, A.; Wu, W.; Wu, X.; Day, A.; Wong, N.; et al. Dual roles of poly(dA:dT) tracts in replication initiation and fork collapse. *Cell* **2018**. [[CrossRef](#)] [[PubMed](#)]
37. Gaillard, H.; Aguilera, A. Transcription as a threat to genome integrity. *Annu. Rev. Biochem.* **2016**, *85*, 291–317. [[CrossRef](#)] [[PubMed](#)]

38. Hamperl, S.; Bocek, M.J.; Saldivar, J.C.; Swigut, T.; Cimprich, K.A. Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell* **2017**, *170*, 774–786.e19. [[CrossRef](#)]
39. Gan, W.; Guan, Z.; Liu, J.; Gui, T.; Shen, K.; Manley, J.L.; Li, X. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **2011**, *25*, 2041–2056. [[CrossRef](#)]
40. Prado, F.; Aguilera, A. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J.* **2005**, *24*, 1267–1276. [[CrossRef](#)]
41. Helmrich, A.; Ballarino, M.; Tora, L. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* **2011**, *44*, 966–977. [[CrossRef](#)] [[PubMed](#)]
42. Oestergaard, V.H.; Lisby, M. Transcription-replication conflicts at chromosomal fragile sites-consequences in M phase and beyond. *Chromosoma* **2017**, *126*, 213–222. [[CrossRef](#)] [[PubMed](#)]
43. Macheret, M.; Halazonetis, T.D. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* **2018**, *555*, 112–116. [[CrossRef](#)] [[PubMed](#)]
44. Gros, J.; Kumar, C.; Lynch, G.; Yadav, T.; Whitehouse, I.; Remus, D. Post-licensing specification of eukaryotic replication origins by facilitated Mcm2-7 sliding along DNA. *Mol. Cell* **2015**, *60*, 797–807. [[CrossRef](#)] [[PubMed](#)]
45. Freudenreich, C.H. R-loops: Targets for nuclease cleavage and repeat instability. *Curr. Genet.* **2018**, *64*, 789–794. [[CrossRef](#)] [[PubMed](#)]
46. Su, X.A.; Freudenreich, C.H. Cytosine deamination and base excision repair cause R-loop-induced CAG repeat fragility and instability in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E8392–E8401. [[CrossRef](#)] [[PubMed](#)]
47. Neil, A.J.; Liang, M.U.; Khristich, A.N.; Shah, K.A.; Mirkin, S.M. RNA-DNA hybrids promote the expansion of Friedreich’s ataxia (GAA)<sub>n</sub> repeats via break-induced replication. *Nucleic Acids Res.* **2018**, *46*, 3487–3497. [[CrossRef](#)] [[PubMed](#)]
48. Meryet-Figuiera, M.; Alaei-Mahabadi, B.; Ali, M.M.; Mitra, S.; Subhash, S.; Pandey, G.K.; Larsson, E.; Kanduri, C. Temporal separation of replication and transcription during S-phase progression. *Cell Cycle* **2014**, *13*, 3241–3248. [[CrossRef](#)]
49. Wei, X.; Samarabandu, J.; Devdhar, R.S.; Siegel, A.J.; Acharya, R.; Berezney, R. Segregation of transcription and replication sites into higher order domains. *Science* **1998**, *281*, 1502–1506. [[CrossRef](#)]
50. Schubeler, D.; Scalzo, D.; Kooperberg, C.; van Steensel, B.; Delrow, J.; Groudine, M. Genome-wide DNA replication profile for *Drosophila melanogaster*: A link between transcription and replication timing. *Nat. Genet.* **2002**, *32*, 438–442. [[CrossRef](#)]
51. Le Beau, M.M.; Rassool, F.V.; Neilly, M.E.; Espinosa, R., 3rd; Glover, T.W.; Smith, D.I.; McKeithan, T.W. 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.* **1998**, *7*, 755–761. [[CrossRef](#)]
52. Techer, H.; Koundrioukoff, S.; Nicolas, A.; Debatisse, M. The impact of replication stress on replication dynamics and DNA damage in vertebrate cells. *Nat. Rev. Genet.* **2017**, *18*, 535–550. [[CrossRef](#)] [[PubMed](#)]
53. Letessier, A.; Millot, G.A.; Koundrioukoff, S.; Lachages, A.M.; Vogt, N.; Hansen, R.S.; Malfoy, B.; Brison, O.; Debatisse, M. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* **2011**, *470*, 120–123. [[CrossRef](#)]
54. Lemmens, B.; Hegarat, N.; Akopyan, K.; Sala-Gaston, J.; Bartek, J.; Hocheegger, H.; Lindqvist, A. DNA replication determines timing of mitosis by restricting CDK1 and PLK1 activation. *Mol. Cell* **2018**, *71*, 117–128. [[CrossRef](#)] [[PubMed](#)]
55. Saldivar, J.C.; Hamperl, S.; Bocek, M.J.; Chung, M.; Bass, T.E.; Cisneros-Soberanis, F.; Samejima, K.; Xie, L.; Paulson, J.R.; Earnshaw, W.C.; et al. An intrinsic S/G2 checkpoint enforced by ATR. *Science* **2018**, *361*, 806–810. [[CrossRef](#)] [[PubMed](#)]
56. Eykelenboom, J.K.; Harte, E.C.; Canavan, L.; Pastor-Pedro, A.; Calvo-Asensio, I.; Llorens-Agost, M.; Lowndes, N.F. ATR activates the S-M checkpoint during unperturbed growth to ensure sufficient replication prior to mitotic onset. *Cell Rep.* **2013**, *5*, 1095–1107. [[CrossRef](#)] [[PubMed](#)]
57. Sorensen, C.S.; Syljuasen, R.G.; Lukas, J.; Bartek, J. ATR, Claspin and the Rad9-Rad1-Hus1 complex regulate Chk1 and Cdc25A in the absence of DNA damage. *Cell Cycle* **2004**, *3*, 941–945. [[CrossRef](#)]
58. Torres-Rosell, J.; De Piccoli, G.; Cordon-Preciado, V.; Farmer, S.; Jarmuz, A.; Machin, F.; Pasero, P.; Lisby, M.; Haber, J.E.; Aragon, L. Anaphase onset before complete DNA replication with intact checkpoint responses. *Science* **2007**, *315*, 1411–1415. [[CrossRef](#)]



59. Casper, A.M.; Nghiem, P.; Arlt, M.F.; Glover, T.W. ATR regulates fragile site stability. *Cell* **2002**, *111*, 779–789. [[CrossRef](#)]
60. Bhat, A.; Andersen, P.L.; Qin, Z.; Xiao, W. Rev<sub>3</sub>, the catalytic subunit of Pol $\zeta$ , is required for maintaining fragile site stability in human cells. *Nucleic Acids Res.* **2013**, *41*, 2328–2339. [[CrossRef](#)]
61. Bergoglio, V.; Boyer, A.S.; Walsh, E.; Naim, V.; Legube, G.; Lee, M.Y.; Rey, L.; Rosselli, F.; Cazaux, C.; Eckert, K.A.; et al. DNA synthesis by Pol  $\eta$  promotes fragile site stability by preventing under-replicated DNA in mitosis. *J. Cell Biol.* **2013**, *201*, 395–408. [[CrossRef](#)] [[PubMed](#)]
62. Gallina, I.; Christiansen, S.K.; Pedersen, R.T.; Lisby, M.; Oestergaard, V.H. TopBP1-mediated DNA processing during mitosis. *Cell Cycle* **2016**, *15*, 176–183. [[CrossRef](#)] [[PubMed](#)]
63. Chan, K.L.; Palmi-Pallag, T.; Ying, S.; Hickson, I.D. Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nat. Cell Biol.* **2009**, *11*, 753–760. [[CrossRef](#)] [[PubMed](#)]
64. Naim, V.; Rosselli, F. The FANCD1 pathway and BLM collaborate during mitosis to prevent micro-nucleation and chromosome abnormalities. *Nat. Cell Biol.* **2009**, *11*, 761–768. [[CrossRef](#)] [[PubMed](#)]
65. Mamrak, N.E.; Shimamura, A.; Howlett, N.G. Recent discoveries in the molecular pathogenesis of the inherited bone marrow failure syndrome Fanconi anemia. *Blood Rev.* **2017**, *31*, 93–99. [[CrossRef](#)] [[PubMed](#)]
66. Okamoto, Y.; Iwasaki, W.M.; Kugou, K.; Takahashi, K.K.; Oda, A.; Sato, K.; Kobayashi, W.; Kawai, H.; Sakasai, R.; Takaori-Kondo, A.; et al. Replication stress induces accumulation of FANCD2 at central region of large fragile genes. *Nucleic Acids Res.* **2018**, *46*, 2932–2944. [[CrossRef](#)] [[PubMed](#)]
67. Durkin, S.G.; Arlt, M.F.; Howlett, N.G.; Glover, T.W. Depletion of CHK1, but not CHK2, induces chromosomal instability and breaks at common fragile sites. *Oncogene* **2006**, *25*, 4381–4388. [[CrossRef](#)]
68. Arlt, M.F.; Xu, B.; Durkin, S.G.; Casper, A.M.; Kastan, M.B.; Glover, T.W. BRCA1 is required for common-fragile-site stability via its G2/M checkpoint function. *Mol. Cell Biol.* **2004**, *24*, 6701–6709. [[CrossRef](#)]
69. Chan, K.L.; North, P.S.; Hickson, I.D. BLM is required for faithful chromosome segregation and its localization defines a class of ultrafine anaphase bridges. *EMBO J.* **2007**, *26*, 3397–3409. [[CrossRef](#)]
70. Naim, V.; Wilhelm, T.; Debatisse, M.; Rosselli, F. ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nat. Cell Biol.* **2013**, *15*, 1008–1015. [[CrossRef](#)]
71. Ying, S.; Minocherhomji, S.; Chan, K.L.; Palmi-Pallag, T.; Chu, W.K.; Wass, T.; Mankouri, H.W.; Liu, Y.; Hickson, I.D. MUS81 promotes common fragile site expression. *Nat. Cell Biol.* **2013**, *15*, 1001–1007. [[CrossRef](#)] [[PubMed](#)]
72. Bhowmick, R.; Minocherhomji, S.; Hickson, I.D. RAD52 facilitates mitotic DNA synthesis following replication stress. *Mol. Cell* **2016**, *64*, 1117–1126. [[CrossRef](#)] [[PubMed](#)]
73. Minocherhomji, S.; Ying, S.; Bjerregaard, V.A.; Bursomanno, S.; Aleliunaite, A.; Wu, W.; Mankouri, H.W.; Shen, H.; Liu, Y.; Hickson, I.D. Replication stress activates DNA repair synthesis in mitosis. *Nature* **2015**, *528*, 286–290. [[CrossRef](#)] [[PubMed](#)]
74. Malkova, A.; Ira, G. Break-induced replication: Functions and molecular mechanism. *Curr. Opin. Genet. Dev.* **2013**, *23*, 271–279. [[CrossRef](#)] [[PubMed](#)]
75. Anand, R.P.; Lovett, S.T.; Haber, J.E. Break-induced DNA replication. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a010397. [[CrossRef](#)] [[PubMed](#)]
76. Di Marco, S.; Hasanova, Z.; Kanagaraj, R.; Chappidi, N.; Altmannova, V.; Menon, S.; Sedlackova, H.; Langhoff, J.; Surendranath, K.; Huhn, D.; et al. RECQ5 helicase cooperates with MUS81 endonuclease in processing stalled replication forks at common fragile sites during mitosis. *Mol. Cell* **2017**, *66*, 658–671.e8. [[CrossRef](#)] [[PubMed](#)]
77. Wyatt, H.D.; Sarbajna, S.; Matos, J.; West, S.C. Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells. *Mol. Cell* **2013**, *52*, 234–247. [[CrossRef](#)]
78. Pedersen, R.T.; Kruse, T.; Nilsson, J.; Oestergaard, V.H.; Lisby, M. TopBP1 is required at mitosis to reduce transmission of DNA damage to G1 daughter cells. *J. Cell Biol.* **2015**, *210*, 565–582. [[CrossRef](#)]
79. Germann, S.M.; Schramke, V.; Pedersen, R.T.; Gallina, I.; Eckert-Boulet, N.; Oestergaard, V.H.; Lisby, M. TopBP1/Dpb11 binds DNA anaphase bridges to prevent genome instability. *J. Cell Biol.* **2014**, *204*, 45–59. [[CrossRef](#)]
80. Gritenaite, D.; Princz, L.N.; Szakal, B.; Bantele, S.C.; Wendeler, L.; Schilbach, S.; Habermann, B.H.; Matos, J.; Lisby, M.; Branzei, D.; et al. A cell cycle-regulated Slx4-Dpb11 complex promotes the resolution of DNA repair intermediates linked to stalled replication. *Genes Dev.* **2014**, *28*, 1604–1619. [[CrossRef](#)]

81. Deng, L.; Wu, R.A.; Kochenova, O.V.; Pellman, D.S.; Walter, J.C. Mitotic CDK promotes replisome disassembly, fork breakage, and complex DNA rearrangements. *BioRxiv* **2018**. [[CrossRef](#)]
82. Glover, T.W.; Stein, C.K. Induction of sister chromatid exchanges at common fragile sites. *Am. J. Hum. Genet.* **1987**, *41*, 882–890. [[PubMed](#)]
83. Lukas, C.; Savic, V.; Bekker-Jensen, S.; Doil, C.; Neumann, B.; Pedersen, R.S.; Grofte, M.; Chan, K.L.; Hickson, I.D.; Bartek, J.; et al. 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nat. Cell Biol.* **2011**, *13*, 243–253. [[CrossRef](#)]
84. Harrigan, J.A.; Belotserkovskaya, R.; Coates, J.; Dimitrova, D.S.; Polo, S.E.; Bradshaw, C.R.; Fraser, P.; Jackson, S.P. Replication stress induces 53BP1-containing OPT domains in G1 cells. *J. Cell Biol.* **2011**, *193*, 97–108. [[CrossRef](#)] [[PubMed](#)]
85. Fernandez-Vidal, A.; Vignard, J.; Mirey, G. Around and beyond 53BP1 Nuclear Bodies. *Int. J. Mol. Sci.* **2017**, *18*. [[CrossRef](#)] [[PubMed](#)]
86. Lezaja, A.; Altmeyer, M. Inherited DNA lesions determine G1 duration in the next cell cycle. *Cell Cycle* **2018**, *17*, 24–32. [[CrossRef](#)] [[PubMed](#)]
87. Barr, A.R.; Cooper, S.; Heldt, F.S.; Butera, F.; Stoy, H.; Mansfeld, J.; Novak, B.; Bakal, C. DNA damage during S-phase mediates the proliferation-quiescence decision in the subsequent G1 via p21 expression. *Nat. Commun.* **2017**, *8*, 14728. [[CrossRef](#)] [[PubMed](#)]
88. Arora, M.; Moser, J.; Phadke, H.; Basha, A.A.; Spencer, S.L. Endogenous replication stress in mother cells leads to quiescence of daughter cells. *Cell Rep.* **2017**, *19*, 1351–1364. [[CrossRef](#)]
89. Zhang, C.Z.; Spektor, A.; Cornils, H.; Francis, J.M.; Jackson, E.K.; Liu, S.; Meyerson, M.; Pellman, D. Chromothripsis from DNA damage in micronuclei. *Nature* **2015**, *522*, 179–184. [[CrossRef](#)]
90. Bonnet, A.; Grosso, A.R.; Elkaoutari, A.; Coleno, E.; Presle, A.; Sridhara, S.C.; Janbon, G.; Geli, V.; de Almeida, S.F.; Palancade, B. Introns protect eukaryotic genomes from transcription-associated genetic instability. *Mol. Cell* **2017**, *67*, 608–621.e6. [[CrossRef](#)]
91. Liu, L.F.; Wang, J.C. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7024–7027. [[CrossRef](#)] [[PubMed](#)]
92. Manzo, S.G.; Hartono, S.R.; Sanz, L.A.; Marinello, J.; De Biasi, S.; Cossarizza, A.; Capranico, G.; Chedin, F. DNA Topoisomerase I differentially modulates R-loops across the human genome. *Genome Biol.* **2018**, *19*, 100. [[CrossRef](#)] [[PubMed](#)]
93. Durkin, S.G.; Glover, T.W. Chromosome fragile sites. *Annu. Rev. Genet.* **2007**, *41*, 169–192. [[CrossRef](#)] [[PubMed](#)]
94. Bartkova, J.; Horejsi, Z.; Koed, K.; Kramer, A.; Tort, F.; Zieger, K.; Guldberg, P.; Sehested, M.; Nesland, J.M.; Lukas, C.; et al. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **2005**, *434*, 864–870. [[CrossRef](#)] [[PubMed](#)]
95. Wright, N.A.; Gregory, T.R.; Witt, C.C. Metabolic ‘engines’ of flight drive genome size reduction in birds. *Proc. Biol. Sci.* **2014**, *281*, 20132780. [[CrossRef](#)]
96. Wei, P.C.; Chang, A.N.; Kao, J.; Du, Z.; Meyers, R.M.; Alt, F.W.; Schwer, B. Long neural genes harbor recurrent dna break clusters in neural stem/progenitor cells. *Cell* **2016**, *164*, 644–655. [[CrossRef](#)]
97. Glover, T.W.; Wilson, T.E. Molecular biology: Breaks in the brain. *Nature* **2016**, *532*, 46–47. [[CrossRef](#)]
98. Finlay, B.L.; Slattery, M. Local differences in the amount of early cell death in neocortex predict adult local specializations. *Science* **1983**, *219*, 1349–1351. [[CrossRef](#)]
99. Ferrer, I.; Soriano, E.; del Rio, J.A.; Alcantara, S.; Auladell, C. Cell death and removal in the cerebral cortex during development. *Prog. Neurobiol.* **1992**, *39*, 1–43. [[CrossRef](#)]
100. Cai, X.; Evrony, G.D.; Lehmann, H.S.; Elhosary, P.C.; Mehta, B.K.; Poduri, A.; Walsh, C.A. Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep.* **2014**, *8*, 1280–1289. [[CrossRef](#)]
101. McConnell, M.J.; Lindberg, M.R.; Brennand, K.J.; Piper, J.C.; Voet, T.; Cowing-Zitron, C.; Shumilina, S.; Lasken, R.S.; Vermeesch, J.R.; Hall, I.M.; et al. Mosaic copy number variation in human neurons. *Science* **2013**, *342*, 632–637. [[CrossRef](#)] [[PubMed](#)]
102. McKinnon, P.J. Maintaining genome stability in the nervous system. *Nat. Neurosci.* **2013**, *16*, 1523–1529. [[CrossRef](#)] [[PubMed](#)]

103. O'Driscoll, M.; Ruiz-Perez, V.L.; Woods, C.G.; Jeggo, P.A.; Goodship, J.A. A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. *Nat. Genet.* **2003**, *33*, 497–501. [[CrossRef](#)] [[PubMed](#)]
104. Lee, Y.; Katyal, S.; Downing, S.M.; Zhao, J.; Russell, H.R.; McKinnon, P.J. Neurogenesis requires TopBP1 to prevent catastrophic replicative DNA damage in early progenitors. *Nat. Neurosci.* **2012**, *15*, 819–826. [[CrossRef](#)] [[PubMed](#)]
105. Roy, N.S.; Wang, S.; Jiang, L.; Kang, J.; Benraiss, A.; Harrison-Restelli, C.; Fraser, R.A.; Couldwell, W.T.; Kawaguchi, A.; Okano, H.; et al. In vitro neurogenesis by progenitor cells isolated from the adult human hippocampus. *Nat. Med.* **2000**, *6*, 271–277. [[CrossRef](#)] [[PubMed](#)]
106. Nunes, M.C.; Roy, N.S.; Keyoung, H.M.; Goodman, R.R.; McKhann, G., 2nd; Jiang, L.; Kang, J.; Nedergaard, M.; Goldman, S.A. Identification and isolation of multipotential neural progenitor cells from the subcortical white matter of the adult human brain. *Nat. Med.* **2003**, *9*, 439–447. [[CrossRef](#)] [[PubMed](#)]
107. Spalding, K.L.; Bergmann, O.; Alkass, K.; Bernard, S.; Salehpour, M.; Huttner, H.B.; Bostrom, E.; Westerlund, I.; Vial, C.; Buchholz, B.A.; et al. Dynamics of hippocampal neurogenesis in adult humans. *Cell* **2013**, *153*, 1219–1227. [[CrossRef](#)] [[PubMed](#)]
108. Ge, X.Q.; Han, J.; Cheng, E.C.; Yamaguchi, S.; Shima, N.; Thomas, J.L.; Lin, H. Embryonic stem cells license a high level of dormant origins to protect the genome against replication stress. *Stem Cell Rep.* **2015**, *5*, 185–194. [[CrossRef](#)] [[PubMed](#)]
109. Erwin, J.A.; Marchetto, M.C.; Gage, F.H. Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat. Rev. Neurosci.* **2014**, *15*, 497–506. [[CrossRef](#)] [[PubMed](#)]
110. Singer, T.; McConnell, M.J.; Marchetto, M.C.; Coufal, N.G.; Gage, F.H. LINE-1 retrotransposons: Mediators of somatic variation in neuronal genomes? *Trends Neurosci.* **2010**, *33*, 345–354. [[CrossRef](#)] [[PubMed](#)]
111. Muotri, A.R.; Chu, V.T.; Marchetto, M.C.; Deng, W.; Moran, J.V.; Gage, F.H. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **2005**, *435*, 903–910. [[CrossRef](#)] [[PubMed](#)]
112. Kuwabara, T.; Hsieh, J.; Muotri, A.; Yeo, G.; Warashina, M.; Lie, D.C.; Moore, L.; Nakashima, K.; Asashima, M.; Gage, F.H. Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nat. Neurosci.* **2009**, *12*, 1097–1105. [[CrossRef](#)] [[PubMed](#)]
113. Gilbert, N.; Lutz, S.; Morrish, T.A.; Moran, J.V. Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol. Cell. Biol.* **2005**, *25*, 7780–7795. [[CrossRef](#)] [[PubMed](#)]
114. Gasior, S.L.; Wakeman, T.P.; Xu, B.; Deininger, P.L. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* **2006**, *357*, 1383–1393. [[CrossRef](#)] [[PubMed](#)]
115. Carruthers, R.D.; Ahmed, S.U.; Ramachandran, S.; Strathdee, K.; Kurian, K.M.; Hedley, A.; Gomez-Roman, N.; Kalna, G.; Neilson, M.; Gilmour, L.; et al. Replication stress drives constitutive activation of the DNA damage response and radioresistance in glioblastoma stem-like cells. *Cancer Res.* **2018**, *78*, 5060–5071. [[CrossRef](#)]
116. Kim, J.; Sturgill, D.; Sebastian, R.; Khurana, S.; Tran, A.D.; Edwards, G.B.; Kruswick, A.; Burkett, S.; Hosogane, E.K.; Hannon, W.W.; et al. Replication stress shapes a protective chromatin environment across fragile genomic regions. *Mol. Cell* **2018**, *69*, 36–47.e7. [[CrossRef](#)] [[PubMed](#)]
117. Chen, H.; Ruiz, P.D.; Novikov, L.; Casill, A.D.; Park, J.W.; Gamble, M.J. MacroH2A1.1 and PARP-1 cooperate to regulate transcription by promoting CBP-mediated H2B acetylation. *Nat. Struct. Mol. Biol.* **2014**, *21*, 981–989. [[CrossRef](#)]
118. Zhang, R.; Poustovoitov, M.V.; Ye, X.; Santos, H.A.; Chen, W.; Daganzo, S.M.; Erzberger, J.P.; Serebriiskii, I.G.; Canutescu, A.A.; Dunbrack, R.L.; et al. Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. *Dev. Cell* **2005**, *8*, 19–30. [[CrossRef](#)]
119. Papadopoulou, C.; Guilbaud, G.; Schiavone, D.; Sale, J.E. Nucleotide pool depletion induces g-quadruplex-dependent perturbation of gene expression. *Cell Rep.* **2015**, *13*, 2491–2503. [[CrossRef](#)]
120. Smith, D.I.; Zhu, Y.; McAvoy, S.; Kuhn, R. Common fragile sites, extremely large genes, neural development and cancer. *Cancer Lett.* **2006**, *232*, 48–57. [[CrossRef](#)]

