

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/29979> holds various files of this Leiden University dissertation

**Author:** Lemmens, Bennie

**Title:** Repair and genetic consequences of DNA double strand breaks during animal development

**Issue Date:** 2014-12-09

# 6

## **Homology-directed repair bypasses polymerase Theta-mediated end joining of G quadruplex- induced DNA breaks**

Lemmens BB, van Schendel R, Tijsterman M

**Abstract**

Damaged DNA bases and DNA secondary structures such as G4 quadruplexes impede DNA replication and promote the occurrence of deleterious DNA double strand breaks (DSBs). In recent years, several alternative repair mechanisms have been found to repair DSBs parallel to the heavily studied pathways non-homologous end joining (NHEJ) and homologous recombination (HR). Yet how these pathways interact and what dictates pathway choice remains poorly understood. We recently identified polymerase Theta mediated end joining (TMEJ) to be the major pathway responsible for mutagenic repair of replication-born DSBs in *C. elegans* and to prevail over NHEJ and HR in the repair of G4-induced DSBs. Here we establish that DNA sequence context can dictate repair pathway choice of G4-induced DSBs and identify a mutagenic homology-driven repair (HDR) mechanism that uses >4 base pair homology at the presumptive DSB ends and can bypass the requirement of polymerase Theta/POLQ-1 for G4-induced deletion formation. Deletion frequency analysis at endogenous G4 sites revealed that HDR can locally dominate over TMEJ and that some TMEJ substrates can be channeled into HDR, illustrating that both mechanisms can repair similar substrates. However, given the specific homology requirements of HDR, TMEJ remains the major repair route genome-wide. We propose that the key role of POLQ-1 polymerase is to create *de novo* homologous sequences at DSB ends, providing a stable double-stranded intermediate to extend 3' DNA ends and seal replication-born DSBs.

## Introduction

Impediments to DNA replication hamper the cell in copying its genome with high fidelity and therefore are a major threat to genome stability (TOURRIERE AND PASERO 2007; BUDZOWSKA AND KANAAR 2009; PRESTON *et al.* 2010). Damaged DNA templates, as well as unresolved DNA secondary structures, can stall replicative polymerases, preventing duplication of the DNA past the lesion. One well-studied DNA secondary structure that is very stable under physiological conditions and is a potent replication block *in vitro* is the G4 quadruplex (HOWELL *et al.* 1996; HAN *et al.* 1999; HUPPERT 2010). Given that the human genome harbors more than 300,000 guanine rich motifs that can adopt G4 quadruplex configurations, G4 DNA poses a serious threat to replication fidelity. Several specialized DNA helicases such as FANCI/DOG-1 can resolve G4 quadruplexes *in vitro* and loss of these helicases results in elevated levels of genomic rearrangements at G4 sites *in vivo* (TARSOUNAS AND TIJSTERMAN 2013; MURAT AND BALASUBRAMANIAN 2014). Recent data indicate that G4 DNA can cause genetic and epigenetic alterations in various model organisms and G4 motifs have been associated with structural genomic variations in human cancers (TARSOUNAS AND TIJSTERMAN 2013; MURAT AND BALASUBRAMANIAN 2014). The molecular mechanisms responsible for G4-induced genomic variations are, however, poorly understood.

In the current model, replication fork stalling at unresolved G4 structures causes DNA double strand breaks (DSBs) that can be repaired in an error-prone fashion, possibly leading to genomic rearrangements that drive malignant transformation (KOOLE *et al.* 2014; MURAT AND BALASUBRAMANIAN 2014). In recent years, several alternative repair pathways have been identified that can repair DSBs parallel to the two major DSB repair pathways: non-homologous end joining (NHEJ) and homologous recombination (HR) (FISHMAN-LOBELL *et al.* 1992; MA *et al.* 2003; CHAN *et al.* 2010; DERIANO AND ROTH 2013; ROERINK *et al.* 2014). While all DSB repair pathways have been shown to act in replicating cells, their relative contribution to DSB repair depends heavily on sequence context. For instance, NHEJ can seal DSBs independent of sequence context (0 nt), while HR requires extensive sequence homology from an undamaged template (>100nt) (SAN FILIPPO *et al.* 2008; LIEBER 2010). The alternative DSB repair routes often require base pairing of complementary DNA to align and seal DSB ends, yet the extent of this DNA template dependence varies among the different pathways. In fact, DSB repair via polymerase Theta mediated end joining (TMEJ) hardly requires homology ( $\leq 1$ nt), while other homology driven repair (HDR) routes typically require longer stretches of complementary sequences to align and repair the break (>4nt) (FISHMAN-LOBELL *et al.* 1992; MA *et al.* 2003; ROERINK *et al.* 2014). To date, the molecular characteristics and genetic requirements of mutagenic HDR are still ill defined and may encompass interconnected mechanisms that have previously been designated as alternative end joining (alt-EJ), micro-homology mediated end joining (MMEJ) or single strand annealing (SSA), all of which use

complementary DNA sequences to seal DSBs. Since the exact nature and genetic distinction of these individual pathways is still controversial, we here define HDR as a mutagenic DSB repair mode that uses >4nt homology.

We recently identified TMEJ as the major repair route responsible for mutagenic repair of replication-associated DSBs in *C. elegans* and found TMEJ to prevail over NHEJ and HR in the repair of G4-induced DSBs (KOOLE *et al.* 2014; ROERINK *et al.* 2014). The highly conserved polymerase theta/POLQ-1 is at the heart of TMEJ and is also required for the frequent flank insertions associated with TMEJ, yet why such a polymerase is crucial for replication-associated DSB repair is still unclear.

Here we investigate alternative DSB repair mechanisms that may act on G4-induced substrates and found that HDR can also result in deletion formation. In fact, we found that deletion formation at endogenous G4 sites that are flanked by short stretches of homology can occur independent of POLQ-1 activity and can locally dominate over TMEJ when homology is readily present at the presumptive DSB ends. All G4 deletions smaller than 600 base pairs (bp) found in *polq-1*-deficient animals used flanking homology, representing an alternative repair product to the >10,000 bp deletion products reported in *polq-1* mutants at non-repetitive loci (KOOLE *et al.* 2014). In accordance with previous studies, we found G4 sites not flanked by apparent homology to depend entirely on POLQ-1 for deletion formation (<1kb). Deletion frequency analysis at endogenous G4 sites revealed that some TMEJ substrates can be channeled into HDR, yet the efficiency is limited and correlates with homology abundance.

All together these data argue that G4-induced DSBs can be repaired by TMEJ and HDR, yet given the specific homology requirements for HDR, TMEJ is the major repair route genome-wide. Nevertheless, the bypass of POLQ-1 requirement by the presence of flanking homology strongly suggests that the major role for POLQ-1 is to create minimal homology between both DSB ends, providing a stable double-stranded intermediate to extent 3' DNA ends and seal replication born DSBs.

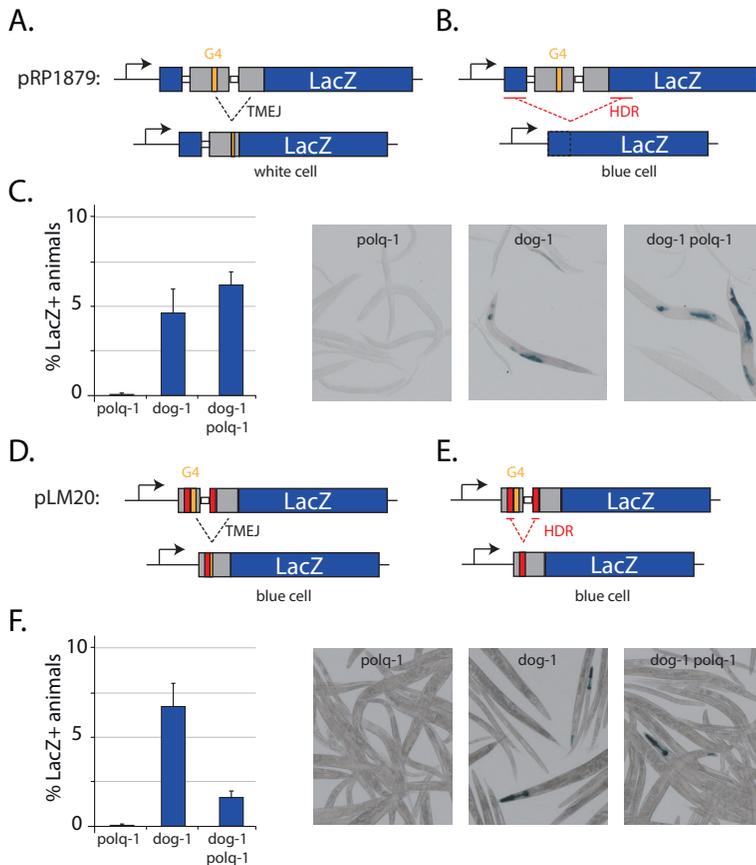
## Results

### Transgenic reporters reveal POLQ-1-independent deletion formation

The first indication that G4-induced DSBs could be repaired by other means than TMEJ came from the analysis of a transgenic reporter assay we had previously employed to identify genes required for G4 instability (Figure 1A). We previously screened randomly mutagenized animals carrying a multi-copy transgene of pRP1879, a LacZ-based reporter construct that reads out mutagenic repair of G4-induced DSBs and found only *dog-1*-deficient animals to express LacZ positive cells (KRUISSELBRINK *et al.* 2008). Given the established role of DOG-1 in suppression of G4 instability at endogenous G4 loci, this screen validated our transgenic reporter setup, however the nature of repair events underlying LacZ expression was unknown.

Deletions at endogenous G4 sites typically rely on TMEJ and are uni-directional, 50-300bp in size, use 0-1bp homology and start directly at the G4 motif (KOOLE *et al.* 2014). Consequently, G4-induced TMEJ events normally leave the sequence upstream of the G4 motif intact (Figure 1A, G4 motif in yellow). The pRP1879 transgene, however, expresses LacZ only when both stop codons flanking the G4 site are removed, which means that a part of the upstream sequence needs to be lost during the repair reaction (Figure 1A, white bars represent stop codon). In fact, HDR using the upstream homologous (230bp) LacZ repeat would result in exclusion of both stop codons and render the LacZ open reading frame (ORF) functional (Figure 1B). Since deletions at G4 sites typically rely on TMEJ, we wondered if these putative HDR events would also rely on POLQ-1. Strikingly, ORF correction of pRP1879 still occurred with high frequency in *polq-1*-deficient animals, arguing that G4-induced deletions can occur in a TMEJ-independent fashion (Figure 1C).

To study whether the position and length of flanking homology could influence *polq-1* dependency, we constructed a novel LacZ-based reporter transgene pLM20, in which we positioned a shorter (50bp) homologous sequence directly upstream (left) and 24bp downstream (right) of the G4 site, allowing both TMEJ and HDR to correct the reporter ORF (Figure 1D and 1E, homology depicted in red). Also here pLM20 transgene expression was specifically induced upon *dog-1* deficiency, suggesting that LacZ expression indeed reflects repair of G4-induced DSBs (Figure 1F). To test the relative contribution of TMEJ in pLM20 expression, we generated *dog-1; polq-1* double mutants and found LacZ expression to be significantly reduced, but not absent, compared to *dog-1* single mutants (Figure 1F). These results imply that most pLM20 ORF correction events depend on TMEJ, yet some level of TMEJ-independent repair is possible. Based on these observations we hypothesized that HDR of G4-induced DSBs is possible and, depending on the degree of flanking homology, may compete with TMEJ for repair.



**Figure 1. Transgenic reporters reveal POLQ-1-independent deletion formation of G4 sites**

**A.** Schematic diagram of G4 instability reporter pRP1879 and repair outcome via TMEJ. pRP1879 is driven by a *hsp-16.41* promoter that allows expression in various somatic tissues. Deletion formation by TMEJ is not expected to remove the upstream stop codon (white blocks) and thus will not result in a blue cell. **B.** Schematic diagram of G4 instability reporter pRP1879 and repair outcome via HDR. HDR using the upstream repeats corrects the LacZ ORF and will result in a blue cell. **C.** Histogram shows quantification of stochastic pRP1879 ORF correction measured by the percentage of LacZ-positive animals of the indicated genotype. Average percentage of LacZ-positive animals of three independent experiments is depicted and error bars represent S.E.M. Representative pictures of the stochastic LacZ expression patterns are shown on the right. **D.** Schematic diagram of G4 instability reporter pLM20 and repair outcome via TMEJ. pLM20 is driven by a *myo-2* promoter that allows specific expression in pharyngeal muscle cells. Deletion formation by TMEJ is expected to remove the downstream stop codon (white blocks) irrespective of the homologous repeats (red) and can result an in-frame LacZ ORF and a blue cell. **E.** Schematic diagram of G4 instability reporter pLM20 and repair outcome via HDR using the 50bp repeats (red). HDR using the homologous repeats corrects the LacZ ORF and will result in a blue cell. **F.** Histogram shows quantification of stochastic pLM20 ORF correction measured by the percentage of LacZ-positive animals of the indicated genotype. Average percentage of LacZ-positive animals of three independent experiments is depicted and error bars represent S.E.M. Representative pictures of the stochastic LacZ expression patterns are shown on the right.

### Flanking homology at endogenous G4 sites allows POLQ-1-independent deletion formation

In order to directly compare TMEJ and HDR events in a single-copy environment, we searched the *C. elegans* genome for endogenous G4 sites that had different degrees of flanking homology. We selected four loci that resided on chromosome I, II, III and IV, respectively: Qua213, Qua375, Qua915 and Qua1277. While Qua375 has no apparent flanking homology, Qua1277, Qua213 and Qua915 have increasing levels of flanking homology, respectively, and harbor short genomic repeats that potentially could support HDR (Figure 2A and S1 for entire sequence context). Especially Qua915 is located in a highly repetitive genomic context and is flanked by many different short repeats and three major repeats of 29bp (Figure 2A).

To analyze deletion formation at these repetitive loci and study TMEJ dependency, we performed nested PCR reactions on genomic DNA lysates of *polq-1*, *dog-1* and *dog-1; polq-1* double mutant animals using primers that flank the G4 motif as well as the surrounding repeat sequences (Figure 2B and S1). While *polq-1* single mutants did not display G4 instability at any of the loci tested, *dog-1*-deficient animals showed many differently sized deletions at all four loci, indicating that also at these loci the DOG-1 helicase is required to prevent the induction of G4-induced deletions. To test if these deletions depended on TMEJ, we analyzed deletion formation in *dog-1; polq-1* double mutants and found that indeed all small deletions at Qua375 depended on POLQ-1 (Figure 2B). In contrast, Qua1277, Qua213 and Qua915 still showed some small deletion in the absence of POLQ-1, indicating that these homology-rich loci can spawn deletions independent of TMEJ (Figure 2B). Interestingly, the residual deletions at Qua1277 or Qua213 were always of identical size, suggestive of a preferred repair outcome specific for each locus. Sequence analysis revealed that all TMEJ-independent deletion events at Qua1277 or Qua213 used HDR based on the two major repeats flanking the G4 sites (Figure 2A).

### HDR footprints imply available 3' overhangs at G4-induced DSBs

While these observations provided further evidence that TMEJ-independent repair mechanisms exist, they also provided clues regarding the molecular nature of the predicted G4-induced substrate: a replication-derived DSB (see Figure 3 for the current model). In TMEJ-proficient *dog-1* animals Qua375 and Qua1277 behaved very similarly: all deletions were uni-directional and started directly at the G4 motif (Figure 2C). The typical position of the upstream deletion breakpoint implies that the location of the G4 structure dictates the position of the resultant DSB end and the strong preservation of the sequences upstream the G4 motif suggests that the 3' end of this DSB is quite stable during the repair process (Figure 2C and 2D).

Furthermore, the HDR footprints at Qua1277 in *dog-1; polq-1* animals imply that the upstream repeat was exposed and able to anneal to its homologous counterpart, suggesting that the upstream DSB end was not blunt ended but instead had >100bp 3' overhangs (Figure

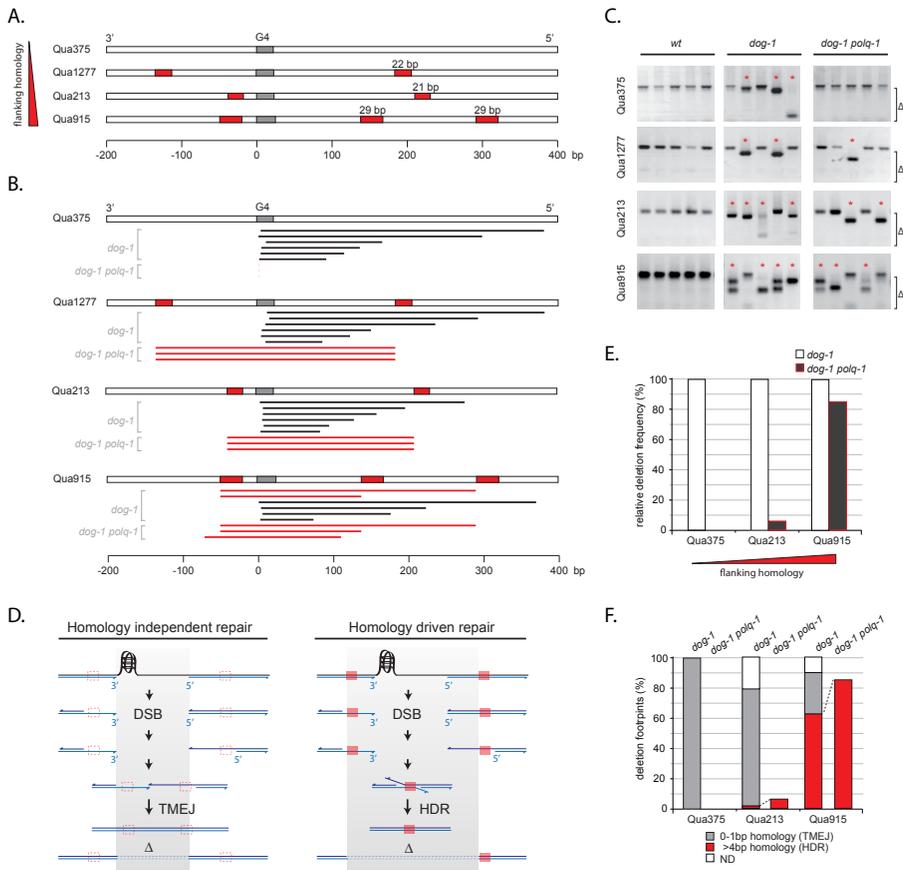
2C and 2D). This observation is supported by the data acquired using the pRP1879 transgene. Similar to the upstream repeat at Qua1277, the upstream LacZ repeat in pRP1879, positioned 270bp away from the G4 motif, needs to be available to allow HDR to occur and the LacZ reporter gene to be expressed (Figure 1A). Likewise, G4 deletions at Qua915 frequently involved annealing of repeats residing 15-80bp upstream of the G4 motif, even in the presence of POLQ-1, suggesting that the upstream DSB end may intrinsically have a 3' overhang that could serve as a substrate for both HDR and TMEJ; a feature consistent with a model for replication-born DSBs (Figure 2C and 3).

### No evidenced for NHEJ activity at G4-induced DSBs

The idea that G4-derived DSBs inherently may have substantial 3' overhangs would be in line with the reported lack of NHEJ activity on these substrates (YOUNDS *et al.* 2006; KOOLE *et al.* 2014). NHEJ can efficiently repair blunt-ended DSBs but not resected DSBs (LIEBER 2010). Although both NHEJ and TMEJ should be able to repair DSBs in the absence of homologous sequences, TMEJ is the pathway of choice to repair G4-induced DSBs (KOOLE *et al.* 2014). Recently several well-conserved factors, including FANCD2/*fcd-2* and CtIP/*com-1*, have been identified that suppress NHEJ activity at endogenous DSBs by initiating DNA end resection (ADAMO *et al.* 2010; LEMMENS *et al.* 2013). To test if NHEJ would be able to repair G4-induced DSBs in the absence of these NHEJ-suppressors, we constructed *dog-1; polq-1; fcd-2* and *dog-1; polq-1; com-1* triple mutants and analyzed deletion formation at Qua375. Similar to the *dog-1; polq-1* double mutant controls, none of the triple mutant animals showed homology-independent deletions, suggesting that also in these genetic backgrounds NHEJ cannot act on G4-derived DSBs (Figure S2). In contrast, we observed many deletion products in the POLQ-1 proficient *dog-1* controls, indicating that TMEJ is the key pathway to repair G4-induced DSBs in the absence of flanking homology (Figure S2).

### A limited number of TMEJ events can channel into HDR

To study the effects of TMEJ deficiency and address the relative contribution of TMEJ and HDR events, we examined deletion formation at Qua375, Qua213 and Qua915 in >360 *dog-1* and >360 *dog-1;polq-1* animals and determined the deletion frequencies at the different G4 loci in the same population (Figure 2E). While in all cases the deletion frequency dropped in *dog-1;polq-1* animals compared to *dog-1* single mutants, the extent of *polq-1* dependency differed substantially between loci. While the deletion frequency at the repetitive locus Qua915 was hardly affected by *polq-1* loss (85%), the deletion frequency at Qua213 dropped drastically in the absence of *polq-1* (6%) and deletions were even completely absent at Qua375 (0%), suggesting that the extent of flanking homology correlates directly to the potency of deletion formation in the absence of POLQ-1 (Figure 2E).

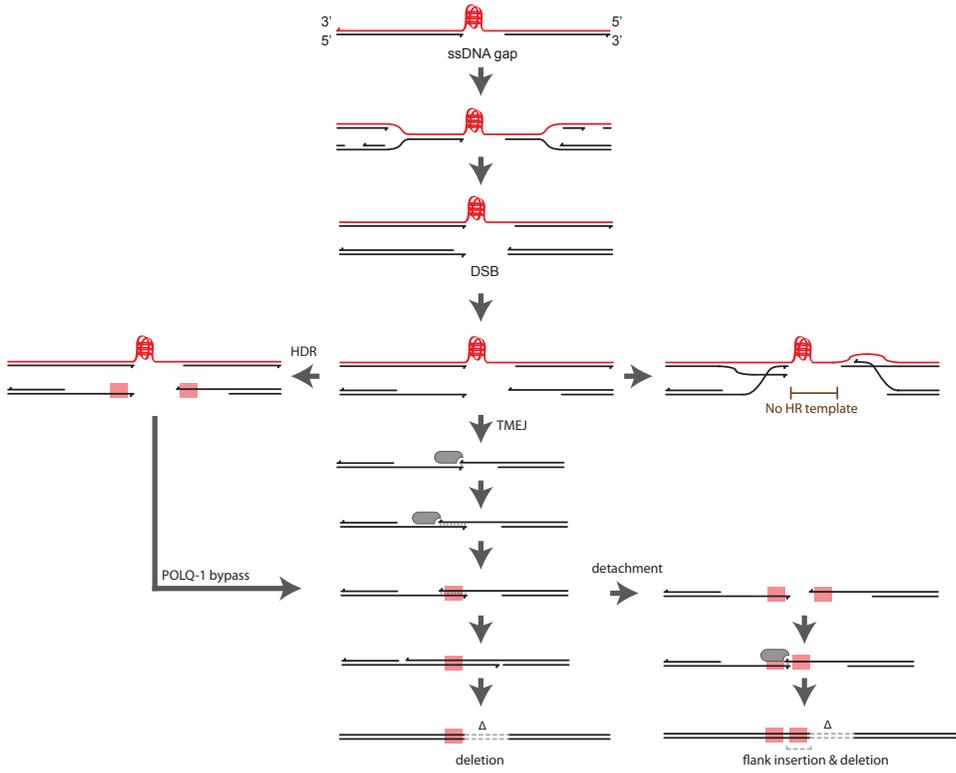


**Figure 2. Flanking homology at endogenous G4 sites allows POLQ-1-independent deletion formation**

**A.** Schematic diagram of four endogenous G4 loci with different degrees of flanking homology. G4 loci are aligned relative to the 5' position of the G4 motif and the most prominent homologous repeats are indicated in red. **B.** Graphic illustration of G4 deletions profiles at three endogenous G4 loci. For each locus six typical G4 deletions in *dog-1* and three typical G4 deletions in *dog-1; polq-1* animals are depicted. Black bars represent homology-independent deletions; red bars represent homology-dependent events. **C.** Representative images of the different PCR-based assays used to identify G4-induced deletions at the indicated G4 loci. Per lane genomic DNA of three adult animals was PCR-amplified using primers flanking the G4 motif and homologous repeats. Asterisks indicate stochastic deletions, which manifest as shorter than wild-type products and  $\Delta$  indicates the size-range of the PCR-amplified deletion products. **D.** Models for G4-induced deletions formation via TMEJ (left) and HDR (right). Filled red boxes indicate homologous sequences. Grey gradients illustrate the association between deletion size/position and the G4-induced ssDNA gap (in case of TMEJ) or the position of the homologous repeats (in case of HDR). **E.** Histogram depicts relative deletion frequencies at the indicated G4 loci as determined by the PCR-based assay on 1% single worm lysates of *dog-1* (white bars) and *dog-1; polq-1* animals (black bars). Depicted frequencies are relative to the deletion frequency in *dog-1* single mutants to allow comparison of loci expressing different stochastic G4 deletion rates ( $n > 360$ , see methods section for details). **F.** Histogram depicts relative frequencies of deletion footprints with  $>4$ bp homology (red) and without homology (grey) as identified by the PCR-based assay at the indicated G4 loci (see methods section for details). Repair footprints were analyzed from PCR-amplified deletion products obtained from 1% single worm lysates of *dog-1* (left) and *dog-1; polq-1* animals (right).

We next extracted the deletion products at Qua375, Qua213 and Qua915 and determined the repair footprints. All residual deletions in *polq-1*-deficient animals used extensive homology (>4bp) and deleted sequences far upstream of the G4 motif, indicative of HDR events (Figure 2F). Importantly, POLQ-1-proficient *dog-1* animals also showed HDR footprints, revealing that TMEJ does not completely suppress HDR. When we directly compared the frequencies of TMEJ and HDR footprints among *dog-1* and *dog-1;polq-1* animals, we observed an increase in HDR events at the expense of TMEJ products in *polq-1* deficient animals, suggesting that some TMEJ substrates can be channeled into HDR (Figure 2F).

6



**Figure 3. Model for G4-induced deletion formation**  
Model describes the origin of replication-born DSBs from G4-induced ssDNA gaps and subsequent DSB repair via HDR and TMEJ. The DNA strand bearing the persistent G4 quadruplex is depicted in red. Transparent red boxes indicate homologous sequences. Polymerase Theta/POLQ-1 is depicted in grey. See main text for further details.

### HDR can locally dominate repair of G4-induced DSBs

The observation that the requirement for POLQ-1 in G4 deletion formation can be bypassed by the presence of homologous sequences lead us to question which repair mode was initiated first. To investigate how sequence context around G4 motifs controlled the choice between HDR and TMEJ, we plotted the distribution of all homology-independent deletions at Qua375, Qua213 and Qua915, and sorted the deletions based on their position relative to the G4 motif (illustrated by the black bars in Figure 4A). Subsequently all deletions were binned in 50bp windows to obtain a 5' deletion junctions distribution relative to the G4 motif (4B). In accordance with previous studies on non-repetitive G4 loci (KOOLE *et al.* 2014), we found the vast majority of Qua375 and Qua213 deletions to be 50-300bp in size (>85%), with most 5' deletion junctions residing 101-150bp downstream of the G4 motif (~30%) (Figure 4B, white and grey bars). The G4 deletions at Qua915 were also typically 50-300bp in size (>85%), however the distribution of 5' deletion junctions was significantly different (Figure 4B, black bars). Strikingly, none of the homology-independent deletions at Qua915 had 5' deletion junctions residing 101-150bp downstream of the G4 motif (0%), clearly contrasting the distribution found at other G4 sites (Figure 4B, highlighted in pink). Also no homology-independent Qua915 deletions were observed 251-300bp downstream of the G4, while these were observed at Qua375 and Qua213 (Figure 4B, highlighted in pink). The fact that the major homologous repeats flanking Qua915 are located exactly downstream of the regions devoid of homology-independent deletions strongly argues for a dominant role of these repeats in sequestering TMEJ substrates. These data imply that G4-derived DSBs that contain homologous repeats of sufficient size at both break ends are preferably repaired via HDR and not TMEJ (Figure 4C, middle panel). Such a dominant effect of homologous sequences also explains why HDR events are frequent at Qua915, even in the presence of functional POLQ-1 (Figure 2F).

We also noted that the dominant effect of the homologous sequences only suppressed the TMEJ events with 5' deletion junctions upstream (left) of the repeat but not those with 5' junctions directly downstream (right) of the repeat (Figure 4B, highlighted in pink and grey, respectively). In fact, the vast majority of the homology-independent deletions at Qua915 had 5' deletion junctions residing 151-200bp downstream of the G4 motif (~45%), which is directly adjacent to the dominant repeat (Figure 4B, highlighted in grey). This directional effect of the homologous sequences is in perfect agreement with our model of G4-derived DSBs (Figure 3 and 4C). This model predicts that the 5' deletion junction of a TMEJ event is determined by the size of the ssDNA gap caused by the replication-blocking G4 structure: replication of the G4-induced ssDNA gap results into a DSB that lacks the DNA sequence covered previously by the ssDNA gap (Figure 3 and 4C). Repair of these DSBs by TMEJ results in deletions with junctions corresponding directly to the position and size of the initial gap (Figure 2D and 3). In cases where the ssDNA gaps are small and reach just up to the homologous sequences,

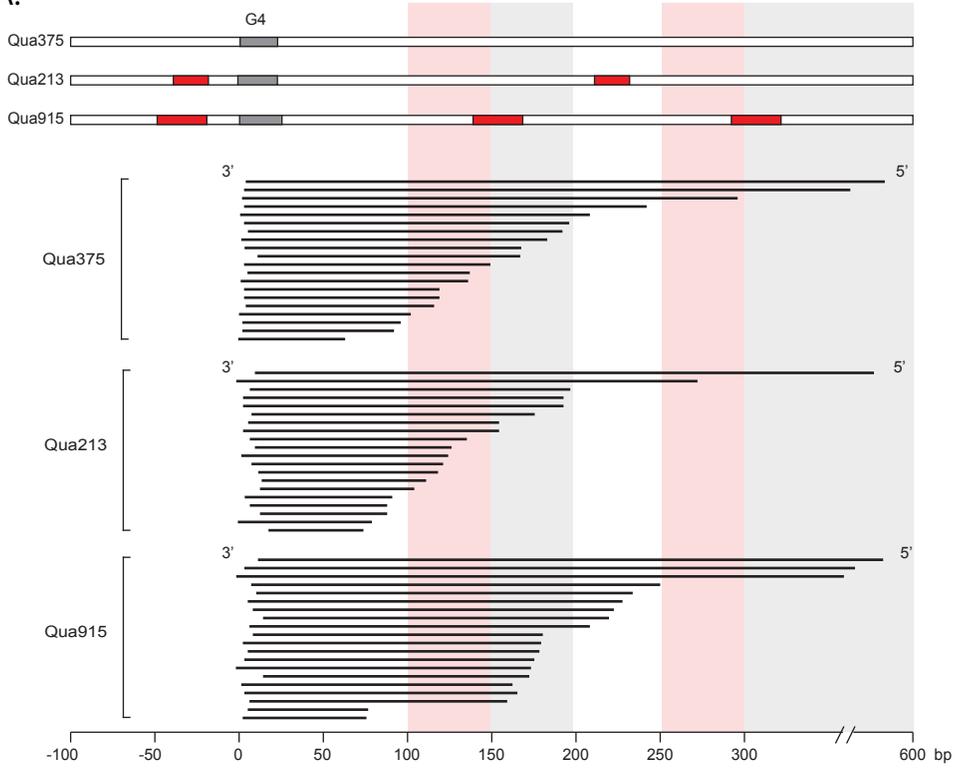
the following DSB ends will contain flanking homology and HDR will be the preferred mode of repair (Figure 4C, middle panel): resulting in a lack of homology-independent Qua915 deletions with 5' junctions upstream (left) of the repeat (Figure 4B, highlighted in pink). However, when the ssDNA gaps are larger and cover the homologous repeats, the DSB will not contain homologous sequences at both ends and TMEJ will be the preferred mode of repair (Figure 4C, right panel): indeed resulting in multiple homology-independent Qua915 deletions with 5' junctions downstream (right) of the repeat (Figure 4B, highlighted in grey).

This model also predicts that repeats that are very close to the G4 site are poor substrates for HDR, even if they were larger than the 29bp repeats at Qua915, given that the ssDNA gaps often would cover the homologous sequences and the following DSBs would not have homologous ends (Figure 4C, right panel). Indeed, the vast majority of pLM20 ORF corrections was still *polq-1*-dependent and did not use the relatively large 50bp repeat positioned just 24bp downstream of the G4 site (Figure 1F and S3). All together these data support a model in which homologous sequences that flank G4 sites can mediate DSB repair but only when they are of sufficient size and located such that they are present in the subsequent DSB ends; in all other cases POLQ-1 activity is required for deletion formation (Figure 4C).

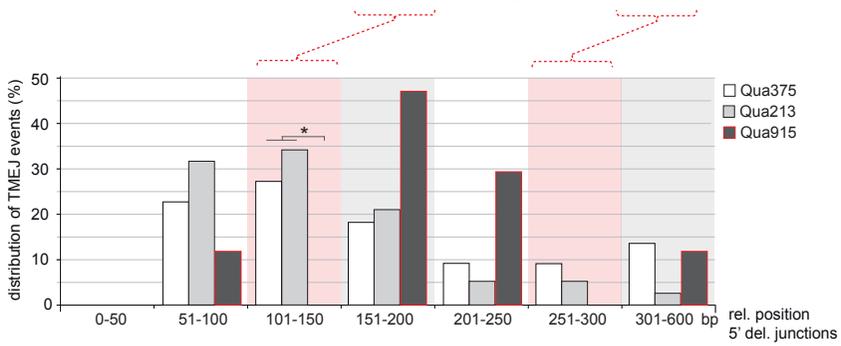
#### Figure 4. HDR can locally dominate repair of G4-induced DSBs →

**A.** Graphic illustration of G4 deletions profiles at Qua375 and Qua915 in *dog-1* deficient animals. Black bars represent homology-independent deletions. Genomic regions that display local dominance of HDR are highlighted in pink and reside directly left (5') of the major repeats at Qua915 (red blocks). Genomic regions that reside directly right (3') of the major repeats at Qua915 are highlighted in grey. **B.** Distribution of homology-independent deletions binned in 50bp windows based on the position of their 5' junctions relative to the G4 motif. Genomic regions devoid of homology-independent events at Qua915 are highlighted in pink and reside directly left (5') of the major repeats present at this locus. Genomic regions that reside directly right (3') of the major repeats at Qua915 are highlighted in grey. Asterisk indicates significant difference between the local deletion frequency at Qua915 and the two other G4 loci that lack the downstream repeat ( $p < 0.01$  by Fisher's exact test, two tailed). **C.** Different models for G4-induced deletion formation depending on the presence of flanking homology. Left model: scenario at G4 sites lacking flanking homology, typically resulting in TMEJ products. Middle model: scenario at G4 sites with flanking repeats that are still present in the following DSB ends, resulting in HDR dominance. Right model: scenario at G4 sites with flanking repeats of which one is lost in the following DSB ends, again resulting in TMEJ products.

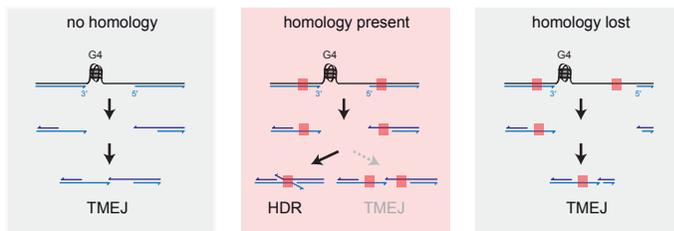
A.



B.



C.



## Discussion

Our genetic analysis revealed that homology-independent deletion events at G4 loci require TMEJ and that a subset of POLQ-1 substrates can be channeled into HDR provided that flanking homology is present. The local dominance of HDR at repetitive loci further illustrates the dynamic balance between TMEJ and HDR events and suggests that the presence of homologous repeats can direct repair pathway choice.

Based on these observations we propose a model in which unresolved G4 structures result in DSBs that can be repaired either via TMEJ or HDR depending on the sequence context (Figure 3 and 4C). The presence of homologous repeats at both DSB ends bypasses the need for POLQ-1 to create complementary 3' DNA ends and allows immediate annealing of the DSB ends: paving the way for TMEJ-independent deletion formation.

When flanking homology is absent at the DSB and the sister-chromatid still harbors a persistent G4 structure, the cell relies on TMEJ to adequately repair G4-induced DSBs (Figure 3). We propose that stalled replication at persistent G4 structures leads to 50-300bp ssDNA gaps that in the next S-phase result in replication-born DSBs that are neither compatible for NHEJ (potentially due to long 3' overhangs) nor have a suitable HR template (due to the persistent G4 structure) and thus require TMEJ or HDR for repair.

Replication-born DSBs are likely to be processed by DNA end resection nucleases in an attempt to initiate error-free HR repair via the sister-chromatid, given that such nucleases are activated during S-phase (FERRETTI *et al.* 2013; TRUONG *et al.* 2013). Moreover, previous studies detected increased RAD-51 foci (a marker for resected DSBs) in genetic backgrounds with increased TMEJ products, including animals deficient for the G4-resolving helicase DOG-1 (KOOLE *et al.* 2014; ROERINK *et al.* 2014). These studies also revealed that G4-induced DSBs lacking obvious flanking homology are processed extensively in TMEJ-deficient animals, ultimately resulting in elevated levels of RAD-51 foci, bi-directional deletions and extensive loss of genetic material (KOOLE *et al.* 2014; ROERINK *et al.* 2014). Repair via TMEJ or HDR would prevent continuous DSB processing and putative deleterious signaling events, promoting proper animal development at the expense of small deletions. The notion that G4-induced DSBs can be repaired via TMEJ or HDR, but not NHEJ, implies that these DSBs contain substantial 3' overhangs that allow annealing of homologous sequences but prevent binding of Ku (a dimeric protein complex that binds blunt DSB ends and initiates NHEJ). In the absence of repair via TMEJ, HDR or HR these 3' overhangs could be a substrate for DNA end resection nucleases, resulting in persistent RAD-51 foci and bi-directional deletions (KOOLE *et al.* 2014).

We propose that the key role of polymerase Theta is to create *de novo* complementary sequences at 3' DNA ends, allowing bridging of DSBs that lack flanking homology: a unique ability needed to repair DSBs genome-wide when canonical repair mechanisms such as HR and NHEJ are not feasible.

## Material and Methods

### Genetics

All strains were cultured according to standard *C. elegans* procedures (BRENNER 1974). Alleles used in this study include: *LGI*; *dog-1(gk10)*, *dog-1(pk2247)*, *fnci-1(tm3081)*, *LGIII*: *polq-1(tm2026)*, *exo-1(tm1842)*, *com-1(t1626)* *LGIV*: *fcd-2(tm1298)*, *mre-11(ok179)*, *LGX*; *pklS2170* [pRP1879], *LG unknown*; *lfls55* [pLM20].

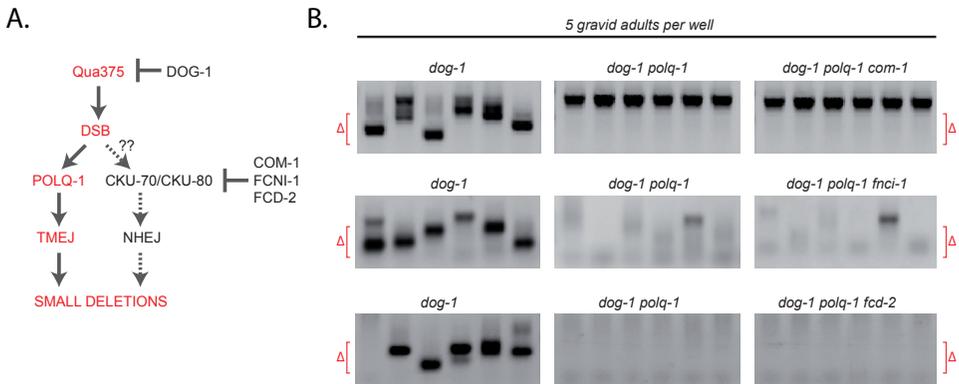
### LacZ-based transgenic reporter assay to visualize G4 deletions

Transgenic strains were obtained by microinjection of reporter construct pLM20 [*myo-2::C23::stops::NLS::LacZ*] and mCherry-based co-expression markers pGH8 pCFJ104 to generate *lfls55* (FROKJAER-JENSEN *et al.* 2008) or by microparticle bombardment of reporter construct pLM20 and *unc-119* expression marker to generate *pklS2170*. To visualize stochastic G4 deletions, >4 clonal lines of *dog-1* deficient *lfls55* or *pklS2170* animals were stained for LacZ expression per experiment (POTHOF *et al.* 2003).

### PCR-based assays to identify G4 deletions at endogenous loci

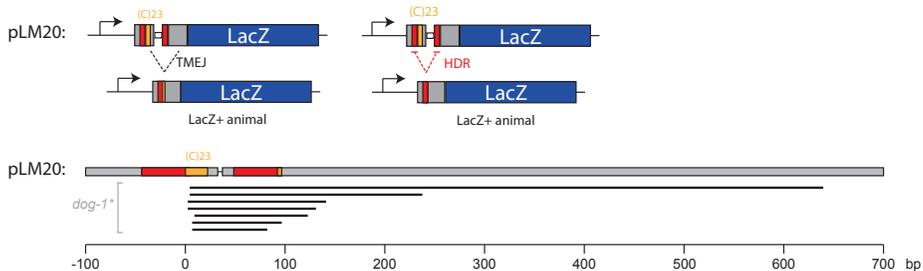
Stochastic deletion formation at endogenous G4 DNA loci was assayed using a PCR-based approach. Genomic DNA was isolated either from single worms or pools of worms and subjected to nested rounds of PCRs with primers that flank the G4 motif and flanking repeats if present (see Figure S1 for detailed sequence context). To obtain stochastic deletions frequencies at Qua375, Qua213 and Qua915, we analyzed two independent PCR reactions on 1% lysate fractions of >190 *dog-1* proficient and >330 *dog-1* deficient animals. L4 stage animals were used (1 worm per 10ul lysis reaction) and ~0.1ul lysate (1%) was transferred into 15ul PCR reactions using a 384 pin replicator (Genetix X5050). Subsequently 0.2ul of PCR product was used for 15ul nested PCR reactions. PCR reactions were typically run for 35 cycles with 54°C primer annealing and 72°C extension for 120 seconds. Unique deletion products were discriminated based on size by gel-electrophoreses and repair footprints were analyzed by Sanger sequencing. Strictly deletions verified by both PCR reactions were analyzed and used to determine the deletion frequency.





**Figure S2. Stochastic G4 deletions at Qua375 depend on TMEJ even in absence of NHEJ suppressors**

**A.** Model for G4-induced deletion formation and hypothetical suppression of NHEJ to act parallel to TMEJ  
**B.** PCR analysis of G4 instability at endogenous G4 site Qua375; each well represents an independent PCR reaction on 10% lysate of five gravid adults of the indicated genotype; size-range of PCR-amplified deletions products is indicated by  $\Delta$ . Upper panels: genomic lysates were obtained from first-generation *com-1* homozygous adults containing several embryo's to minimize the contribution of maternal COM-1. Middle and lower panels: PCR reactions were run with shorter extension times (1min) to hinder the formation of abundant wild-type products and enrich for G4 deletion products. In all cases, PCR conditions allowed efficient detection of multiple deletion events in *polq-1* proficient controls (left panels).



\* Out of 2000 independent populations of *dog-1* deficient pLM20 animals, 9 were identified that carried a germline LacZ ORF restoration event. Deletion footprint analysis revealed that 7 out of 7 tested populations contained non-homologous deletions resulting in an in-frame LacZ product. Although the sample size is small, these data suggest that only a minor fraction of repair events use the flanking repeat positioned very close to the G4 motif. This observation is in line with the substantial reduction in LacZ ORF correction events in TMEJ deficient pLM20 animals (Figure 1).

**Figure S3. Frequent non-homologous germline G4 deletions (TMEJ events) at pLM20 reporter locus**

Upper panels depict the pLM20 reporter construct and the anticipated TMEJ (left) and HDR (right) outcomes. Lower panel depicts germline G4 deletions profiles at the pLM20 reporter locus. \*Out of 2000 independent populations of *dog-1* deficient pLM20 animals, 9 were identified that carried a germline LacZ ORF restoration event. Deletion footprint analysis revealed that 7 out of 7 tested populations contained non-homologous deletions resulting in an in-frame LacZ product (black bars). Although the sample size is small, these data suggest that only a minor fraction of repair events use the flanking repeat positioned very close to the G4 motif. This observation is in line with the substantial reduction in LacZ ORF correction events in TMEJ deficient pLM20 animals (Figure 1).

### **Acknowledgments**

The authors thank Shohei Mitani (National Bioresource Project, Japan) and the Caenorhabditis Genetics Center for strains; Jane van Heteren for technical support and valuable comments on the manuscript

## References

- Adamo, A., S. J. Collis, C. A. Adelman, N. Silva, Z. Horejsi *et al.*, 2010 Preventing nonhomologous end joining suppresses DNA repair defects of Fanconi anemia. *Mol Cell* 39: 25-35.
- Brenner, S., 1974 The genetics of *Caenorhabditis elegans*. *Genetics* 77: 71-94.
- Budzowska, M., and R. Kanaar, 2009 Mechanisms of dealing with DNA damage-induced replication problems. *Cell Biochem Biophys* 53: 17-31.
- Chan, S. H., A. M. Yu and M. McVey, 2010 Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet* 6: e1001005.
- Deriano, L., and D. B. Roth, 2013 Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu Rev Genet* 47: 433-455.
- Ferretti, L. P., L. Lafranchi and A. A. Sartori, 2013 Controlling DNA-end resection: a new task for CDKs. *Front Genet* 4: 99.
- Fishman-Lobell, J., N. Rudin and J. E. Haber, 1992 Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Mol Cell Biol* 12: 1292-1303.
- Frokjaer-Jensen, C., M. W. Davis, C. E. Hopkins, B. J. Newman, J. M. Thummel *et al.*, 2008 Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat Genet* 40: 1375-1383.
- Han, H., L. H. Hurley and M. Salazar, 1999 A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res* 27: 537-542.
- Howell, R. M., K. J. Woodford, M. N. Weitzmann and K. Usdin, 1996 The chicken beta-globin gene promoter forms a novel "cinched" tetrahelical structure. *J Biol Chem* 271: 5208-5214.
- Huppert, J. L., 2010 Structure, location and interactions of G-quadruplexes. *FEBS J* 277: 3452-3458.
- Koole, W., R. van Schendel, A. E. Karambelas, J. T. van Heteren, K. L. Okihara *et al.*, 2014 A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat Commun* 5: 3216.
- Kruisselbrink, E., V. Guryev, K. Brouwer, D. B. Pontier, E. Cuppen *et al.*, 2008 Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C. elegans*. *Curr Biol* 18: 900-905.
- Lemmens, B. B., N. M. Johnson and M. Tijsterman, 2013 COM-1 promotes homologous recombination during *Caenorhabditis elegans* meiosis by antagonizing Ku-mediated non-homologous end joining. *PLoS Genet* 9: e1003276.
- Lieber, M. R., 2010 The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 79: 181-211.
- Ma, J. L., E. M. Kim, J. E. Haber and S. E. Lee, 2003 Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol Cell Biol* 23: 8820-8828.
- Murat, P., and S. Balasubramanian, 2014 Existence and consequences of G-quadruplex structures in DNA. *Curr Opin Genet Dev* 25C: 22-29.
- Pothof, J., G. van Haaften, K. Thijssen, R. S. Kamath, A. G. Fraser *et al.*, 2003 Identification of genes that protect the *C. elegans* genome against mutations by genome-wide RNAi. *Genes Dev* 17: 443-448.
- Preston, B. D., T. M. Albertson and A. J. Herr, 2010 DNA replication fidelity and cancer. *Semin Cancer Biol* 20: 281-293.

Roerink, S. F., R. van Schendel and M. Tijsterman, 2014 Polymerase theta-mediated end joining of replication-associated DNA breaks in *C. elegans*. *Genome Res* 24: 954-962.

San Filippo, J., P. Sung and H. Klein, 2008 Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* 77: 229-257.

Tarsounas, M., and M. Tijsterman, 2013 Genomes and G-quadruplexes: for better or for worse. *J Mol Biol* 425: 4782-4789.

Tourriere, H., and P. Pasero, 2007 Maintenance of fork integrity at damaged DNA and natural pause sites. *DNA Repair (Amst)* 6: 900-913.

Truong, L. N., Y. Li, L. Z. Shi, P. Y. Hwang, J. He *et al.*, 2013 Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc Natl Acad Sci U S A* 110: 7720-7725.

Youds, J. L., N. J. O'Neil and A. M. Rose, 2006 Homologous recombination is required for genome stability in the absence of DOG-1 in *Caenorhabditis elegans*. *Genetics* 173: 697-708.