Trends in Genetics

CellPress
REVIEWS

Review

Templated Insertions: A Smoking Gun for Polymerase Theta-Mediated End Joining

Joost Schimmel,[1] Robin van Schendel,[1] Johan T. den Dunnen,[1] and Marcel Tijsterman[1,*]

A recognized source of disease-causing genome alterations is erroneous repair of broken chromosomes, which can be executed by two distinct mechanisms: non-homologous end joining (NHEJ) and the recently discovered polymerase theta-mediated end joining (TMEJ) pathway. While TMEJ has previously been considered to act as an alternative mechanism backing up NHEJ, recent work points to a role for TMEJ in the repair of replication-associated DNA breaks that are excluded from repair through homologous recombination. Because of its mode of action, TMEJ is intrinsically mutagenic and sometimes leaves behind a recognizable genomic scar when joining chromosome break ends (i.e., 'templated insertions'). This review article focuses on the intriguing observation that this polymerase theta signature is frequently observed in disease alleles, arguing for a prominent role of this double-strand break repair pathway in genome diversification and disease-causing spontaneous mutagenesis in humans.

Highlights

DNA rearrangements that cause human pathologies frequently display small duplications of neighboring sequences.

Templated insertions represent an evolutionarily conserved polymerase theta-mediated end joining (TMEJ)-specific mutational signature.

Polymerase TMEJ has emerged as a stand-alone, intrinsically mutagenic repair pathway that could serve as a back-up but also acts on chromosomal breaks that are refractory to non-homologous end joining and homologous recombination pathways.

Cells of certain types of tumors rely on TMEJ for their survival.

Three Instead of Two Pathways Acting on Chromosomal Breaks

Faithful transmission of genetic information to daughter cells can be threatened by replication-impeding DNA damage that results from both environmental as well as metabolic sources. Although cells use a wide variety of DNA repair mechanisms to maintain genome integrity, DNA damage can result in permanent changes in the nucleotide sequence (i.e., DNA mutations). Mutations are causative to human disease; if inflicted in germ cells, they become heritable and can result in genetic disorders. In somatic cells the accumulation of mutations drives cancer, a process that can be accelerated by germline mutations in cancer predisposition genes. Specific types of DNA mutations found in human disease [especially deletions and **translocations** (see Glossary)] are attributed to the erroneous repair of DNA **double-strand breaks (DSBs)** that can be introduced by obstruction of DNA replication, by reactive oxygen species, and by exposure to chemicals or radiation [1–3].

In the predominant view, repair of DSBs is categorized into two distinct pathways: **non-homologous end joining (NHEJ)** and **homologous recombination (HR)**. The primary difference between these two pathways is that HR requires an undamaged template, generally the sister chromatid, to guide error-free repair, while NHEJ is template independent and thus able to repair DSBs in situations where there is no undamaged copy available. NHEJ is a protein-guided pathway: DSB end binding and protection by the heterodimer Ku70-Ku80 as well as their ligation by the XRCC4-DNA ligase IV complex is sequence independent. In contrast, HR is sequence guided: to create a priming end that copies the template, HR commences by generating stretches of 3′-overhanging single-stranded DNA (ssDNA). **End resection** is mainly executed by the nuclease activity of the Mre11, Rad50, and Nbs1 complex (**MRN complex**), which is stimulated by CtIP. Resulting ssDNA overhangs are first bound by **replication protein A (RPA)** but then replaced through BRCA2 action by the recombinase protein RAD51, forming a RAD51-ssDNA nucleoprotein filament. This filament initiates invasion of an undamaged homologous copy that will serve as the repair template [4,5].

[1]Department of Human Genetics, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, The Netherlands

*Correspondence:
m.tijsterman@lumc.nl (M. Tijsterman).

Mechanisms that determine use of HR versus NHEJ are becoming increasingly clear. First and foremost, pre- and post-replicative specific chromatin-marks direct repair-pathway usage by serving as docking sites for pathway-specific repair factors that regulate end resection, a process that is also under cell cycle control [6–14] (reviewed in [5] and [15]). In addition, binding of Ku70-Ku80 to both ends of a chromosomal break restricts end resection [16]. While HR repairs DSBs with high fidelity, NHEJ can be both error free and error prone. The accuracy of NHEJ greatly depends on the architecture of the DSB: at 'clean' blunt breaks or breaks with complementary (sticky) ends, NHEJ acts error-free through binding of Ku70-Ku80 and direct ligation by XRCC4-DNA ligase IV. At incompatible ends, however, additional processing by nucleases and **DNA polymerases** is required, which results in the deletion and/or addition of a few nucleotides [17,18]. It is presently unknown what fraction of the physiologically relevant DSBs that use NHEJ require such polishing. Thus, although NHEJ is template independent it can function as a highly accurate error-free repair mechanism preventing loss of genetic information. This is in stark contrast to what is observed at many mutations in human disease, where deletions can be large, occasionally contain insertions of *de novo* DNA, and frequently display **microhomology** at the deletion junctions, indicating that sequence information was used to guide repair. These features imply the existence of an alternative DSB repair pathway that is responsible for a yet undefined fraction of disease alleles.

Discovery of an Intrinsically Mutagenic DSB Repair Pathway

The notion of a third pathway acting on chromosomal breaks first became apparent two decades ago, when residual end-joining activity was observed in NHEJ deficient yeast and hamster cells [19,20]. Compared to NHEJ-proficient controls, repair products recovered from these backgrounds had lost more DNA and repair seemed to rely on small stretches of repetitive sequences located up- and downstream of the break sites (so-called microhomology). Similar findings emerged from research on DNA translocations in murine cells, demonstrating that the frequency of translocation formation actually increases in the absence of NHEJ factors, and these joints are also characterized by microhomology [21–24]. These studies, together with biochemical work [25], provided the first insight into which factors facilitated non-canonical end joining, as translocation formation was greatly reduced in the absence of CtIP, Lig3, and PARP1. Robust end-joining activity has also been observed in human cells compromised for NHEJ [26,27]. Two key features, non-reliance on classical NHEJ factors and microhomology at the repair junction, inspired the usage of the terms **alternative end joining (Alt-EJ)** and microhomology-mediated end joining (MMEJ) to describe this end-joining activity. Recently, this pathway gained tremendous impetus by the identification of a specialized DNA polymerase, **DNA polymerase theta** (Polϴ, encoded by the *Polq* gene), which is essential for various repair activities attributed to Alt-EJ; and by the discovery that this polymerase is upregulated in several human cancers [28,29] and can be crucial for their viability [30,31].

The first indication that Polϴ acts in DSB repair came from a phenotype-driven mutagenesis screen in mice, where a mutation in *Polq* was linked to increased numbers of spontaneous and DNA-damage-induced **micronuclei** [32] in cells from the hematopoietic lineage. Studies on *mus308*, the *Drosophila melanogaster* homolog of *Polq*, revealed a role in DSB repair independent of NHEJ, and placed Polϴ in a form of Alt-EJ that was called synthesis-dependent MMEJ [33,34]. Subsequently, Polϴ was found to be essential for the majority of Alt-EJ/MMEJ in roundworms, zebrafish, and plants and in mouse and human cells [30,31,35–41]. The term **theta-mediated end joining (TMEJ)** was coined to more precisely define a pathway having distinct genetic requirements, as opposed to the more comprehensive term Alt-EJ that may encompass additional biology. In addition, the worm and plant work revealed that in certain biological contexts TMEJ is the only option to repair DSBs, thus in fact not acting as an 'alternative' or back-up pathway.

Glossary

Alternative end joining (Alt-EJ): umbrella term for end-joining pathways that do not rely on NHEJ.
BRCAness: term to describe tumors that show similar features to BRCA-deficient tumors.
Chromothripsis: phenomenon where a single catastrophic event results in tens to thousands of chromosomal rearrangements within one chromosome.
ClinVar database: public archive of reports of the relationship among human variations and phenotypes.
Complex genomic rearrangements (CGRs): structural variations that are caused by more than two chromosome breaks resulting in exchanges of chromosomal segments.
COSMIC database: Catalogue Of Somatic Mutations In Cancer, online database of somatic mutations found in human cancer.
CRISPR/Cas9: Cas9 is an endonuclease enzyme that recognizes and cleaves specific strands of DNA that are complementary to clustered regularly interspaced short palindromic repeat (CRISPR) sequences; it is an antiviral defense system of prokaryotes, now used as a gene-editing technique in many species.
DNA helicase domain: domain conserved in helicase enzymes, which regulates the separation of annealed DNA strands by using energy derived from ATP hydrolysis.
DNA polymerases: enzymes that synthesize DNA molecules by adding deoxyribonucleotides to the 3′ end of a DNA strand.
DNA polymerase theta (Polϴ): A-family polymerase encoded by the *Polq* gene; key factor in TM-EJ of DSBs thereby relying on microhomology.
Double-strand breaks (DSBs): physical interruption of the DNA double helix due to nearby lesions of the phosphodiester backbone in opposing complementary DNA strands.
End resection: nucleolytic degradation of the 5′ strands at DSBs to expose 3′ ssDNA stretches.
Homologous recombination (HR): error-free DSB repair pathway that utilizes the undamaged sister chromatid as a template to guide repair.
Illegitimate recombination: untargeted random integration of exogenous DNA in a genome, not relying on sequence homology.

CellPress
REVIEWS

Key Figure

Conceptual Model for Separation of Function in DNA Double-Strand Break Repair

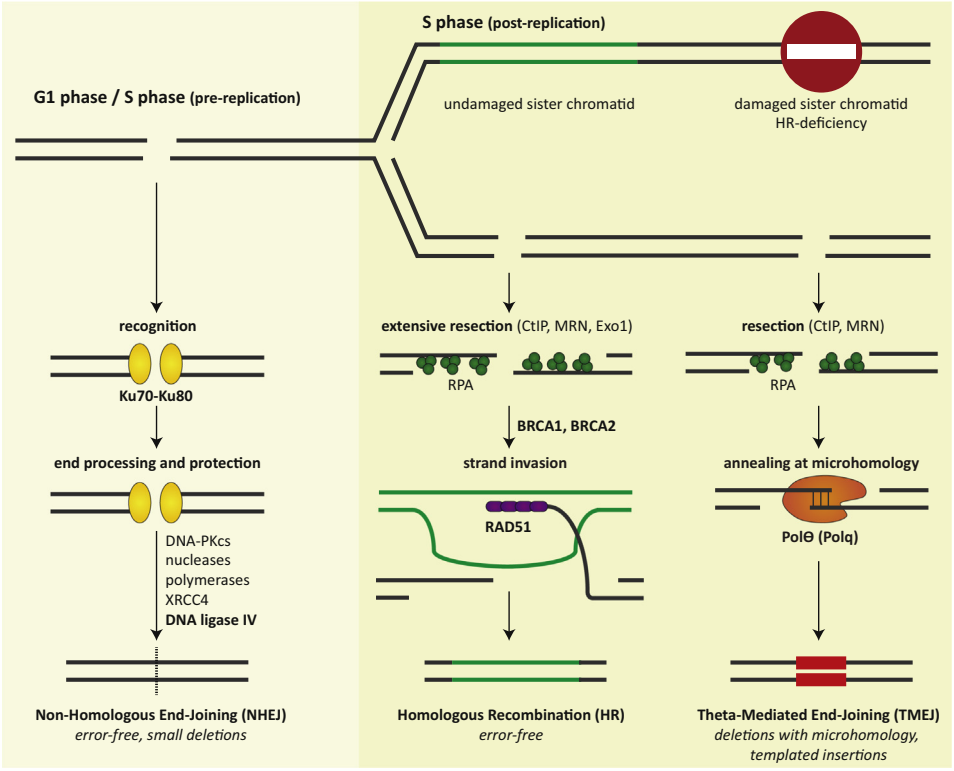

Figure 1. Breaks that occur pre-replication (G1, early S phase) are predominantly repaired by non-homologous end joining (NHEJ), in which no or minimal processing results in either error-free repair or in small deletions, respectively. Breaks that occur post-replication are preferentially repaired by homologous recombination (HR), where end resection exposes long stretches of single-stranded DNA that are protected by replication protein A (RPA) coating, which is subsequently exchanged for RAD51 to facilitate strand invasion of the sister chromatid. This undamaged template is then used as a blueprint to guide error-free repair. Theta-mediated end joining (TMEJ) may be particularly vital to repair DNA double-strand breaks that have been resected but for which a template for error-free repair is unavailable, for example because of a damaged sister chromatid, or because of a genetic defect in HR. DNA polymerase theta (Polθ, encoded by the *Polq* gene) uses annealed microhomologous sequences in the immediate vicinity of the break ends as a starting point for DNA extension. This biochemical requirement makes this pathway error prone by default, resulting in deletions with microhomology and templated insertions.

TMEJ's *Raison d'être*

Given the accuracy of NHEJ and the fidelity of HR, one could wonder why cells have evolved TMEJ, a highly mutagenic DSB repair pathway. *De novo* rearrangements in mammalian genomes frequently display hallmarks of TMEJ (see Templated Insertions in Human Disease Alleles), arguing that a sufficient number of DSBs are produced during the lifetime of an organism that require this route of repair. This leads to the following question: In which contexts are NHEJ and HR insufficient? Or, in other words, what types of DSBs cannot be repaired by the two most well-studied and evolutionarily conserved pathways?

Immunoglobulin class switch recombination (CSR): break-induced genomic rearrangement in the heavy chain locus of B cells needed for antibody diversification.
Microhomology: the presence of two short stretches of identical sequences in a given piece of DNA.
Micronuclei: small nucleus-like structures that form in addition to two daughter nuclei, containing chromosome fragments as the result of DNA breaks.
MRN complex: protein complex (consisting of Mre11, Rad50, and Nbs1) that initiates resection of DSBs.
Non-homologous end joining (NHEJ): DSB repair pathway that directly joins two ends by ligation without the use of a repair template or homology; depending on the break NHEJ can be error-free or result in small deletions at the break site.
Polymerase domain: hand shaped structure conserved in DNA polymerase enzymes, where the 'palm' domain provides the catalytic active site, the 'fingers' regulate the positioning of the template, and the 'thumb' is important for processivity.
Replication protein A (RPA): a heterotrimeric protein that binds to (3') ssDNA to maintain its unwound state and protect the DNA against endonucleases.
S phase: synthesis phase, term to describe the period in the cell cycle during which DNA replication takes place.
T-DNA: transfer-DNA, a discrete segment of *Agrobacterium tumefaciens*' DNA, which is integrated into the host plant's genome upon infection.
Templated insertions: small DNA insertions found at DSB-induced mutagenesis that are synthesized by Polθ by using nearby DNA as a template.
Theta-mediated end joining (TMEJ): intrinsic mutagenic DSB repair pathway in which Polθ uses microhomology in opposing 3' ssDNA to anneal and repair chromosomal breaks.
Translocation: chromosomal rearrangement where chromosomal regions of two non-homologous chromosomes are joined.

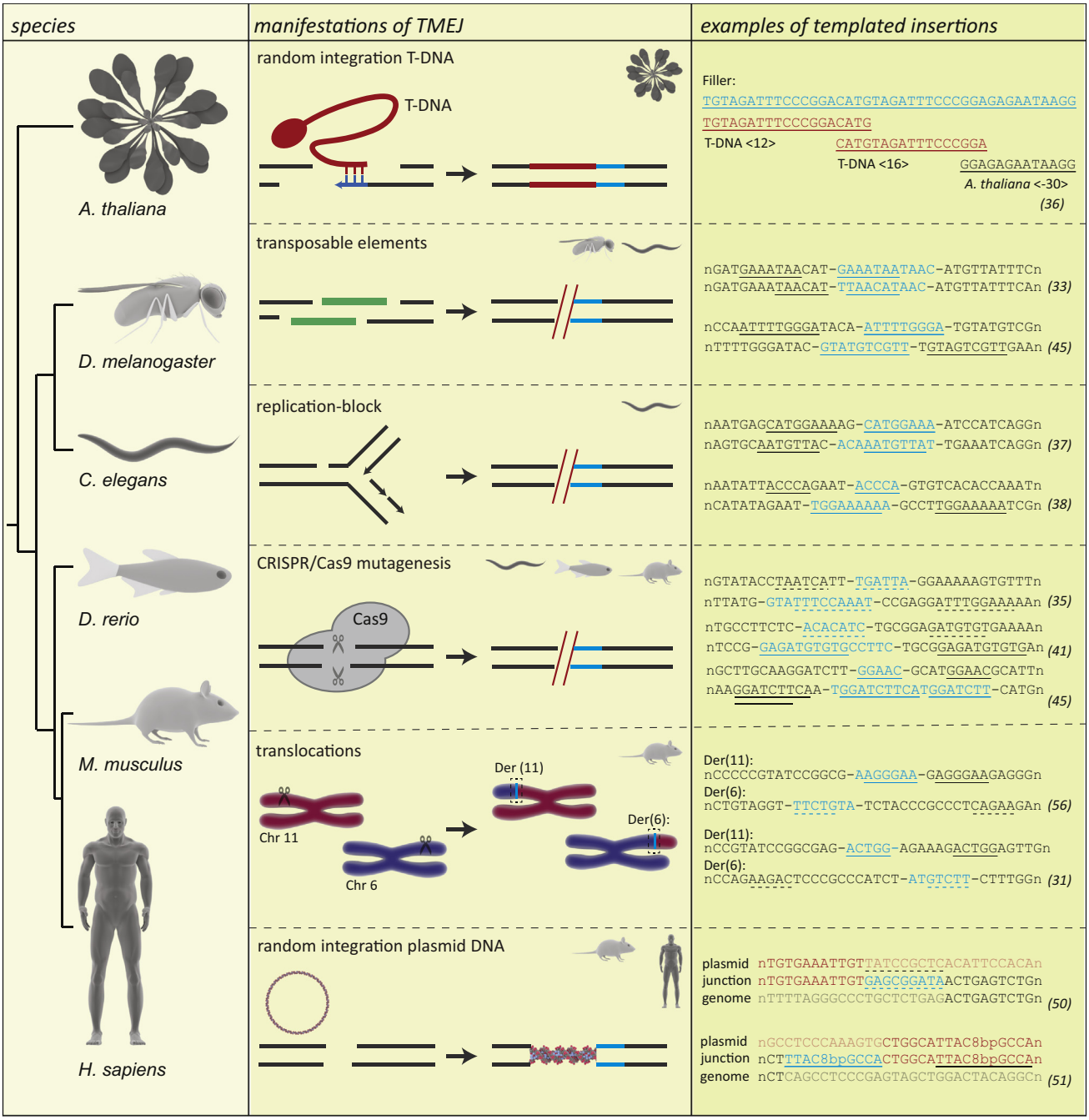

Figure 2. Templated Insertions: Evolutionary Conserved Hallmark of Theta-Mediated End Joining. Summary of theta-mediated end-joining (TMEJ) manifestations and examples of associated templated insertions in *Arabidopsis thaliana* [36], *Drosophila melanogaster* [33], *Caenorhabditis elegans* [35,37,38,45], *Danio rerio* [41], *Mus musculus* [31,45,50], and *Homo sapiens* [51] (first column). Only double-strand break repair phenomena were included in which the involvement of polymerase theta in the formation of templated insertions has been demonstrated. The second column depicts the biological context for which TMEJ activity has been observed in different species (indicated by the small pictograms). The third column illustrates typical examples of TMEJ-dependent templated insertions. Original sequences are shown in black; inserted *de novo* DNA sequences appear in blue. Identical sequences in the insertions and in the flanking sequence are underlined (straight lines for templated insertions with a forward orientation, and dotted lines for templated insertions with a reverse complement orientation). In cases of random

(Figure legend continued at the bottom of the next page.)

One important hallmark of TMEJ outcomes provide clues as to which DSBs specifically require this pathway: microhomology at the junction argues for ssDNA tails to bestow annealing of complementary sequence, thus pointing to the requirement of (at least minimal) end resection. TMEJ may thus mainly act when resection at DSBs is activated in **S phase**. In support, Polθ has been shown to create deletions at DSBs resulting from persistent replication fork-blocking structures or lesions [37,38,42,43]. Also, HR and a mutagenic TMEJ-type of repair pathway have been shown to share the initial step of end resection [44]. In fact, HR-deficient tumor-cells isolated from patients with a BRCA mutation turned out to rely on TMEJ for proliferation and survival [30,31].

TMEJ can thus be considered an alternative to HR, providing a framework for three conceptually different DSB repair pathways operating in cells (Figure 1, Key Figure): (i) NHEJ repairing DSBs in a cellular context where no sister chromatid is available, (ii) HR repairing DSBs that occur post or during replication using an undamaged sister-chromatid as template, and (iii) TMEJ repairing HR intermediates (DSBs that are resected and thus are disfavored substrates for NHEJ) in cases where the sister chromatid cannot serve as a template. Such a pathway will become important in cases where both sisters have sustained a break (e.g., by DNA transposition, ionizing radiation, or persistent replication blocks) or when one of the other two pathways (HR or NHEJ) is genetically compromised. This conceptual model is supported by a more prominent role of TMEJ in fast-cycling embryonic and germ cells, by the observation that Polθ expression is upregulated in a wide range of rapidly dividing human cancer cell lines [29,35,41,45,46], and by the finding that TMEJ ensures cellular viability in NHEJ- and HR-compromised cells. Whether TMEJ also acts in nondividing cells on replication-independent DSBs is not yet clear, but the observation that microhomology usage during end joining is markedly reduced in senescent cells argues against this possibility [47,48].

Templated Insertions Are an Evolutionary Conserved TMEJ-Specific Mutational Signature

TMEJ-Dependent Templated Insertions in Different Species

The original work on *Drosophila* Polθ [33] presented two distinct mutational outcomes: (i) simple deletions, frequently containing microhomology at the junction; and (ii) more complex deletions that contained insertions of *de novo* DNA. The sequence of these insertions was frequently at least in part identical to sequences flanking the original break site, arguing that the latter had served as a template for DNA synthesis. A substantial body of work in a range of genetic systems and using different sources of DSBs has since firmly established a causal relationship between these **templated insertions** and Polθ and also established this repair outcome as a TMEJ-specific mutational signature (Figure 2). TMEJ proved to be a major driver of mutagenesis and genome diversification in *Caenorhabditis elegans;* templated insertions were abundantly present in different natural isolates, and their frequency increased in worms that were propagated for many generations in the laboratory or in animals that were exposed to mutagens [38,45,49].

In mouse cells, TMEJ-dependent insertions were first observed in the DNA joints produced by **immunoglobulin class switch recombination (CSR)** [39]. Junctions in B cells have >1 bp insertions in ~10% of the events, yet none were present in B cells lacking Polθ expression. The larger insertions (approximately >10 bp) were identical to sequences present relatively far (2–5 kbp) from the

integration of T-DNA and plasmid DNA, templated insertions find their origin in exogenous DNA (in red) or in genomic DNA. The example for T-DNA integration shows a large insert (filler) present at a T-DNA–plant genome junction, the sequence identical to different flanking parts of the T-DNA and genomic ends are depicted underneath the filler (T-DNA in red, *Arabidopsis thaliana* genome in black; see [36] for details). The context of the illustrated examples can be found in the indicated literature (in parentheses). Abbreviation: CRISPR, clustered regularly interspaced short palindromic repeats. See also [56].

junction site, arguing that more distal sites can also be used as a template for Polθ action. Additionally, Polθ was found to be responsible for nucleotide insertions at unprotected telomeres in mouse embryonic fibroblasts (MEFs) and at translocation junctions in induced pluripotent stem cells derived from those MEFs [31]; at **CRISPR/Cas9**-induced DSBs, templated insertions were found in both mouse embryonic stem (mES) cells and MEFs, but were absent in *Polq* knockout cells [35,40].

Another demonstration of TMEJ-induced templated insertions is at sites where foreign DNA integrates at random locations into a host genome, a phenomenon also known as **illegitimate recombination**. In the plant *Arabidopsis thaliana*, **T-DNA** integration completely relies on Polθ functionality; in mES cells, TMEJ is the dominant pathway for random integration (but NHEJ also contributes); and in human cells, integration of exogenous DNA becomes more reliant on TMEJ when NHEJ is compromised [36,50,51]. In all cases, templated insertions are frequently observed at the borders of integration but only in Polθ-proficient conditions. Together, these studies have provided evidence that templated insertions are an evolutionary conserved outcome of DSB repair, acting as a 'smoking-gun' for TMEJ activity.

Modus operandi of DNA Polθ in End-Joining Repair

Ever since its identification [52,53], Polθ has also been linked to DNA damage repair pathways other than DSB repair [54]. However, because templated insertions have only been observed at TMEJ of DSBs, here we only discuss Polθ's mode of action in this pathway. Polθ is unique in that it is the only eukaryotic polymerase that contains a superfamily 2 **DNA helicase domain** in addition to an A-family type **polymerase domain**. While both domains are evolutionarily well conserved, an intervening large central domain has diverged among species and its function remains to be elucidated [55]. Important insight into how Polθ operates during DSB repair comes from structural and biochemical analysis of the C-terminal polymerase domain [56–59]. The thumb, palm, and finger subdomains of Polθ induce a canonical right-hand configuration, common for DNA polymerases, which is needed for DNA-binding and template recognition [59]. Interestingly, and most prominently in vertebrates, Polθ contains three extra insertion loops in the polymerase domain, the function of which remain elusive. By using DNA substrates modeled after partially resected DSBs, it was shown that the purified human Polθ polymerase domain (hereafter referred to as Polθ-polymerase) performs end joining of 3′ ssDNA overhangs [57]. On these substrates, Polθ requires ≥2 bp of homology located at or near the end of 3′ overhangs to form a DNA synapse. The opposing overhangs are then used as a template for DNA-synthesis [57–59]. Similar to NHEJ polymerases, purified Polθ-polymerase also shows non-templated terminal transferase, potentially increasing versatility by generating additional options to find microhomologous sequences [56,60]. Proof for such activity *in vivo*, however, is still missing. *In vitro*, Polθ-polymerase is most active on small 3′ overhangs, but the *in vivo* configuration of break ends with respect to the degree of resection and the occupation of ssDNA binding proteins is not understood. Cell-based assays revealed potent TMEJ on episomal DNA substrates with 3′ overhangs of more than 45 nucleotides [40]; however, in cases where TMEJ acts on HR intermediates, end resection may be much more extensive. The notion that most templated insertions are derived from sequences in very close proximity of the DSB end (pointing towards limited availability of nearby sequence to act as a template for DNA synthesis) may be explained by stereo-chemical inhibition of ssDNA binding proteins or RAD51 coating-resected DNA.

Mutational signatures derived from biochemical and cellular experiments revealed that templated insertions not only come from the opposing strand on the other side of the break (in *trans*), but also can be formed via a snap-back mechanism of the protruding ssDNA itself (in *cis*), resulting in a templated insertion with a reverse complement orientation (Figures 3 and 4) [35,49,56]. It can, however, not be excluded that a broken sister chromatid has served as a template for insertions of reverse compliment orientation (Figure 3).

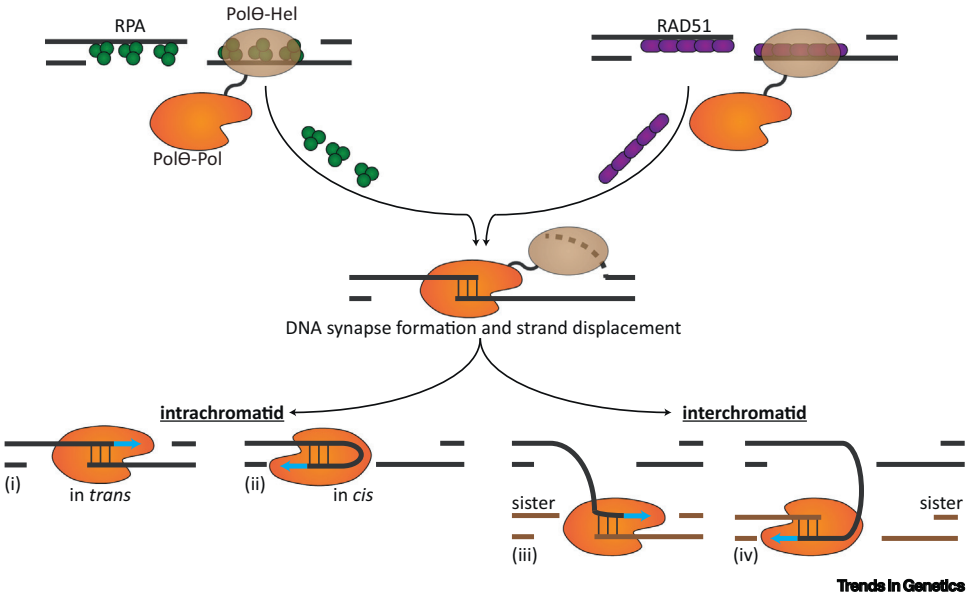

Figure 3. A Molecular Model for Theta-Mediated End Joining. A tentative model for theta-mediated end joining. The N-terminal helicase domain of polymerase theta (PolΘ) is suggested to displace replication protein A (RPA) and/or RAD51 molecules from 3′ single-stranded DNA to facilitate DNA synapsis at microhomologous sequences. Subsequently, PolΘ-polymerase (PolΘ-Pol) extends one end of the break by using the opposing strand of the other break end as a template, potentially facilitated by the strand displacement activity of PolΘ-helicase (PolΘ-Hel). In situations where there is no sister chromatid available, two substrates can potentially serve as a template for extension: (i) the opposing strand of the same chromatid (in *trans*) and (ii) upstream sequences of the same molecule (in *cis*). In situations where both chromatids are broken (e.g., in CRISPR-induced or transposon-induced DNA double-strand break formation), two additional substrates need consideration: (iii) the opposing strand and (iv) the identical strand of the other chromatid. Scenarios (ii) and (iv) reflect different biochemistry but can both explain templated insertions of reverse complement orientation.

While PolΘ polymerase is sufficient to perform end joining *in vitro*, several studies have pointed towards an important role for the N-terminal helicase domain in TMEJ. For example, this domain is able to displace RAD51 as well as RPA from ssDNA, which may be needed to expose microhomologous sequences that promote annealing and/or polymerase action (Figure 3) [30,61]. Such activity also supports the notion that TMEJ can act on HR-intermediates. In mouse cells, mutations in the ATP-binding sites of PolΘ-helicase disrupt TMEJ activity on chromosomal translocations and random integration [50,61]; templated insertions are still observed at translocation junctions albeit at 3-fold reduced levels [61]. Flies with a mutation of a highly conserved proline in the N terminus display increased frequencies of templated insertions, while ATPase-dead flies show reduced numbers of templated insertions [33,62]. It thus appears that the conserved N terminus directly affects the biochemistry of templated insertion formation, but it is not yet clear how [33,62].

These studies, together with detailed analysis of *in vivo* repair products, provided a model that explains templated insertions as the result of aborted template-dependent extension by PolΘ, followed by reannealing at secondary, perhaps more stable, homologous sequences (which includes template-switching) and a second step of extension (Figure 4). What makes TMEJ abort once PolΘ has engaged in DNA synthesis is an intriguing subject for future research; is it a polymerase error, a damaged template, mechanical forces, or just low processivity (see Outstanding Questions)?

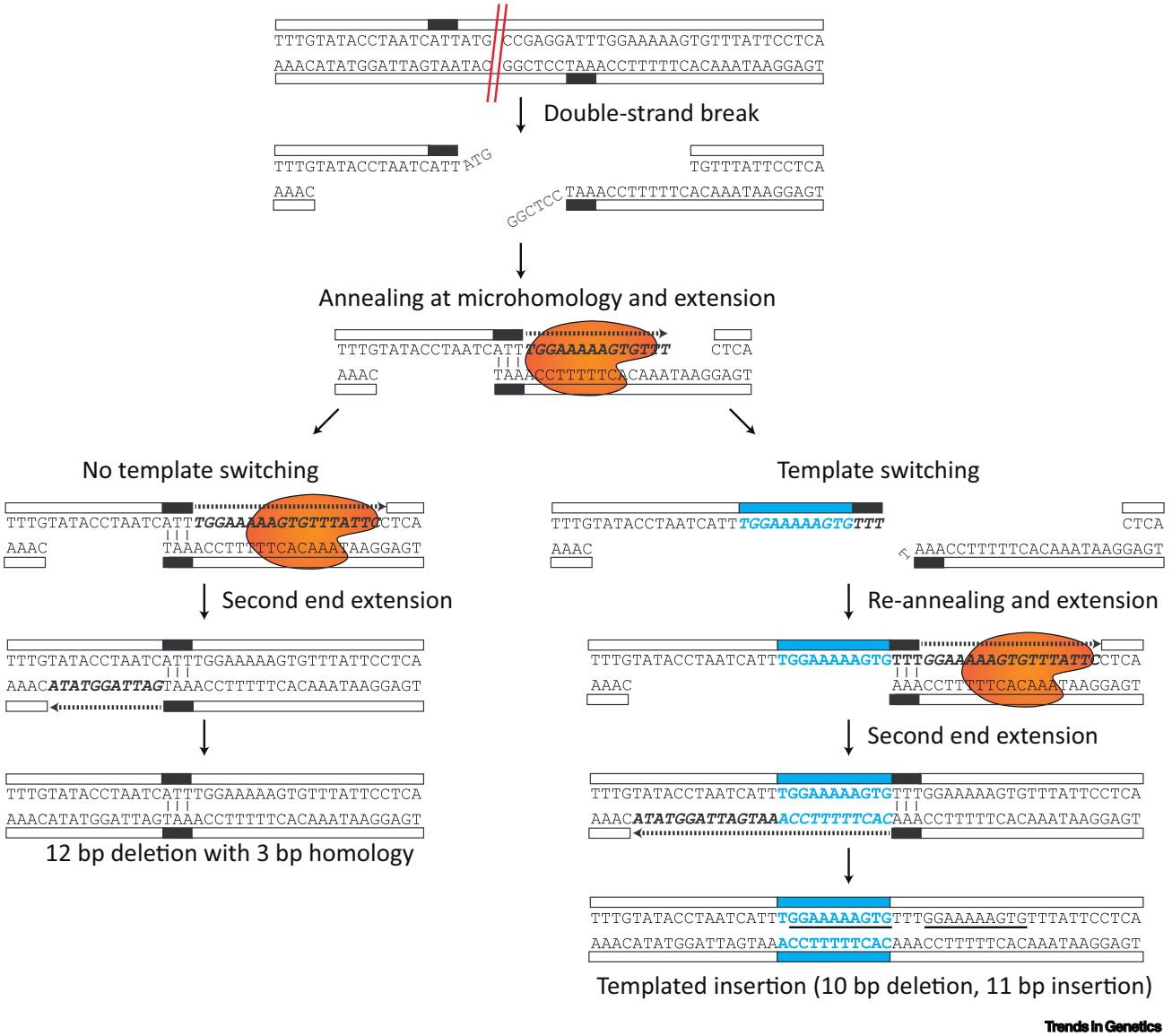

Figure 4. Schematic Illustration to Explain the Two Typical Mutational Signatures of Theta-Mediated End Joining. Annealing at microhomologous sequences followed by polymerase theta-polymerase-mediated extension results in a deletion hallmarked by microhomology (left). Primer template switching to another microhomologous sequence, induced by yet-unknown sources, results in a typical theta-mediated end-joining-specific templated insertion (right). Multiple iterations can result in more complex insertions.

Templated Insertions in Human Disease Alleles

Research in the fields of DNA repair and mutagenesis has led to increased insight into the mechanisms that are responsible for maintaining genetic integrity. However, which processes drive *de novo* mutations and thus contribute to inherited diseases are still largely unclear. Assigning DNA break-induced mutagenesis to a specific mechanism is further complicated by the fact that the outcomes of NHEJ and TMEJ, acting concurrently in mammalian cells, can resemble each other greatly [35,40]. As described in this review article, templated insertions form an exception as these constitute a clearly recognizable and evolutionary conserved mutational signature of TMEJ, different from more complex rearrangements attributed to the

DNA replication-based mechanisms fork stalling and template switching or break-induced replication, and different from minute (non)templated insertions coming from NHEJ [63,64]. The presence of this signature in mammalian genomes serves as a smoking gun for repair of DNA breaks mediated by Polθ.

Templated Insertions in Human Variations

The increased abundancy of TMEJ in cell culture experiments when NHEJ is disturbed has led to the idea that this pathway mainly acts as a back-up to an error-prone NHEJ pathway. However, in *C. elegans* it is TMEJ, not NHEJ, that underlies genomic diversification and spontaneous mutagenesis. The compelling question thus arises: Is this also true for humans? Is this intrinsically mutagenic route to repair broken chromosomes responsible for genomic alterations observed between individuals and causative to disease alleles? Using templated insertions as a TMEJ fingerprint now allows us to address these questions. Indeed, Figure 5 illustrates an abundant presence of templated insertions in alleles that are collected in the **ClinVar database**, a public archive containing human variations and associated phenotypes [65]. About 14% of 25,000 deletions contain an insertion (delins). Of the delins that are of sufficient size to query their origins, about 60% represent typical TMEJ-cases (Figure 5A). Inspecting their distribution reveals that templated insertions are found throughout the human genome affecting all types of disease alleles, hence contributing to a wide variety of human diseases, including familial breast cancer, Lynch syndrome, Cornelia de Lange syndrome, and Ehler–Danlos syndrome (Figure 5B). Apart from appearing in relatively simple deletion alleles, templated insertions are also found at junctions

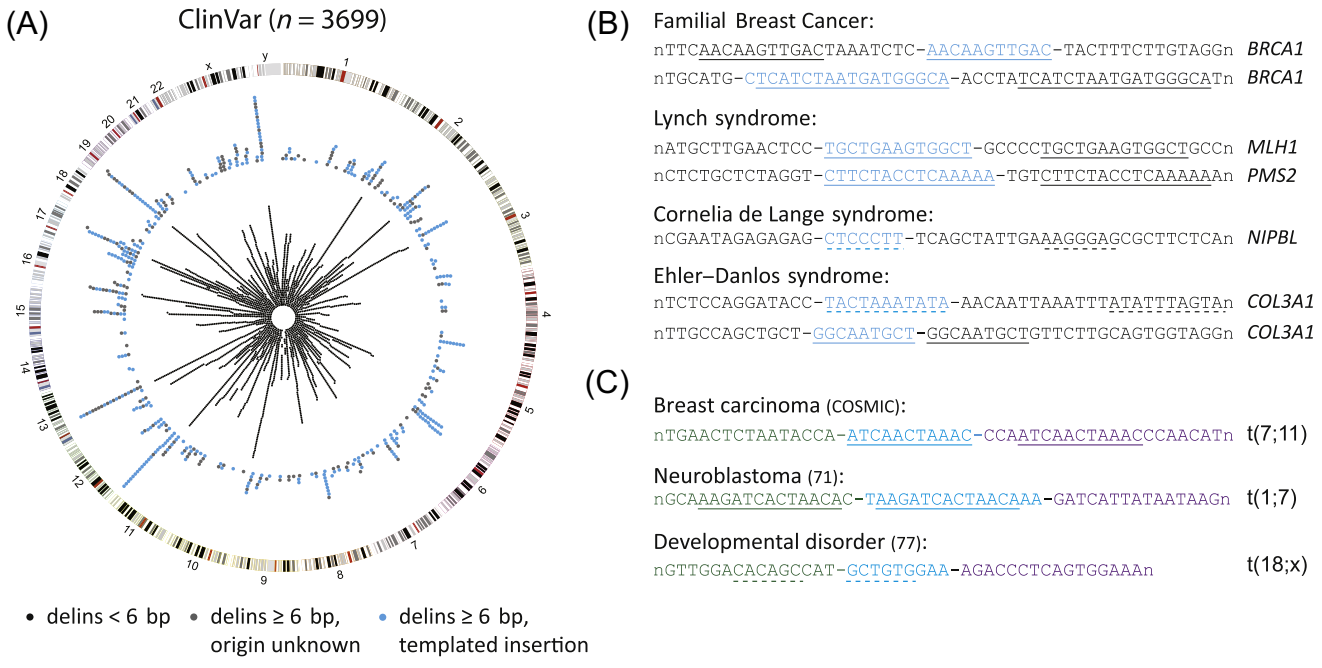

Figure 5. Theta-Mediated End-Joining Signatures in Human Disease Alleles. (A) Circos plot of deletions accompanied by insertions (delins) found in the ClinVar database. Black dots represent delins with insert too small (<6 bp) to reliable determine their origin; gray dots represent delins with insert ≥6 bp but whose origins are not found in close proximity of the event; blue dots represent delins that are (partly) identical to sequences located in close proximity. (B and C) Examples of templated insertions observed in the ClinVar database (B) and in translocation data (C). Original sequences are in black, inserted *de novo*-generated DNA sequence in blue. Identical sequences in the insertions and in the flanking sequence are underlined (straight lines for templated insertions with a forward orientation, dotted lines for templated insertions with a reverse complement orientation). The human condition that is associated with these disease alleles is indicated, as well as the cognate genes (B) and translocations (C). Translocated chromosomes have different colors (green and purple).

of more complex genomic rearrangement that are either disease causing or neutral [66–68]. As disease alleles only represent a small subset of genomic alterations, it is obvious that TMEJ not only contributes to the onset of human pathology but also to genetic diversity.

The total contribution of TMEJ to genetic variation can only be very roughly estimated because templated insertions only represent the tip of the iceberg of TMEJ-activity; simple deletions with (but also without) microhomology at the junction are the predominant outcome of TMEJ. In fact, only 4–8% of TMEJ-exclusive repair events in mES cells contain a templated insertion [35]. Extrapolating a similar outcome profile for human germ and/or embryonic cells predicts a profound contribution of TMEJ (~18%) in the etiology of disease alleles and genome rearrangements in the human population.

Templated Insertions at Translocation Junctions and in Complex Rearrangements

End joining of breaks that occur concurrently in the genome can result in translocations and other **complex genomic rearrangements (CGRs)**. As mentioned, DNA translocations are often characterized by microhomology at their junctions. While this is highly suggestive of TMEJ involvement, attributing outcomes to a specific mode of end joining cannot be made on a case-by-case basis; NHEJ can also result in deletions with microhomology at the junction and is known to contribute to translocation formation in human cells [17,69,70]. TMEJ-specific templated insertions, however, can be abundantly found at translocation junctions that are archived in the **COSMIC database** and also in the published literature (Figure 5B) [71–77]. These cases encompass translocations that drive different types of cancer (such as lymphomas) as well as cognitive disorders, and also translocations that are phenotypically neutral. Finally, CGRs typifying **chromothripsis**, a break-induced shattering of chromosomes caused by a single catastrophic event, also display at their junctions the signatures typical of TMEJ [66,71,78–80].

Templated Insertions in BRCAness Signature

The observation that tumor cells deficient in HR have become reliant on Polθ for their survival [30,31] suggests that in those cells, HR substrates are rerouted towards TMEJ. The obvious question whether HR-deficient tumors accumulate TMEJ signatures in their genome is yet unanswered, but this seems highly probable. Indeed, whole genome sequence analysis of tumors identified a distinct mutational signature that typifies tumors that are potentially HR-deficient [81–85]. This mutational signature number 3 (also termed **BRCAness**) is associated with small insertion and deletion signature 6, characterized by deletions with microhomology and frequent occurrence of >3 bp insertions. Future work will reveal whether Polθ action is indeed causally implicated in cancer-promoting mutation accumulation in BRCA-deficient cells and also whether TMEJ signatures can be used as a classifier to stratify tumors (see Outstanding Questions).

Concluding Remarks

Alt-EJ has long been considered to be of less relevance to human health than NHEJ. While Alt-EJ has been suggested to be causal to microhomology-mediated DNA translocations, the NHEJ pathway and DNA slippage mechanisms were thought to be primarily responsible for deletion mutagenesis. The identification of Polθ as a quintessential component of a stand-alone, intrinsically mutagenic DSB repair pathway, and the recognition of a unique and evolutionary conserved mutational outcome (templated insertions) calls for a re-evaluation. Figure 5 highlights the profound contribution of this pathway to spontaneous mutagenesis in the human population, showing that TMEJ is thus a direct contributor to human disease etiology and genome diversification. There are still many questions to be answered concerning the biology of TMEJ; perhaps first and foremost is how cells prevent this mutagenic pathway to act on the substrates that it can act upon when HR is compromised. In other words, how do cells establish hierarchy in DSB repair,

Outstanding Questions

How do cells establish a favorable order of events in DSB repair, preferring error-free homologous recombination repair over mutagenic TMEJ? Do signaling cascades exist, potentially involving chromatin marks, that discriminate between nondamaged and thus available sister chromatids versus those that are damaged and unusable?

Is TMEJ restricted to the S/G2 phases of the cell cycle, or can it also repair DSBs in G1 phase and/or nondividing cells? How is this regulated?

What triggers template switching during Polθ action and thus results in templated insertions? Are the loop domains in vertebrate Polθ modifying this behavior? Why do different species have a different requirement for the extent of microhomology (1 bp in worms versus 2–3 bp in mammals)?

Which enzymes act up- and downstream of Polθ in TMEJ? Are these required for each manifestation of TMEJ? Does Polθ hand over the substrate to more accurate DNA polymerase once a stable intermediate is established?

Is the suggested prominent role of TMEJ mostly germline intrinsic and the result of fast proliferation?

Do cells of patients with an HR deficiency accumulate templated insertions in their genome? And if so, can templated insertions be used as a 'biomarker' for HR-efficiency in patients that carry variants of unknown significance and/or to identify patients that might benefit from future Polθ inhibitors?

prevailing error-free HR over mutagenic end joining? Another question concerns the seemingly varying contributions of NHEJ and TMEJ to end- joining outcomes in different cell types: several studies have pointed towards a more prominent role of TMEJ in germ cells and during embryogenesis [35,41,45]. It was shown that breakpoints of germline variants show significantly more templated insertions and usage of microhomology as compared to those of somatic variants [66]. To the question why do cells need another pathway beside the versatile classical NHEJ and HR, the following conceptual explanation has been proposed [38,42]: to repair replication-associated breaks that are committed to perform HR but, for different reasons, cannot use the sister chromatid as a template. These potentially also include breaks that have escaped NHEJ and interrupt both sister chromatids upon DNA replication, providing an explanation for the observed back-up activity of TMEJ in NHEJ-compromised cells. In such defined physiological contexts, 'alternative' end joining is not alternative to anything, but the only solution to a specific problem, unfortunately with mutations as a (small) price to pay.

Acknowledgments

J.S. is supported by a Veni grant from The Netherlands Organization for Scientific Research for Earth and Life Sciences. M.T. is supported by the Dutch Cancer Society (Kankerbestrijding).

References

1. Bunting, S.F. and Nussenzweig, A. (2013) End-joining, translocations and cancer. *Nat. Rev. Cancer* 13, 443–454
2. Helleday, T. *et al.* (2014) Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* 15, 585–598
3. Carvalho, C.M. and Lupski, J.R. (2016) Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* 17, 224–238
4. Jasin, M. and Rothstein, R. (2013) Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.* 5, a012740
5. Hustedt, N. and Durocher, D. (2016) The control of DNA repair by the cell cycle. *Nat. Cell Biol.* 19, 1–9
6. Nakamura, K. *et al.* (2019) H4K20me0 recognition by BRCA1-BARD1 directs homologous recombination to sister chromatids. *Nat. Cell Biol.* 21, 311–318
7. Saredi, G. *et al.* (2016) H4K20me0 marks post-replicative chromatin and recruits the TONSL-MMS22L DNA repair complex. *Nature* 534, 714–718
8. Pellegrino, S. *et al.* (2017) Replication-coupled dilution of H4K20me2 guides 53BP1 to pre-replicative chromatin. *Cell Rep.* 19, 1819–1831
9. Fradet-Turcotte, A. *et al.* (2013) 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature* 499, 50–54
10. Ghezraoui, H. *et al.* (2018) 53BP1 cooperation with the REV7-shieldin complex underpins DNA structure-specific NHEJ. *Nature* 560, 122–127
11. Noordermeer, S.M. *et al.* (2018) The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* 560, 117–121
12. Dev, H. *et al.* (2018) Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. *Nat. Cell Biol.* 20, 954–965
13. Gupta, R. *et al.* (2018) DNA repair network analysis reveals shieldin as a key regulator of NHEJ and PARP inhibitor sensitivity. *Cell* 173, 972–988
14. Mirman, Z. *et al.* (2018) 53BP1-RIF1-shieldin counteracts DSB resection through CST- and Polalpha-dependent fill-in. *Nature* 560, 112–116
15. Setiaputra, D. and Durocher, D. (2019) Shieldin – the protector of DNA ends. *EMBO Rep.* 20, e47560
16. Krasner, D.S. *et al.* (2015) Interplay between Ku and replication protein A in the restriction of Exo1-mediated DNA break end resection. *J. Biol. Chem.* 290, 18806–18816
17. Chang, H.H.Y. *et al.* (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* 18, 495–506
18. Betermier, M. *et al.* (2014) Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet.* 10, e1004086
19. Boulton, S.J. and Jackson, S.P. (1996) *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.* 15, 5093–5103
20. Kabotyanski, E.B. *et al.* (1998) Double-strand break repair in Ku86- and XRCC4-deficient cells. *Nucleic Acids Res.* 26, 5333–5342
21. Weinstock, D.M. *et al.* (2007) Formation of NHEJ-derived reciprocal chromosomal translocations does not require Ku70. *Nat. Cell Biol.* 9, 978–981
22. Simsek, D. *et al.* (2011) DNA ligase III promotes alternative non-homologous end-joining during chromosomal translocation formation. *PLoS Genet.* 7, e1002080
23. Zhang, Y. and Jasin, M. (2011) An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nat. Struct. Mol. Biol.* 18, 80–84
24. Simsek, D. and Jasin, M. (2010) Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat. Struct. Mol. Biol.* 17, 410–416
25. Audebert, M. *et al.* (2004) Involvement of poly (ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *J. Biol. Chem.* 279, 55117–55126
26. Wang, H. *et al.* (2003) Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Res.* 31, 5377–5388
27. Wang, M. *et al.* (2006) PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res.* 34, 6170–6182
28. Higgins, G.S. *et al.* (2010) A small interfering RNA screen of genes involved in DNA repair identifies tumor-specific radiosensitization by POLQ knockdown. *Cancer Res.* 70, 2984–2993
29. Lemee, F. *et al.* (2010) DNA polymerase θ up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proc. Natl. Acad. Sci. U. S. A.* 107, 13390–13395
30. Ceccaldi, R. *et al.* (2015) Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature* 518, 258–262
31. Mateos-Gomez, P.A. *et al.* (2015) Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* 518, 254–257
32. Shima, N. *et al.* (2003) Phenotype-based identification of mouse chromosome instability mutants. *Genetics* 163, 1031–1040

33. Chan, S.H. *et al.* (2010) Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet.* 6, e1001005
34. Yu, A.M. and McVey, M. (2010) Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic Acids Res.* 38, 5706–5717
35. Schimmel, J. *et al.* (2017) Mutational signatures of non-homologous and polymerase theta-mediated end-joining in embryonic stem cells. *EMBO J.* 36, 3634–3649
36. van Kregten, M. *et al.* (2016) T-DNA integration in plants results from polymerase-θ-mediated DNA repair. *Nat. Plants* 2, 16164
37. Koole, W. *et al.* (2014) A polymerase theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat. Commun.* 5, 3216
38. Roerink, S.F. *et al.* (2014) Polymerase theta-mediated end joining of replication-associated DNA breaks in *C. elegans*. *Genome Res.* 24, 954–962
39. Yousefzadeh, M.J. *et al.* (2014) Mechanism of suppression of chromosomal instability by DNA polymerase POLQ. *PLoS Genet.* 10, e1004654
40. Wyatt, D.W. *et al.* (2016) Essential roles for polymerase θ-mediated end joining in the repair of chromosome breaks. *Mol. Cell* 63, 662–673
41. Thyme, S.B. and Schier, A.F. (2016) Polq-mediated end joining is essential for surviving DNA double-strand breaks during early zebrafish development. *Cell Rep.* 15, 707–714
42. Lemmens, B. *et al.* (2015) Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. *Nat. Commun.* 6, 8909
43. Wang, Z. *et al.* (2019) DNA polymerase θ (POLQ) is important for repair of DNA double-strand breaks caused by fork collapse. *J. Biol. Chem.* 294, 3909–3919
44. Truong, L.N. *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
45. van Schendel, R. *et al.* (2015) Polymerase theta is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis. *Nat. Commun.* 6, 7394
46. Kawamura, K. *et al.* (2004) DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers. *Int. J. Cancer* 109, 9–16
47. Schneider, L. *et al.* (2012) Terminally differentiated astrocytes lack DNA damage response signaling and are radioresistant but retain DNA repair proficiency. *Cell Death Differ.* 19, 582–591
48. Iyama, T. and Wilson III., D.M. (2013) DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst)* 12, 620–636
49. van Schendel, R. *et al.* (2016) Genomic scars generated by polymerase theta reveal the versatile mechanism of alternative end-joining. *PLoS Genet.* 12, e1006368
50. Zelensky, A.N. *et al.* (2017) Inactivation of Pol θ and C-NHEJ eliminates off-target integration of exogenous DNA. *Nat. Commun.* 8, 66
51. Saito, S. *et al.* (2017) Dual loss of human POLQ and LIG4 abolishes random integration. *Nat. Commun.* 8, 16112
52. Seki, M. *et al.* (2003) POLQ (Pol θ), a DNA polymerase and DNA-dependent ATPase in human cells. *Nucleic Acids Res.* 31, 6117–6126
53. Sharief, F.S. *et al.* (1999) Cloning and chromosomal mapping of the human DNA polymerase theta (POLQ), the eighth human DNA polymerase. *Genomics* 59, 90–96
54. Beagan, K. and McVey, M. (2016) Linking DNA polymerase theta structure and function in health and disease. *Cell. Mol. Life Sci.* 73, 603–615
55. Black, S.J. *et al.* (2016) DNA polymerase θ: a unique multifunctional end-joining machine. *Genes (Basel)* 7, 67
56. Kent, T. *et al.* (2016) Polymerase θ is a robust terminal transferase that oscillates between three different mechanisms during end-joining. *Elife* 5, e13740
57. Kent, T. *et al.* (2015) Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. *Nat. Struct. Mol. Biol.* 22, 230–237
58. He, P. and Yang, W. (2018) Template and primer requirements for DNA Pol θ-mediated end joining. *Proc. Natl. Acad. Sci. U. S. A.* 115, 7747–7752
59. Zahn, K.E. *et al.* (2015) Human DNA polymerase θ grasps the primer terminus to mediate DNA repair. *Nat. Struct. Mol. Biol.* 22, 304–311
60. Ramsden, D.A. and Asagoshi, K. (2012) DNA polymerases in nonhomologous end joining: are there any benefits to standing out from the crowd? *Environ. Mol. Mutagen.* 53, 741–751
61. Mateos-Gomez, P.A. *et al.* (2017) The helicase domain of Polθ counteracts RPA to promote alt-NHEJ. *Nat. Struct. Mol. Biol.* 24, 1116–1123
62. Beagan, K. *et al.* (2017) *Drosophila* DNA polymerase theta utilizes both helicase-like and polymerase domains during microhomology-mediated end joining and interstrand crosslink repair. *PLoS Genet.* 13, e1006813
63. Zhang, F. *et al.* (2009) The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat. Genet.* 41, 849–853
64. Lieber, M.R. *et al.* (1988) Lymphoid V (D) J recombination: nucleotide insertion at signal joints as well as coding joints. *Proc. Natl. Acad. Sci. U. S. A.* 85, 8588–8592
65. Landrum, M.J. *et al.* (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 42 (Database issue), D980–D985
66. Malhotra, A. *et al.* (2013) Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome Res.* 23, 762–776
67. Kloosterman, W.P. *et al.* (2015) Characteristics of *de novo* structural changes in the human genome. *Genome Res.* 25, 792–801
68. Abyzov, A. *et al.* (2015) Analysis of deletion breakpoints from 1,092 humans reveals details of mutation mechanisms. *Nat. Commun.* 6, 7256
69. Pannunzio, N.R. *et al.* (2014) Non-homologous end joining often uses microhomology: implications for alternative end joining. *DNA Repair (Amst)* 17, 74–80
70. Ghezraoui, H. *et al.* (2014) Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Mol. Cell* 55, 829–842
71. Boeva, V. *et al.* (2013) Breakpoint features of genomic rearrangements in neuroblastoma with unbalanced translocations and chromothripsis. *PLoS One* 8, e72182
72. Jager, U. *et al.* (2000) Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. *Blood* 95, 3520–3529
73. Murga Penas, E.M. *et al.* (2010) The t(14;18)(q32;q21)/IGH-MALT1 translocation in MALT lymphomas contains templated nucleotide insertions and a major breakpoint region similar to follicular and mantle cell lymphoma. *Blood* 115, 2214–2219
74. Nilsson, D. *et al.* (2017) Whole-genome sequencing of cytogenetically balanced chromosome translocations identifies potentially pathological gene disruptions and highlights the importance of microhomology in the mechanism of formation. *Hum. Mutat.* 38, 180–192
75. Weckselblatt, B. *et al.* (2015) Unbalanced translocations arise from diverse mutational mechanisms including chromothripsis. *Genome Res.* 25, 937–947
76. Welzel, N. *et al.* (2001) Templated nucleotide addition and immunoglobulin JH-gene utilization in t(11;14) junctions: implications for the mechanism of translocation and the origin of mantle cell lymphoma. *Cancer Res.* 61, 1629–1636
77. Schluth-Bolard, C. *et al.* (2019) Whole genome paired-end sequencing elucidates functional and phenotypic consequences of balanced chromosomal rearrangement in patients with developmental disorders. *J. Med. Genet.* Published online March 28, 2019. https://dx.doi.org/10.1136/jmedgenet-2018-105778
78. Cleal, K. *et al.* (2019) Chromothripsis during telomere crisis is independent of NHEJ and consistent with a replicative origin. *Genome Res.* 29, 737–749

79. Liu, P. *et al.* (2011) Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* 146, 889–903
80. Stephens, P.J. *et al.* (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144, 27–40
81. Davies, H. *et al.* (2017) HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* 23, 517–525
82. Nik-Zainal, S. *et al.* (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54
83. Lord, C.J. and Ashworth, A. (2016) BRCAness revisited. *Nat. Rev. Cancer* 16, 110–120
84. Turner, N. *et al.* (2004) Hallmarks of 'BRCAness' in sporadic cancers. *Nat. Rev. Cancer* 4, 814–819
85. Alexandrov, L.B. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500, 415–421