Templated Sequence Insertion Polymorphisms in the Human Genome

Masahiro Onozawa[1,2] and Peter D. Aplan[1]*

[1] Genetics Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA, [2] Department of Hematology, Hokkaido University Graduate School of Medicine, Sapporo, Japan

Templated Sequence Insertion Polymorphism (TSIP) is a recently described form of polymorphism recognized in the human genome, in which a sequence that is templated from a distant genomic region is inserted into the genome, seemingly at random. TSIPs can be grouped into two classes based on nucleotide sequence features at the insertion junctions; Class 1 TSIPs show features of insertions that are mediated via the LINE-1 ORF2 protein, including (1) target-site duplication (TSD), (2) polyadenylation 10–30 nucleotides downstream of a "cryptic" polyadenylation signal, and (3) preference for insertion at a 5′-TTTT/A-3′ sequence. In contrast, class 2 TSIPs show features consistent with repair of a DNA double-strand break (DSB) via insertion of a DNA "patch" that is derived from a distant genomic region. Survey of a large number of normal human volunteers demonstrates that most individuals have 25–30 TSIPs, and that these TSIPs track with specific geographic regions. Similar to other forms of human polymorphism, we suspect that these TSIPs may be important for the generation of human diversity and genetic diseases.

Keywords: templated sequence insertion polymorphisms (TSIPs), mitochondria, polymorphism, human migration, DNA repair, LINE-1 retrotransposon

INTRODUCTION

Maintenance of chromosomal integrity is required for the survival of all organisms, from simple prokaryotes to complex eukaryotes. This maintenance of chromosomal integrity is accomplished by DNA repair enzymes. There are a number of DNA repair systems that operate in eukaryotes, including DNA mismatch repair, DNA single-strand break repair, and DNA double-strand break (DSB) repair. DNA DSB repair can be further subdivided into repair by homologous recombination, "canonical" non-homologous end-joining (NHEJ), and "non-canonical" NHEJ (Chiruvella et al., 2013; Deriano and Roth, 2013).

Transfected plasmid DNA can be captured and used as a patch at the site experimentally induced DNA DSBs; the DNA patches typically show signs of NHEJ, such as micro-deletion, microhomology, and non-templated nucleotide addition (Lin and Waldman, 2001; Varga and Aplan, 2005; Cheng et al., 2010). Moreover, Yu and Gabriel detected mitochondrial DNA fragments at the site of HO endonuclease-induced DNA DSBs (Yu and Gabriel, 1999), demonstrating that a DNA DSB can be repaired by insertion of DNA sequences in yeast.

RNA can provide a template for DNA synthesis during telomere elongation (Autexier and Lue, 2006) or reverse transcription of retrotransposons (Baltimore, 1985). Several lines of evidence suggest that endogenous retrotransposons may also have a role in DNA DSB repair in human cells.

When introduced into yeast, the human LINE-1 ORF2 can mediate repair of HO endonuclease-induced DNA DSB via insertion of retrotransposon sequences (Teng et al., 1996), through a cDNA intermediate. Subsequently, synthetic RNA oligonucleotides were shown to be a template for DNA synthesis during repair of HO endonuclease induced DNA DSB in yeast, although the efficiency of repair with RNA oligonucleotides was orders of magnitude lower than DNA oligonucleotides (Storici et al., 2007). Finally, a role for LINE-1 retrotransposons in DNA DSB repair in mammalian cells has been predicted (Morrish et al., 2002, 2007). This prediction was based on clever experiments that showed new integrations of an endonuclease-incompetent LINE-1 retrotransposon could be identified in rodent cell lines; this form of LINE-1 integration was designated endonuclease-independent (ENi) retrotransposition (Morrish et al., 2002). Because these new integration sites lacked the typical features of LINE-1 endonuclease mediated insertions, such as Target-Site Duplications (TSDs) and poly(A) tracts, the authors hypothesized that LINE-1 sequences were used as a "patch" to repair a spontaneous DNA DSB (as opposed to one introduced by the LINE-1 endonuclease).

With the advent of next generation sequencing technologies, millions of germline variants in mammalian genomes within a species have been identified. Most characterized variants, known as polymorphisms, fall into three large categories; single nucleotide polymorphisms (SNPs), small (<50 bp) insertions or deletions, referred to collectively as short indels, and large deletions (>50 bp) (Genomes Project et al., 2012). More recently, a smaller number of polymorphic insertions of retro-elements (such as LINE-1 or Alu) have been identified through the use of sophisticated methods to detect mobile element insertions (Beck et al., 2010; Huang et al., 2010; Iskow et al., 2010). In addition, insertion of processed gene transcripts into the germline have been identified and referred to as polymorphic pseudogenes (Ewing et al., 2013). Recent reports have demonstrated that insertions of retroelements and pseudogenes represent only a fraction of the insertional polymorphisms in the human genome (Onozawa et al., 2014, 2015).

DNA DSBS CAN BE REPAIRED BY INSERTION OF SEQUENCES DERIVED FROM DISTANT REGIONS OF THE GENOME IN AN EXPERIMENTAL SYSTEM
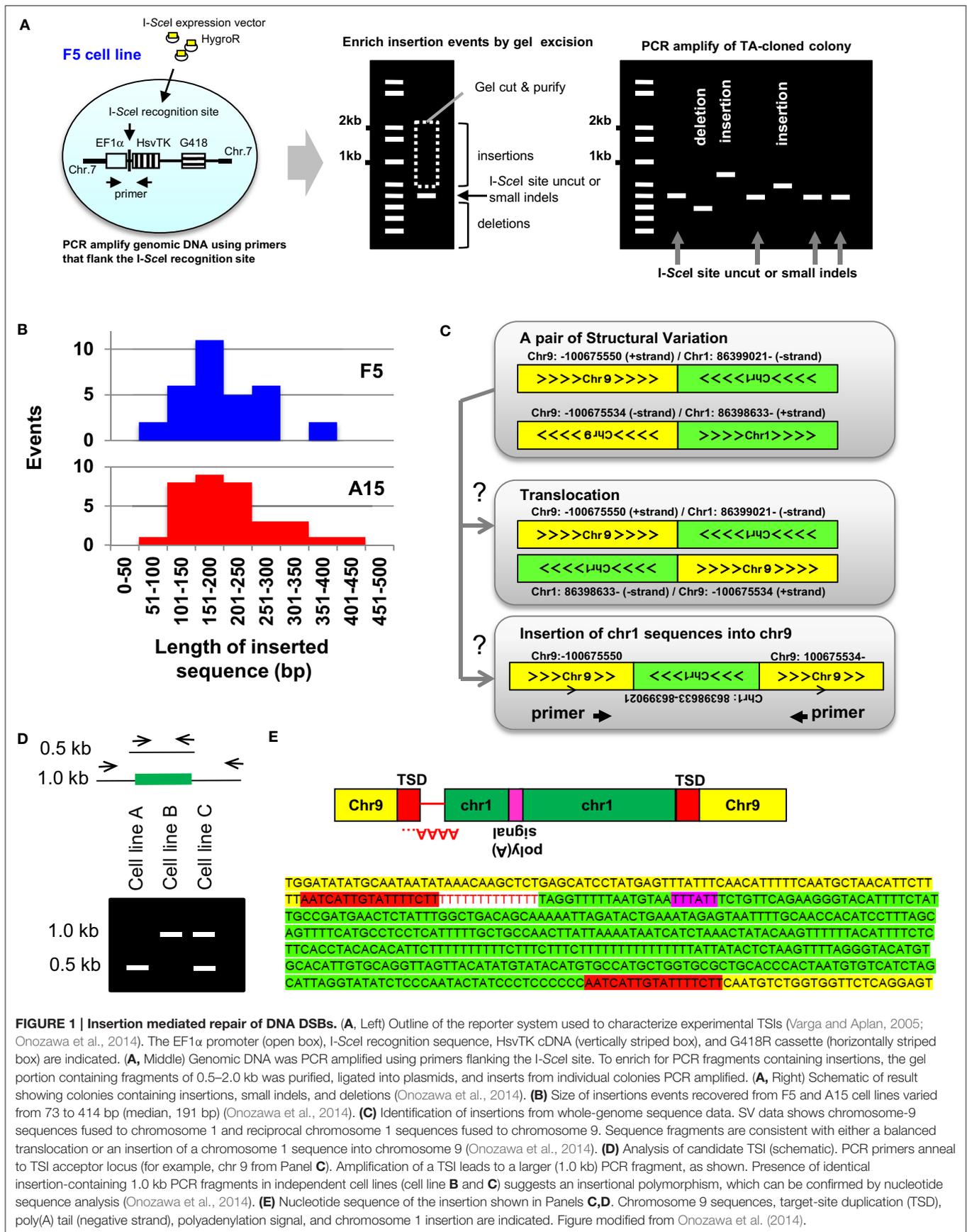
Templated Sequence Insertions (TSI) were first characterized using a cell culture based approach to study DNA DSB (Varga and Aplan, 2005). These studies modified an experimental system that had been popularized by Jasin and colleagues (Jasin and Haber, 2016), and employed a vector (designated EF1αTK), that contained the EF1α promoter driving expression of the herpes simplex virus thymidine kinase (HsvTK); the recognition site for the rare-cutting meganuclease I-SceI was inserted between the EF1α promoter and HsvTK cDNA (Varga and Aplan, 2005). Expression of the HsvTK enzyme in mammalian cells confers sensitivity to the nucleoside analog ganciclovir; thus, millions of cells can quickly be screened for loss of HsvTK

expression by treatment with ganciclovir. The EF1αTK vector was electroporated into the human leukemia cell line U937, or ovarian cancer cell line OVCAR8. Sub-clones that had integrated a single copy of the EF1αTK vector, designated "F5" for U937 (Varga and Aplan, 2005) or "A15" for OVCAR8 (Cheng et al., 2010), were isolated (**Figure 1A**). In an attempt to induce gross chromosomal rearrangements, an I-SceI expression vector was introduced into the cell lines, followed by ganciclovir (GCV) selection of clones that had lost expression of HsvTK due to mis-repair of a DNA DSB.

Although most of the GCV-resistant clones had short deletions encompassing the HsvTK start codon, rare clones that had undergone an insertion at the DNA DSB site were identified (Varga and Aplan, 2005; Cheng et al., 2010; Onozawa et al., 2014). In these clones, the inserted sequence was not derived from nearby genomic sequences, but instead mapped to a distant region of the genome. After modifying the procedure to enrich for insertions (**Figure 1A**), 32 insertions of sequences derived from distant regions of the genome were identified in F5 subclones and 34 in A15 subclones (**Figure 1B**). The origin of these insertions mapped to 18 of 24 human chromosomes, without an obvious preference for any chromosome or chromosomal region (Onozawa et al., 2014). These insertions were designated "TSI," or TSI, in contrast to the short, non-TSI seen at the site of NHEJ-mediated repair of a DNA DSB (Onozawa et al., 2014). The TSI junctions often showed features of NHEJ, such as microhomology and non-templated nucleotide addition. Generation of TSIs seemed to be generalizable, as they were not restricted to I-SceI induced cleavage but also found at TALEN-mediated cleavage sites (Onozawa et al., 2014).

TSIs ARE DERIVED FROM RNA

The TSIs inserted at the site of NHEJ mediated DNA DSB were not excised from the genome, since the donor sites were intact, and the donor (inserted) sequence had an additional copy compared to genomic regions flanking the donor site (Onozawa et al., 2014). The TSI sequences were enriched for transcribed sequences, suggesting that the TSI may have originated via reverse transcription of RNA. In addition, treatment of cells that expressed the I-SceI enzyme with reverse-transcriptase inhibitors suppressed the frequency of TSIs at DNA DSB site by more than three-fold. Moreover, co-transfection of murine RNA with an I-SceI expression vector into the human F5 cell line showed that reverse transcribed murine sequences were used as insertions at the I-SceI cleavage site. Finally, three insertions displayed mammalian telomere repeat (TTAGGG)n sequences, suggesting that telomerase RNA can also be used as a TSI template. Taken together, although other mechanisms, such as template switching by DNA polymerases or break-induced repair remain possible (Malkova and Haber, 2012; Morrish et al., 2013), the above observations support the hypothesis that reverse transcribed RNA can be used as a template to patch a DNA DSB (Onozawa et al., 2014). Potential sources of this reverse transcriptase activity include LINE-1 ORF2, HERVs (Hohn et al., 2013), as well as the

FIGURE 1 | Insertion mediated repair of DNA DSBs. (A, Left) Outline of the reporter system used to characterize experimental TSIs (Varga and Aplan, 2005; Onozawa et al., 2014). The EF1α promoter (open box), I-*Sce*I recognition sequence, HsvTK cDNA (vertically striped box), and G418R cassette (horizontally striped box) are indicated. (**A,** Middle) Genomic DNA was PCR amplified using primers flanking the I-*Sce*I site. To enrich for PCR fragments containing insertions, the gel portion containing fragments of 0.5–2.0 kb was purified, ligated into plasmids, and inserts from individual colonies PCR amplified. (**A,** Right) Schematic of result showing colonies containing insertions, small indels, and deletions (Onozawa et al., 2014). **(B)** Size of insertions events recovered from F5 and A15 cell lines varied from 73 to 414 bp (median, 191 bp) (Onozawa et al., 2014). **(C)** Identification of insertions from whole-genome sequence data. SV data shows chromosome-9 sequences fused to chromosome 1 and reciprocal chromosome 1 sequences fused to chromosome 9. Sequence fragments are consistent with either a balanced translocation or an insertion of a chromosome 1 sequence into chromosome 9 (Onozawa et al., 2014). **(D)** Analysis of candidate TSI (schematic). PCR primers anneal to TSI acceptor locus (for example, chr 9 from Panel **C**). Amplification of a TSI leads to a larger (1.0 kb) PCR fragment, as shown. Presence of identical insertion-containing 1.0 kb PCR fragments in independent cell lines (cell line **B** and **C**) suggests an insertional polymorphism, which can be confirmed by nucleotide sequence analysis (Onozawa et al., 2014). **(E)** Nucleotide sequence of the insertion shown in Panels **C,D**. Chromosome 9 sequences, target-site duplication (TSD), poly(A) tail (negative strand), polyadenylation signal, and chromosome 1 insertion are indicated. Figure modified from Onozawa et al. (2014).

TERT subunit of telomerase, which has recently been suggested to possess hTR-independent RT activity (Sharma et al., 2012).

TSIs ARE NOT AN ARTEFACT OF EXPERIMENTAL, INDUCED DNA DSB, BUT CAN BE IDENTIFIED IN UNMANIPULATED HUMAN CELLS

TSIs cannot be detected from routine analysis of whole genome sequence (WGS) reads, which are typically <150 bp. However, TSIs can be identified from WGS using the principles outlined below. First, the junction of two non-homologous chromosomes is designated as a "structural variant" (SV) on short read WGS; pairs of SVs that map very closely can be ascertained by inspection of SV files (**Figure 1C**). For a pair of SVs to represent a Templated Sequence Insertion Polymorphism (TSIP), both fusion junctions must be located within 50 kb of one another. Second, the strand polarity must align such that an insertion is feasible. Third, each end of the SV needs to be localized to a single, unique genomic loci; SV that show multiple or imperfect alignments (<95% sequence identity) are excluded. Fourth, all highly repetitive alpha satellite sequences are discarded.

The authors screened SVs obtained from two multiple myeloma cell lines (KP6, MC1286PE1) using these criteria and identified 23 unique, verified TSIs (Onozawa et al., 2014). A typical example is shown in **Figures 1C–E**. Briefly, although this pair of SVs is consistent with a balanced translocation, it is also consistent with an insertion of chromosome 1 sequences into chromosome 9 (**Figure 1C**). Primers were generated that could amplify the putative insertion, including flanking sequences and both junctions, on a single PCR fragment (**Figures 1D,E**). Nucleotide sequence of the PCR product verified that the SVs were indeed produced by insertion of chromosome 1 derived sequences into chromosome 9, as opposed to a reciprocal translocation between chromosomes 1 and 9 (**Figure 1E**). Interestingly 8 out of 23 insertions were identical or nearly identical in the two cell lines (Onozawa et al., 2014) (a schematic example of this phenomenon is shown in **Figure 1D**), suggesting that these insertions represent polymorphisms in the human genome as opposed to SV acquired by the tumor cells. Consistent with the TSI definition above, these polymorphic insertions were designated templated-sequence insertion polymorphisms (TSIPs).

## MOST TSIs IDENTIFIED IN NORMAL HUMAN SUBJECTS ARE POLYMORPHIC

A publicly available database of WGS from 52 normal individuals of defined ethnic/geographic groups ("SV baseline genome set," filename B37baselinejunctions.tsv, available at http://www.completegenomics.com/sequence-data/download-data/) contained a total of 39,595 SVs from the 52 individuals (Onozawa et al., 2015). Using the criteria set forth in section IV above, 171 candidate TSIPs were identified (Onozawa et al., 2015). Each individual had an average of 25–30 TSIPs, and TSIPs could be classified as "common" (26%; present in at least 20%

of individuals), or "rare" (74%; present in <20% of individuals). Interestingly, three TSIPs had a frequency of almost 100%, suggesting that the reference human genome (GRCh37/hg19) is based on an uncommon variant that lacks these three TSIPs. When divided into four regional "super groups" of specified geographic origin (African, Asian, European, North American), common TSIPs were present in individuals from all regions, whereas rare TSIPs tended to be restricted to individuals from a single region (Onozawa et al., 2015). There were more TSIPs per individual of African origin than other geographic regions. All of these findings are consistent with diversity identified in previous studies of mitochondrial and Y-chromosome sequences, and are consistent with patterns of human migration and the hypothesis that *Homo sapiens* originated in Africa (Cann et al., 1987; Hammer, 1995; Underhill et al., 2000).

The investigators obtained genomic DNA from eight of the normal individuals, who had a total of 89 candidate TSIs and successfully validated 69/89 (77.5%) candidate TSIs (Onozawa et al., 2015). Since these insertions can be polymorphic, they must be heritable, leading to the conclusion that the insertion event must have taken place in either a germ cell (sperm or egg), or early-stage embryo.

SEQUENCES USED AS TEMPLATES FOR TSIPs

Nucleotide sequence analysis of the insertion sequences revealed that partial LINE-1 elements, cDNAs (with several spliced exons), non-annotated intergenic or intronic sequences, and mitochondrial sequences were used as templates for TSIPs identified in normal individuals (Onozawa et al., 2015). Of note, although mitochondrial fragment insertions were commonly identified as TSIP donor sequences, no TSIs derived from mitochondrion were identified in experimentally induced TSIs using the F5 and A15 cell lines described above (Varga and Aplan, 2005; Cheng et al., 2010; Onozawa et al., 2014). Although speculative, it is possible that mitochondrial sequence insertions may be reproduction-specific events that take place in germ cells or early stage embryos, but do not occur, or occur only rarely, in somatic cells. Interestingly, sperm mitochondria are known to be ubiquitinated and destroyed shortly after fertilization (Sutovsky et al., 2000), leading to the hypothesis that fragmented paternal mitochondrial DNA [or reverse transcribed RNA that was encoded by mitochondrial DNA (Sharma et al., 2012)] can be used to patch a DNA DSB in a fertilized embryo, leading to a TSIP which contained mitochondrial sequence in all cells of the individual, including germ cells (Woischnik and Moraes, 2002; Onozawa et al., 2015; Zhou et al., 2016). Although no TSIPs contained telomere sequences (Onozawa et al., 2014, 2015), interstitial telomeric sequences (ITSs) have been identified in several species (Ruiz-Herrera et al., 2002), and were identified as insertions in the I-*Sce*I experimental system (Onozawa et al., 2014), leading to the speculation that the ITSs may have resulted from telomere patches used to repair a DNA DSB (Nergadze et al., 2004, 2007; Onozawa et al., 2014).
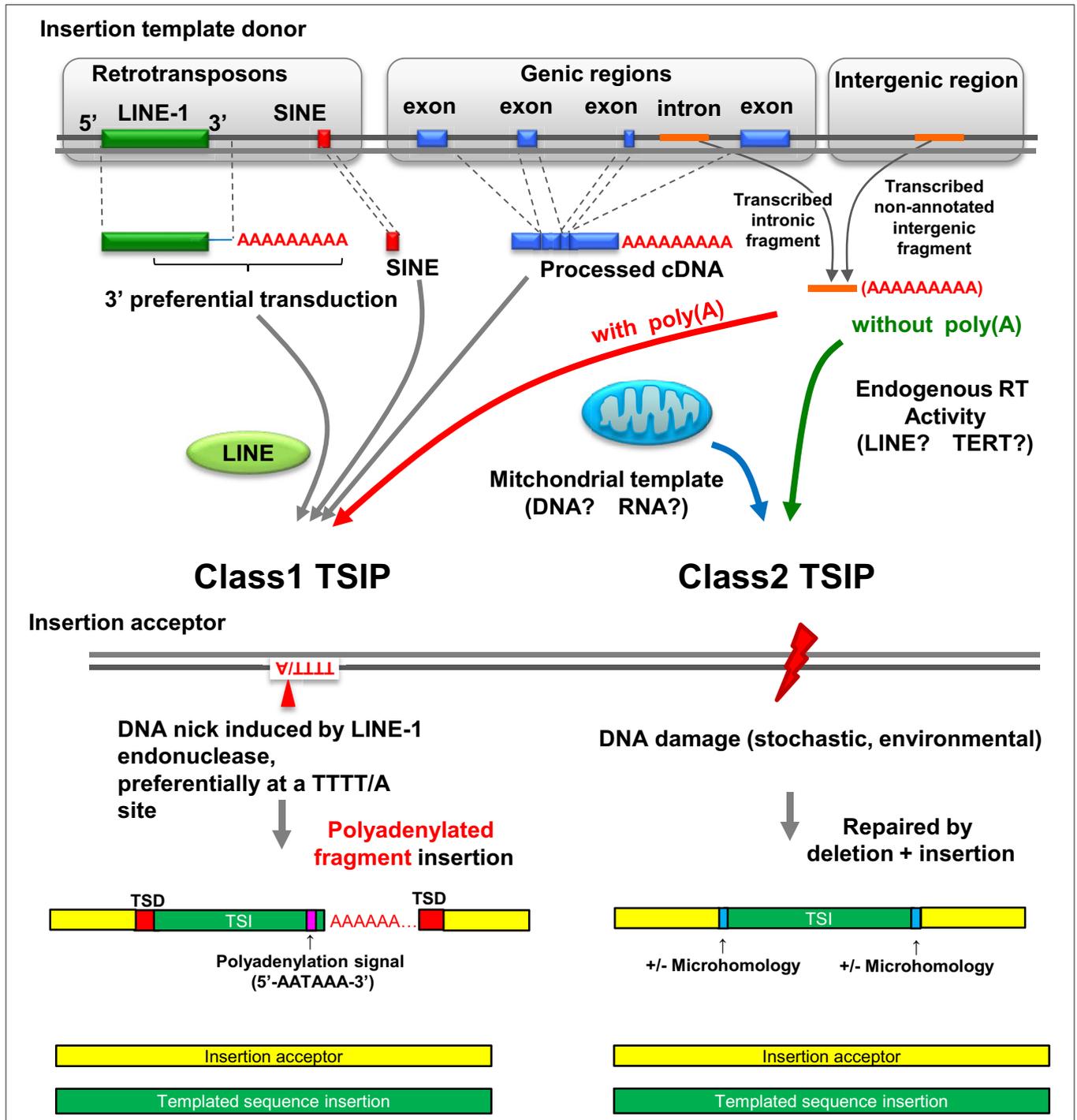
FIGURE 2 | Landscape of insertion polymorphisms in the human genome. LINE-1 mediated integration of LINE-1/SINE sequences, LINE-1 sequences (which may include additional 3′ transduced sequences), and processed cDNA insertions are known to create insertion polymorphisms (Beck et al., 2010; Huang et al., 2010; Iskow et al., 2010; Ewing et al., 2013). Polyadenylated intronic or intergenic fragments can also be acted upon in *trans* by LINE-1 ORF2 and integrate at the site of a nick created by LINE-1 ORF2, resulting in a class 1 TSIP. Class 2 TSIPs can be generated by reverse transcription of RNA transcripts into a cDNA patch that is used to repair a DNA DSB via a NHEJ mechanism. Alternatively, RNA could be inserted in the DNA DSB and used directly as a patch template, as reported for yeast (Storici et al., 2007). Finally a DNA DSB can be repaired by fragments of mitochondrial DNA or cDNA; mitochondrial insertions seem to be unique to germ cells or embryos. Figure modified from Onozawa et al. (2015).

CLASS 1 AND CLASS 2 TSIPs

TSIPs can be placed into two classes (class 1 and class 2) based on nucleotide sequences at the insertion site (**Figure 2**). Class 1 TSIPs show a duplication of recipient sequences at both insertion junctions of at least 5 bp; this feature is reminiscent of a TSD that is characteristic of insertions caused by cleavage and insertion of LINE-1 sequences. In addition, class 1 TSIs typically inserted at a preferred LINE-1 integration site (consensus sequence 5′-TTTT/A-3′), and contained a non-templated addition of 10–40 "A" residues, as well as a polyadenylation signal (5′-AATAAA-3′) located 10–20 nucleotides upstream of the poly(A) track. These features strongly support an insertion mediated by LINE-1 ORF2 protein, which contains both endonuclease and reverse transcriptase activity, acting upon non-LINE-1 RNA, and inserting the sequence into a distant region of the genome (Luan et al., 1993; Piskareva et al., 2013). Moreover, since these events must have occurred in germ cells or embryos to be transmitted, it is interesting to note that LINE-1 elements are de-repressed and active in embryos (Castro-Diaz et al., 2014). We can detect no obvious physiologic function for the Class 1 TSIPs, and suspect that these are caused by a careless LINE-1 ORF2 protein causing mischief throughout the genome. Class 2 TSIPs had none of these features (i.e., no TSDs, cryptic poly(A) signal, or poly(A) tract) but instead displayed NHEJ features such as microdeletion, microhomology, and non-templated nucleotide addition at the insertion junction, similar to what one would predict if an ENi retrotransposition event (Morrish et al., 2002) used non-LINE-1 RNA as a template. Consistent with this prediction, it is well established that LINE-1 ORF2 can act in *trans* (Wei et al., 2001). All experimental DNA DSB repair events were class 2 events, and we speculate that class 2 TSIPs are caused by DNA DSB repair events that occurred in a germ cell or early stage embryo of an ancestral individual.

POTENTIAL TO CAUSE GENETIC DISEASE

There is potential for this mechanism of DNA DSB repair to cause genetic disease. Several TSIPs disrupted the coding region of a gene (Onozawa et al., 2015). Furthermore, a recent report described a constitutional 72-bp insertion of mitochondrial sequence into the coding region of *GLI3*, leading to Pallister-Hall syndrome (Turner et al., 2003). Of note, the conception of this patient was temporally and geographically associated with high-level radioactive contamination following the Chernobyl accident (Turner et al., 2003). Although speculative, it is conceivable that a DNA DSB in the germ cell, caused by ionizing radiation, was repaired by a TSI derived from mitochondrial DNA in this individual.

CONCLUSION

TSIPs encompass several forms of insertion polymorphisms in human genomes, and are mediated via a combination of mechanisms. Class 1 TSIPs are retrotransposon-mediated events that insert polyadenylated, reverse-transcribed cDNA into seemingly random regions of the genome, whereas class 2 TSIPs are consistent with DNA DSB repair events, in which a short fragment of reverse-transcribed RNA is used as a patch to repair a DNA DSB. These TSIPs provide unique polymorphic markers, similar to SNPs and variable tandem repeats, and can be used to track population migration and evolution. Similar to retrotransposon insertions, we suspect that TSIPs, which can disrupt coding regions of the genome, may play a role in both the etiology of genetic diseases as well as mammalian evolution.

AUTHOR CONTRIBUTIONS

PA conceived and edited mini-review. MO conceived mini-review, generated figures, and wrote the first draft.

ACKNOWLEDGMENTS

REFERENCES

Autexier, C., and Lue, N. F. (2006). The structure and function of telomerase reverse transcriptase. *Annu. Rev. Biochem.* 75, 493–517. doi: 10.1146/annurev.biochem.75.103004.142412

Baltimore, D. (1985). Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. *Cell* 40, 481–482. doi: 10.1016/0092-8674(85)90190-4

Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., et al. (2010). LINE-1 retrotransposition activity in human genomes. *Cell* 141, 1159–1170. doi: 10.1016/j.cell.2010.05.021

Cann, R. L., Stoneking, M., and Wilson, A. C. (1987). Mitochondrial, D. N. A., and human evolution. *Nature* 325, 31–36. doi: 10.1038/325031a0

Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., et al. (2014). Evolutionarily dynamic L1 regulation in embryonic stem cells. *Genes Dev.* 28, 1397–1409. doi: 10.1101/gad.241661.114

Cheng, Y., Zhang, Z., Keenan, B., Roschke, A. V., Nakahara, K., and Aplan, P. D. (2010). Efficient repair of DNA double-strand breaks in malignant cells with structural instability. *Mutat. Res.* 683, 115–122. doi: 10.1016/j.mrfmmm.2009.10.016

Chiruvella, K. K., Liang, Z., and Wilson, T. E. (2013). Repair of double-strand breaks by end joining. *Cold Spring Harb. Perspect. Biol.* 5:a012757. doi: 10.1101/cshperspect.a012757

Deriano, L., and Roth, D. B. (2013). Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu. Rev. Genet.* 47, 433–455. doi: 10.1146/annurev-genet-110711-155540

Ewing, A. D., Ballinger, T. J., Earl, D., Broad Institute Genome Sequencing and Analysis Program and Platform, Harris, C. C., Ding, L., et al. (2013). Retrotransposition of gene transcripts leads to structural variation in mammalian genomes. *Genome Biol.* 14:R22. doi: 10.1186/gb-2013-14-3-r22

Genomes Project, C., Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi: 10.1038/nature11632

Hammer, M. F. (1995). A recent common ancestry for human Y chromosomes. *Nature* 378, 376–378. doi: 10.1038/378376a0

Hohn, O., Hanke, K., and Bannert, N. (2013). HERV-K(HML-2), the Best Preserved Family of HERVs: endogenization, expression, and implications in health and disease. *Front. Oncol.* 3:246. doi: 10.3389/fonc.2013.00246

Huang, C. R., Schneider, A. M., Lu, Y., Niranjan, T., Shen, P., Robinson, M. A., et al. (2010). Mobile interspersed repeats are major structural variants in the human genome. *Cell* 141, 1171–1182. doi: 10.1016/j.cell.2010.05.026

Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., et al. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261. doi: 10.1016/j.cell.2010.05.020

Jasin, M., and Haber, J. E. (2016). The democratization of gene editing: insights from site-specific cleavage and double-strand break repair. *DNA Repair (Amst).* 44, 6–16. doi: 10.1016/j.dnarep.2016.05.001

Lin, Y., and Waldman, A. S. (2001). Capture of DNA sequences at double-strand breaks in mammalian chromosomes. *Genetics* 158, 1665–1674.

Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605. doi: 10.1016/0092-8674(93)90078-5

Malkova, A., and Haber, J. E. (2012). Mutations arising during repair of chromosome breaks. *Annu. Rev. Genet.* 46, 455–473. doi: 10.1146/annurev-genet-110711-155547

Morrish, T. A., Bekbolysnov, D., Velliquette, D., Morgan, M., Ross, B., Wang, Y., et al. (2013). Multiple mechanisms contribute to telomere maintenance. *J. Cancer Biol. Res.* 1. Available online at: https://www.jscimedcentral.com/CancerBiology/Articles/cancerbiology-1-1012.pdf

Morrish, T. A., Garcia-Perez, J. L., Stamato, T. D., Taccioli, G. E., Sekiguchi, J., and Moran, J. V. (2007). Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* 446, 208–212. doi: 10.1038/nature05560

Morrish, T. A., Gilbert, N., Myers, J. S., Vincent, B. J., Stamato, T. D., Taccioli, G. E., et al. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* 31, 159–165. doi: 10.1038/ng898

Nergadze, S. G., Rocchi, M., Azzalin, C. M., Mondello, C., and Giulotto, E. (2004). Insertion of telomeric repeats at intrachromosomal break sites during primate evolution. *Genome Res.* 14, 1704–1710. doi: 10.1101/gr.2778904

Nergadze, S. G., Santagostino, M. A., Salzano, A., Mondello, C., and Giulotto, E. (2007). Contribution of telomerase RNA retrotranscription to DNA double-strand break repair during mammalian genome evolution. *Genome Biol.* 8:R260. doi: 10.1186/gb-2007-8-12-r260

Onozawa, M., Goldberg, L., and Aplan, P. D. (2015). Landscape of insertion polymorphisms in the human genome. *Genome Biol. Evol.* 7, 960–968. doi: 10.1093/gbe/evv043

Onozawa, M., Zhang, Z., Kim, Y. J., Goldberg, L., Varga, T., Bergsagel, P. L., et al. (2014). Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7729–7734. doi: 10.1073/pnas.1321889111

Piskareva, O., Ernst, C., Higgins, N., and Schmatchenko, V. (2013). The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding. *FEBS Open Bio* 3, 433–437. doi: 10.1016/j.fob.2013.09.005

Ruiz-Herrera, A., García, F., Azzalin, C., Giulotto, E., Egozcue, J., Ponsà, M., et al. (2002). Distribution of intrachromosomal telomeric sequences (ITS) on Macaca fascicularis (Primates) chromosomes and their implication for chromosome evolution. *Hum. Genet.* 110, 578–586. doi: 10.1007/s00439-002-0730-6

Sharma, N. K., Reyes, A., Green, P., Caron, M. J., Bonini, M. G., Gordon, D. M., et al. (2012). Human telomerase acts as a hTR-independent reverse transcriptase in mitochondria. *Nucleic Acids Res.* 40, 712–725. doi: 10.1093/nar/gkr758

Storici, F., Bebenek, K., Kunkel, T. A., Gordenin, D. A., and Resnick, M. A. (2007). RNA-templated DNA repair. *Nature* 447, 338–341. doi: 10.1038/nature05720

Sutovsky, P., Moreno, R. D., Ramalho-Santos, J., Dominko, T., Simerly, C., and Schatten, G. (2000). Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol. Reprod.* 63, 582–590. doi: 10.1095/biolreprod63.2.582

Teng, S. C., Kim, B., and Gabriel, A. (1996). Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature* 383, 641–644. doi: 10.1038/383641a0

Turner, C., Killoran, C., Thomas, N. S., Rosenberg, M., Chuzhanova, N. A., Johnston, J., et al. (2003). Human genetic disease caused by *de novo* mitochondrial-nuclear DNA transfer. *Hum. Genet.* 112, 303–309. doi: 10.1007/s00439-002-0892-2

Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., et al. (2000). Y chromosome sequence variation and the history of human populations. *Nat. Genet.* 26, 358–361. doi: 10.1038/81685

Varga, T., and Aplan, P. D. (2005). Chromosomal aberrations induced by double-strand DNA breaks. *DNA Repair (Amst).* 4, 1038–1046. doi: 10.1016/j.dnarep.2005.05.004

Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., et al. (2001). Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Mol. Cell. Biol.* 21, 1429–1439. doi: 10.1128/MCB.21.4.1429-1439.2001

Woischnik, M., and Moraes, C. T. (2002). Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res.* 12, 885–893. doi: 10.1101/gr.227202

Yu, X., and Gabriel, A. (1999). Patching broken chromosomes with extranuclear cellular DNA. *Mol. Cell* 4, 873–881. doi: 10.1016/S1097-2765(00)80397-4

Zhou, Q., Li, H., Li, H., Nakagawa, A., Lin, J. L., Lee, E. S., et al. (2016). Mitochondrial endonuclease G mediates breakdown of paternal mitochondria upon fertilization. *Science* 353, 394–399. doi: 10.1126/science.aaf4777