Published in final edited form as:

Trends Genet. 2018 August ; 34(8): 572–574. doi:10.1016/j.tig.2018.05.003.

Long-Distance Relationships: Suppression of Repeat-Mediated Deletions

Maria E. Morales, Ph.D., Tiffany Kaul, and Prescott Deininger, Ph.D.

Tulane Cancer Center, Tulane University, New Orleans, LA 70122

Abstract

The high proportion of repetitive DNA sequences in the human genome provides tremendous opportunities for DNA rearrangements between non-allelic repetitive elements. The genome must use multiple competing and collaborating repair mechanisms to minimize these types of DNA rearrangements, some of which fail in cancer cells where DNA repair pathways are suppressed.

Double-strand breaks (DSB) in DNA represent one of the most devastating forms of DNA damage. DSBs occur naturally from stalled replication complexes, as well as from a number of environmental insults, such as radiation. DSBs must be repaired for cell survival, and there are many competing pathways to resolve these breaks. Some of these pathways repair the DNA accurately, while others can lead to mutations that contribute to cancer and disease (reviewed in [1]).

A critical initial decision on how to repair a DSB involves whether to utilize canonical Non-Homologous End Joining (c-NHEJ) which typically creates small mutations, or a number of alternative pathways that first require DNA resection to expose a long stretch of single-stranded DNA around the breaks (see Figure 1). These resected segments can then participate in Homologous Recombination (HR), which is the most conservative repair pathway, alternative Non-Homologous End Joining (alt-NHEJ), or Single-Strand Annealing (SSA)-like mechanisms that typically create significant deletions (reviewed in [2]).

The DSB repair pathway utilized will influence the nature of damage to the genome, which is regulated by a number of factors. These include the stage of the cell cycle, as well as expression levels of many of the proteins sensing the damage or participating in the repair pathways (Figure 1)[1]. The choice of repair pathways involving resection can also be influenced by the presence of repetitive DNA sequences in the vicinity of the DSB that may interact between the exposed single-stranded regions. Although this will primarily be in the SSA sub-pathway, HR may also contribute if there are non-allelic sequences. (Figure 1).

About half of the human genome is made up of repetitive DNA sequences [3]. The two primary types of repetitive sequences that contribute to DSB repair and resulting human disease are the segmental duplications and Alu elements [4, 5]. Segmental duplications represent long regions of DNA that are highly homologous, can exceed 100 kb in length and are usually in complicated tandem arrangements. Changes in copy number through recombination alter genes in these segments leading to specific human disorders. Alu elements on the other hand, are short (less than 300 bp) segments of DNA that are spread

fairly randomly throughout the human genome and are present on the average every 3 kb. Thus, recombination between non-allelic dispersed loci, like Alu elements, can delete or duplicate the sequences between the Alu elements. Furthermore each Alu element can participate in recombination events with many other Alu elements over varying distances (Figure 1). Another unique property of Alu elements is that they are not identical, but generally have 5–30% mismatch relative to one another [3] that can contribute to alternative repair mechanisms as well [6].

In their recent article [7], the Stark laboratory has developed a new reporter system that has allowed assessment, for the first time, of the relevance of the distance between the DSB and the repetitive sequences. They utilized a variant of a common EGFP reporter approach [8], but made a homology region about 300 bp upstream of the EGFP to mimic the size of an Alu element and they added up to 3% mismatch to assess the impact of mismatches on the recombination efficiency. Their reporter was integrated in the genome in such a way that they could use one Crispr/Cas9 guide RNA to cut near the upstream homology region and a second guide RNA to cut varying distances (3.3 to 28.4 kb) from the downstream homology region.

Using their assay [7], the Stark lab showed that the frequency of Repeat-Mediated Deletions (RMDs) increased when the DSB is close to the repetitive elements and declines rapidly with distance followed by a very gradual decline up to at least 28.4 kb. At the longer distances, the RMD was much more dependent on *rad52* than at shorter distances. The role of *rad52* in SSA is to help anneal homologous or homeologous sequences to form heteroduplexes [1], whereas NHEJ and alt-NHEJ pathways are *rad52*-independent. They also showed that even a modest level of mismatch between the repetitive sequences (1–3%) significantly inhibited RMD at all distances. This inhibition by mismatches between repetitive elements was largely dependent on *msh2*, which would be expected for a process mediated by heteroduplex rejection [9].

This new study also provided further insight on the role of Ku70 and XRCC4, which are involved in the c-NHEJ pathway, demonstrating that their loss favors RMDs. This seems likely to represent the competition between this non-resection pathway and those depending on resection (Figure 1). Decreases in the pro-resection proteins CtIP and BRCA1, were found to decrease RMD. The authors [7] find that the direct regulation by heteroduplex rejection (*msh2*-mediated) is a stronger inhibitor of RMD than competition with c-NHEJ (as measured by Ku70 knockout).

The presence of dispersed Alu elements throughout the genome [3], and other repetitive DNA sequences, provides endless opportunities for RMDs that disrupt genome function. The relatively efficient excision of tens of kilobases of DNA at DSBs means that many copies of Alu may be exposed in single-strand regions leading to many possible rearrangements. The human genome could not survive the resulting instability if it did not have strong regulators inhibiting these RMDs.

There are a number of situations in which some of the DSB repair pathways are either suppressed or defective in humans. Cancer is the most obvious situation in which some of

these genes, such as msh2, brca1 and brca2 are commonly defective. Thus, in various cancers one might expect to see increases in the abundance of RMDs and possibly other changes in the nature of repeat-mediated genomic rearrangements that might lead to more genetic instability, which in turn helps the cancer evolve to a more malignant form.

References

- 1Ceccaldi R, Rondinelli B, D'Andrea AD. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol.* 2016; 26(1):52–64. [PubMed: 26437586]
- 2Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. *Annu Rev Genet.* 2011; 45:247–71. [PubMed: 21910633]
- 3Lander ES, et al. Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium. *Nature.* 2001; 409(6822):860–921. [PubMed: 11237011]
- 4Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet.* 2016; 17(4):224–38. [PubMed: 26924765]
- 5Belancio VP, Hedges DJ, Deininger P. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res.* 2008; 18(3):343–58. [PubMed: 18256243]
- 6Morales ME, et al. The contribution of alu elements to mutagenic DNA double-strand break repair. *PLoS Genet.* 2015; 11(3):e1005016. [PubMed: 25761216]
- 7Mendez-Dorantes C, Bhargava R, Stark JM. Repeat-mediated deletions can be induced by a chromosomal break far from a repeat, but multiple pathways suppress such rearrangements. *Genes Dev.* 2018; 32(7–8):524–536. [PubMed: 29636371]
- 8Gunn A, Stark JM. I-SceI-based assays to examine distinct repair outcomes of mammalian chromosomal double strand breaks. *Methods Mol Biol.* 2012; 920:379–91. [PubMed: 22941618]
- 9Sugawara N, et al. Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1. *Proc Natl Acad Sci U S A.* 2004; 101(25): 9315–20. [PubMed: 15199178]

Figure 1.

Double-Strand Break Repair. The major pathways related to DSB repair are highlighted with some of the relevant proteins marked. After the critical decision whether to resect the DNA ends, the choice of which competing pathway to use on the resected ends depends on many variables. These include various proteins involved in the sub-pathways, and also depend on exposure of repetitive elements (shown as colored arrows in the resection to indicate different levels of mismatch) that may interact in various ways within the SSA pathway. Different levels of resection at the DSB ends will result in exposure of repetitive elements at different distances from the ends, as well as exposing multiple elements on the same resected strand.

Pathway (A) represents the c-NHEJ pathway, which generally causes little or no mutation. Pathway (A) involves no resection of the DSB, but the other repair pathways often involve long resections (tens of kilobases). Pathway (B) represents HR, which requires the presence of a sister chromatid for the resected DNA to invade. Invasion at a non-allelic repetitive sequence may cause deletion or duplication. Pathway (C) represents alt-NHEJ, which results in deletions of various sizes. Pathway (D) represents the SSA, which allows repetitive sequences within the resected segments to interact leading to deletion of the segments between repeat regions.