Mechanisms for Structural Variation in the Human Genome

Benjamin B. Currall · Colby Chiangmai ·
Michael E. Talkowski · Cynthia C. Morton

Abstract It has been known for several decades that genetic variation involving changes to chromosomal structure (i.e., structural variants) can contribute to disease; however this relationship has been brought into acute focus in recent years largely based on innovative new genomics approaches and technology. Structural variants (SVs) arise from improperly repaired DNA double-strand breaks (DSB). DSBs are a frequent occurrence in all cells and two major pathways are involved in their repair: homologous recombination and non-homologous end joining. Errors during these repair mechanisms can result in SVs that involve losses, gains and rearrangements ranging from a few nucleotides to entire chromosomal arms. Factors such as rearrangements, hotspots and induced DSBs are implicated in the formation of SVs. While de novo SVs are often associated with disease, some SVs are conserved within human subpopulations and may have had a meaningful influence on primate evolution. As the ability to sequence the whole human genome rapidly evolves, the diversity of SVs is illuminated, including very complex rearrangements involving multiple DSBs in a process recently designated as "chromothripsis". Elucidating mechanisms involved in the etiology of SVs informs disease pathogenesis as well as the dynamic function associated with the biology and evolution of human genomes.

Keywords Structural variants · Chromosomal rearrangement · Homologous recombination (HR) · Nonhomologous end joining (NHEJ) · Chromothripsis

B. B. Currall · C. C. Morton (✉)
Departments of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital and Harvard Medical School, New Research Building, Room 160D, 77 Avenue Louis Pasteur, Boston, MA 02115, USA
e-mail: cmorton@partners.org

B. B. Currall · C. C. Morton
Harvard Medical School, Boston, MA, USA

C. Chiangmai
University of Virginia School of Medicine, Charlottesville, VA, USA

M. E. Talkowski
Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA

M. E. Talkowski
Department of Neurology, Harvard Medical School, Boston, MA, USA

M. E. Talkowski · C. C. Morton
Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA

C. C. Morton
Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA

Introduction

For more than a century, it has been possible to image and classify chromosomes using microscopy. As karyotyping techniques were refined, changes in chromosome number as well as structure were recognized and it became possible to correlate these changes in genomic composition and architecture to disease. In the last 5–10 years, the advent of technologies such as array comparative genomic hybridization (aCGH) and next-generation sequencing (NGS) has allowed finer resolution of structural variants (SVs) ranging in size from entire chromosomal arms to a single base pair and has demonstrated that SV frequencies can range from

extremely rare events to common population polymorphisms. In this review, we examine the causes, repair mechanisms, and errors that result in SV as well as some of the biological factors that play a role in the development of SV within the human genome.

Causes of SVs

DNA damage is a common insult to cells. It is estimated that the genomes in every cell of the human body normally undergo as many as 10,000 lesions per day [1–3]. In contrast to many forms of single nucleotide lesions, SVs result as an improper repair of a double-strand break (DSB), a lesion in which both strands of the phosphodiester backbone of DNA are broken, with frequencies of up to 50 DSB per cell cycle (Fig. 1a) [4]. DSBs are caused by factors that arise from internal cellular function (i.e., endogenous) and environmental insults (i.e., exogenous). Endogenous factors include reactive oxygen species, improperly repaired single nucleotide lesions, unrepaired single strand breaks, DNA replication stress, and self-induced DSB in meiotic and lymphoid cells [5–12]. Exogenous DNA damage can be caused by chemical mutagens, which bind to or directly change the DNA structure, and by ionizing radiation (i.e., high energy photons including X-rays and gamma rays), which can break nucleotide bonds or produce nearby free radicals [13–16]. In addition to these insults, DSB can also occur as a result of transposable elements within the genomic DNA [17]. If not repaired expeditiously, DSBs can quickly lead to destabilization of the genome and result in catastrophic events for the cell and organism [18, 19].

Repair of DSBs

Eukaryotic cells have developed at least two types of repair pathways to resolve DSBs: homologous recombination (HR) and nonhomologous end joining (NHEJ) (Fig. 1b). HR is the primordial mechanism of double-strand break repair (DSBR) first discovered as part of meiotic crossover [20]. Homologous regions of the genome (e.g., homologous chromosomes and sister chromatids) serve as a template to prime HR, which can result in error-free replication of the original DNA sequence. In addition to meiotic crossover events in gametes, HR is most prevalent between S and M phases (when sister chromatids are available as a repair template) in eukaryotic diploid cells [21]. DSBR also occurs in the complete absence of homology in a process known as NHEJ. In NHEJ, the exposed DNA ends created by the DSB are directly religated to each other. Because NHEJ does not rely on a homologous chromosome as a

template, it can occur during any phase of the cell cycle. Problematically, NHEJ ligation can occur between any two exposed DNA ends, whether or not they result from the same DSB or whether any nuclease activity has occurred [22]. This lack of specificity can result in a loss of normal genome structure suggesting that NHEJ is an error-prone mechanism. Despite the nonspecificity of DNA end joining, NHEJ is the predominant DSBR mechanism in higher order mammals [23]. While most of the DSBR mechanism can fall under these two headings, HR and NHEJ pathways differ among species and between cells. Starting with NHEJ as the default mechanism in mammals, a brief description of the DSBR molecular pathways is provided below (see also Table 1) [23]. Excellent in-depth descriptions of proteins involved in DSBR can also be found in several recent reviews [24••, 25••, 26].

## Nonhomologous End Joining (NHEJ)

c-NHEJ

After a DSB in mammals, NHEJ is often the initial DSBR mechanism due in part to its fast kinetics [26]. In the canonical, or classic, form of NHEJ (c-NHEJ), DNA ends created by the DSB are bound and processed by heterodimers of the *Ku* proteins (*Ku70* and *Ku80* in humans) [27]. The *Ku* heterodimer binds to and activates $DNA–PK_{cs}$ to form the *DNA–PK* complex. *DNA–PK* is able to bond across the gap between DNA ends, tethering them together and forming the synaptic complex [28]. *DNA–PK* can recruit *ARTEMIS*, an enzyme with both exo- and endonuclease activity, to process non-ligatable DNA ends [29, 30]. Processed DNA ends are then phosphorylated by *DNA–PK*, while the synaptic complex protects against additional nuclease activity [31, 32]. Religation of the DNA ends is performed by the *DNA repair protein* and *DNA ligase IV* (*XRCC4/LIG4*) complex, the activity of which is promoted by XLF to repair the DSB [33–35].

a-NHEJ

When c-NHEJ is unable to process the DNA ends for ligation or Ku binding is inhibited, an alternate form of NHEJ (a-NHEJ, also known as microhomology-mediated end joining or MMEJ) can be employed. a-NHEJ is initiated by *PARP1*, which can also bind directly to DNA ends and compete with *Ku* proteins [36]. *PARP1* is able to modify histones to create a favorable repair environment as well as recruit the *MRN* complex (*Mre11, Rad50, Nbs1*), which in turn displaces the *PARPs* on the DNA ends [37]. *MRN* serves to tether and process the DNA ends and recruit additional DNA repair enzymes including *CtIP* (also

Fig. 1 Schematic of mechanism involved in SVs. **a** Causes of double-strand breaks (DSBs) include several factors including endogenous, exogenous, and programmed sources, and results in breakage of both phosphodiester DNA helix backbones. **b** DSB repair (DSBR) occurs through two major pathways: 1) non-homologous end joining (NHEJ), which occurs throughout the cell cycle; or 2) homologous recombination (HR), which occurs primarily in S and G_2 phases of the cell cycle. In NHEJ, either classical NHEJ (c-NHEJ), where no resectioning has occurred, or alternative NHEJ (a-NHEJ), where minor resectioning has occurred, is employed. HR requires major resectioning and then can repair through single strand annealing (SSA) or form a D-loop which is rectified either by synthesis dependent strand annealing (SDSA) or the canonical HR pathway, which forms a double Holliday junction (dHJ) that can dissolve as a non-crossover event (NCO) or resolve through cleavage in either a NCO or a crossover (CO) event. **c** DSBR can result in no normal repair, where the original chromosomal architecture is intact or result in structural variants (SVs) that can have a combination of losses, gains and/or rearrangements
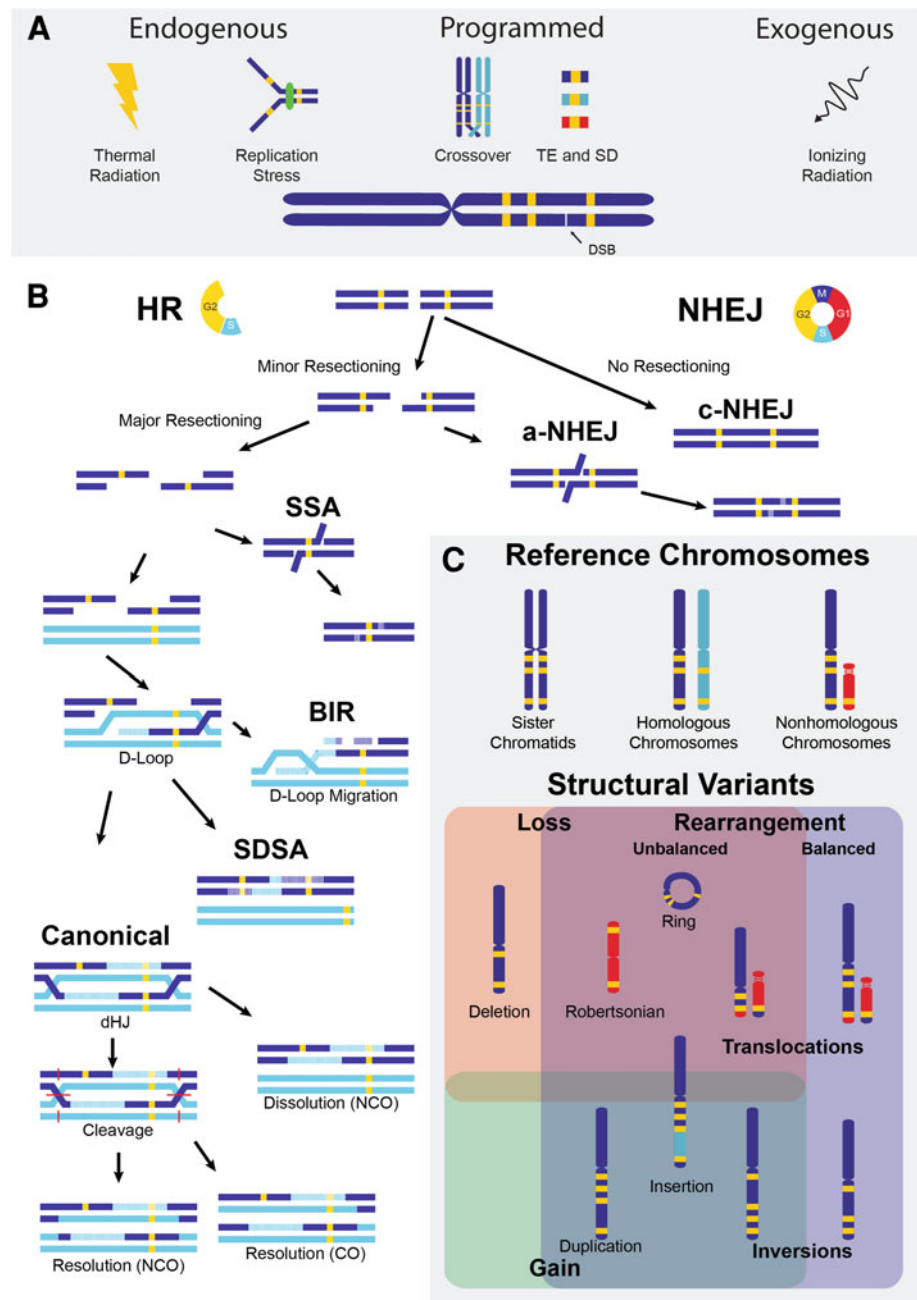
known as *RBBP8*) [38]. *MRN/CtIP* performs exonuclease activity to resect the 5′ strand resulting in an exposed 3′ single stranded DNA (ssDNA) overhang [39, 40]. In a-NHEJ, resection of DSB ends is limited and followed by annealing of the resected ssDNA to each other through microhomology sequences (5–25 nucleotides) [39]. Ligation is then performed in a process similar to single strand break repair mechanisms using the *DNA repair protein XRCC1* and *DNA ligase III* (*XRCC1/LIG3*) complex [41]. Nucleotides 3′ to the microhomology sequence are removed in this process resulting in the loss of several base pairs at the initial DSB.

Homologous Recombination (HR)

Initiation

The HR pathway is initiated after major resectioning of DNA ends to expose ssDNA. Initial binding of DNA ends can occur by *PARPs*, as in a-NHEJ, or by direct binding of the *MRN* complex [42]. *MRN* again binds *CtIP* and this complex promotes nuclease activity [43]. In S and G_2 phases, *BRCA1* binds to *CtIP* and may participate in end resectioning, and when combined with other nucleases like *EXO1* leads to more extensive resection [39, 44–47]. The

Table 1 Essential proteins involved in double-strand break repair

Pathway	Protein(s)	Gene	Function
c-NHEJ	*Ku70/Ku80*	*XRCC6/XRCC5*	End binding and processing
	DNA–PK$_{cs}$	*PRKDC*	End processing, alignment, phosphorylation, and protection
	ARTEMIS	*DCLRE1C*	End processing
	XRCC4/LIG4	*XRCC4/LIG4*	Ligation
	XLF	*NHEJ1*	Promotes ligation
a-NHEJ	*PARP*	*PARP1*	End binding, *MRN* recruitment
	MRN	*MRE11A/RAD50/NBN*	End alignment and resectioning, *BRCA1* recruitment
	CtIP (RBBP8)	*BCL11B*	End resectioning, *BRCA1* recruitment
	XRCC1/LIG3	*XRCC1/LIG3*	Ligation
HR			
Initiation	*MRN*	*MRE11A/RAD50/NBN*	End binding, *MRN* recruitment
	CtIP (RBBP8)	*BCL11B*	End alignment and resectioning, *BRCA1* recruitment
	EXO1	*EXO1*	End resectioning
	BRCA1	*BRCA1*	PALB2 recruitment
	RPA	*RPA1/RPA2/RPA3*	ssDNA binding and stabilization
SSA	*RAD52*	*RAD52*	Annealing of complementary ssDNA
	ERCC1/XPF	*ERCC1/ERCC4*	Endonuclease activity
D-loop	*RAD51/BRCA2/ PALB2*	*RAD51/BRCA2/PALB2*	Stimulate strand invasion
SDSA	*RTEL1*	*RTEL1*	Disassembly of D-loop
C-HR			
	BLM/TOPOIII	*RMI1/RMI2/BLM/ TOP3A*	Unwind HJ in dissolution
	GEN1	*GEN1*	Endonuclease of dHJ in resolution (CO and NCO)
	MUS81/EME1	*MUS81/EME1*	Endonuclease of dHJ in resolution (CO and NCO)
	SLX1/SLX4	*SLX1/SLX4*	Endonuclease of dHJ in resolution (CO and NCO)
	EXO1	*EXO1*	Endonuclease of dHJ in resolution (CO and NCO)
Control			
Kinases	*ATM*	*ATM*	Activates Chk2, p53, H2AX, *MRN*, *ARTEMIS*, *BRCA1*, *BRCA2*
	ATR	*ATR*	Activates Chk1, p53, H2AX, *MRN* RPA
	DNA–PK$_{cs}$	*PRKDC*	Activates *ATM*
Chromatid remodeling	*H2AX/γH2AX*	*H2AFX*	Scaffolding for DSBR chromatid remodeling
Checkpoint	*CHK1*	*CHEK1*	Prevents cell cycle progression
	CHK2	*CHEK2*	Prevents cell cycle progression
Apoptosis	*p53*	*TP53*	Promotes apoptosis

extended ssDNA produced by resection is stabilized by *RPA* to prevent nuclease activity on the exposed 3′ strand [48].

Single Strand Annealing (SSA)

SSA can occur in areas of the genome where DNA repeat sequences (e.g., tandem repeats, interspersed repetitive DNA) are highly concentrated. Rather than a sister chromatid or homologous chromosome, repeat sequences in the ssDNA serve as a template for HR. *Rad52* propagates annealing between the 3′ ssDNA [49–51]. After annealing, a complex of *ERCC1* and *XPF* binds to *Rad52* and cleaves nucleotides 3′ to the repetitive sequences in a process similar to nucleotide excision repair [52, 53].

Displacement Loop (D-loop)

In lieu of SSA, *BRCA2* facilitates *Rad51* binding to a 3′ ssDNA overhang, displacing *RPA* and forming a nucleoprotein filament [54, 55]. The *Rad51* filament facilitates invasion and annealing of the 3′ ssDNA overhang to

homologous sequences and forms a structure known as the D-loop [56]. Once the D-loop has formed, template-based DNA synthesis can be initiated by DNA polymerase [57]. The D-loop can then be resolved through several different paths [56].

Synthesis-Dependent Strand Annealing (SDSA)

In SDSA, the D-loop collapses and the newly synthesized 3′ DNA strand is annealed back to the original DSB. This is mediated by *RETL1*, a helicase, which promotes displacement of the newly synthesized 3′ DNA strand [58]. Once released, the newly synthesized strand anneals back to the reverse complementary exposed ssDNA created from the resection of the origin DSB, repairing one side. The newly synthesized strand can then serve as the template for polymerase-based synthesis on the complementary strand. SDSA does not result in a chromosomal crossover (i.e., exchange of genetic material); however, since the homologous sequence served as a template for synthesis, the sequence at the original DSB will be converted to the homologous sequence, a process known as gene conversion (GC) [59, 60].

Canonical Homologous Recombination (c-HR)

In the c-HR pathway, the newly synthesized DNA is not immediately released. Instead, the complementary 3′ ssDNA also invades the D-loop and begins template-based synthesis on the complementary strands of the homologous sequences, forming a structure known as the double Holiday junction (dHJ). Both synthesizing strands ligate back to the resected 5′ DNA ends from the initial DSB [61]. The dHJ can either resolve in a dissolution state, where the dHJ unwinds and maintains the original chromosomal architecture, or in a resolution state, where the ends of the dHJ are cleaved and genetic material is transferred between the homologous sequences. Dissolution is mediated by a complex of *BLM* and *TOPOIIIα* proteins and can result in GC without exchange of flanking sequences [62]. Resolution of the dHJ does result in exchange of flanking materials and is performed by endonucleases such as *MUS81, EXO1, EME1,* and *SLX4/SLX1*. Resolution can result in GC and both crossover and non-crossover events [63].

Break Induced Repair (BIR)

BIR is a form of DSBR most often associated with replication stress and shares many features in common with HR. In BIR only one end of the DSB is involved. As in HR, the 3′ ssDNA overhang invades the homologous sequence forming a D-loop and primes DNA synthesis by DNA polymerase. The D-loop then migrates along the chromosome as the 3′ end extends. DNA synthesis in the opposite direction can then occur either on the newly synthesized leading strand or on the complementary ssDNA exposed by the D-loop. BIR, and a form of BIR that relies on microhomology (MMBIR), is often associated with breaks that occur during DNA replication in S phase. The DNA replication fork can stall due to such factors as errant nucleotide repair, single strand break, and helicase inhibition. Replication is restarted by inducing a DSB that is repaired by BIR [64–66]. A similar mechanism, known as fork stalling and template switching (FoSTeS), also suggests a mechanism for restarting replication forks, but without a DSB [67, 68].

Control of DSBR

Coordination of DSBR is essential for proper repair and prevention of further damage. After a DSB, the cell delays progression through the cell cycle, to allow time for DSBR, or, if the damage is too extensive, signals apoptosis. Two master control kinases are employed to organize these events: *ATM* and *ATR*. *ATM* can activate both *p53*, a pro-apoptotic factor, and *CHK2*, which delays cell cycle progression. *ATR* delays cell cycle progression through the *CHK1* cascade. Both *ATM* and *ATR* have extensive phosphorylation cascades involving essential DSBR proteins including *ARTEMIS*, *BRCA1*, *BRCA2*, and *MRN* subunits [69, 70]. *ATM* and *ATR*, as well as *DNA–PK$_{cs}$*, can also phosphorylate *H2AX*, converting it to its active state γ-*H2AX*. γ-*H2AX* serves as a focal point in DSBR through chromatid remodeling and as a platform to assemble other enzymes involved in DSBR [71]. *ATM* is one of the first proteins activated in the DSBR pathway and though initial activation is still unclear, *ATM* is phosphorylated by the *Nbs1* subunit of the *MRN* complex during early stages of a-NHEJ and HR [69]. *ATR* is activated later through an interaction with *RPA* proteins associated with ssDNA [70]. Many other proteins are involved in the control of DSBR, depending on cell stage, age, and type, and many of the proteins involved in these pathways are still being elucidated.

SVs Resulting from DSBs

For the most part, DSBR results in faithful reproduction of the original genetic architecture before the DSB. In some cases, errors do occur in DSBR that may result in many different forms of SVs. Several SV classification systems have been generated depending on the mode of detection. Microscopic techniques (e.g., karyotyping) describe SVs at the resolution of chromosomal banding patterns (>3–10 Mb, depending on banding resolution), including

large deletions, insertions, and rearrangements, but are limited in resolution. Higher resolution molecular techniques (e.g., array comparative genome hybridization or aCGH, and single nucleotide polymorphism arrays or SNP arrays) describe SVs characterized by genomic imbalance, or copy number variations (CNVs), but are often unable to detect dosage imbalances below the resolution of probe density, and are blind to some rearrangements of chromosomes [72•, 73, 74]. As NGS techniques are developed offering both architecture characterization and high resolution, these traditional classification systems are being redefined [75]. A comprehensive classification system can describe SVs by (1) losses, (2) gains, or (3) rearrangements of segments of nucleotides compared to the reference genome (Fig. 1c). Rearrangements can be further subdivided into inversions, when the segment is in an opposite orientation, and translocations, when a genomic segment is displaced to a different part of the genome. Many traditional classification systems fit within this rubric. For example, deletions, copy number losses, microdeletions, and indel losses are all SV losses, while duplications, copy number gains, microduplications, and indel gains are all SV gains. DSBs that result in rearrangement SVs can also be accompanied by gains or losses and can be considered unbalanced. Some examples of unbalanced rearrangements include many forms of nonhomologous translocations as well as translocations caused by telomere erosion such as ring chromosomes, intrachromosomal ligation of telomeres, or Robertsonian translocations, the ligation of beta-satellite sequences after loss of the majority of the short arm in acrocentric chromosomes (i.e., human chromosomes 13–15, 21, and 22). More complex SVs may be characterized by all three properties; for example, an insertion can arise from the duplication of a segment of one chromosome (SV gain) that is inserted into DNA sequence of a separate chromosome (SV rearrangement) replacing the original DNA sequence (SV loss). As the detection of SVs by NGS now offers nucleotide resolution, more definitive classification systems are required to describe accurately the losses, gains, and rearrangements that result from SVs.

Gains, losses, and rearrangements can arise due to errors in any one of the DSBR mechanisms; however, certain mechanisms are more prone to certain types of errors than others. NHEJ is often considered an error-prone mechanism because no template is used in DSBR. Instead, any DSB end protected by *Ku* heterodimers can be used for ligation in c-NHEJ. Even in the presence of a single DSB, involving only two *Ku* protected ends, nucleotides may be lost either through the initial break event or during end processing by *ARTEMIS*. If the presence of multiple DSBs, c-NHEJ will ligate *Ku* protected ends regardless of whether they originated from the same break or from different

chromosomes resulting in dramatic changes to chromosomal architecture, including rearrangements [22, 76]. Although microhomology is used in a-NHEJ, this mechanism typically results in SVs. Losses will almost always be generated as nucleotides are excised 3′ to the microhomology site. In addition, microhomology sites can bind promiscuously and while advantageous in cases of simple DSBs, in the presence of multiple DSBs, a-NHEJ can result in many forms of rearrangements [40, 77]. On the other end of the spectrum, HR is considered to be a high fidelity mechanism, because homologous sequences serve as a template for repair. This is often true when sister chromatids, which are faithful duplications of each other, serve as the homologous template. When homologous chromosomes are used as a template in HR, differences (e.g., single nucleotide variants, or SNVs) between the invading strand and template will result in mismatches, which are rectified by DNA repair mechanisms. This results in a loss of heterozygosity between the homologous chromosomes and is referred to as GC. The effects are more dramatic when templates are used that share high sequence homology, but are non-allelic. This non-allelic homologous recombination (NAHR) can occur on templates that are within the same chromosome, in sister chromatids, in homologous sequences, or in entirely different chromosomes. NAHR can result in complex SVs with unbalanced rearrangements resulting in losses or gains of entire chromosomal arms. NAHR is likely to occur if HR is employed during G_0 and G_1 phases of the cell cycle, when sister chromatids are not available [26, 78]. The rate of DSB and repair mechanism are also influenced by the development state and cell type.

DSBs can occur at any point during development from pre-conception to adult. During meiosis prophase I DSBs are repaired by HR. In mammals, as the cell begins rapidly dividing, NHEJ becomes the predominant DSBR throughout the cell cycle [21, 23]. After mitosis, the cell enters G_1 and NHEJ is the dominant DSBR mechanism [23]. HR in G_1 is likely suppressed in that there are no sister chromatids available, meaning any HR would result in either GC of homologous chromosomes or SVs resulting from NAHR. As cells progress towards S/G_2 phase, and sister chromatids become available, HR is upregulated. *CtIP* modification, by *ATM* and *BRCA1*, may play a central role in this transition from NHEJ to HR in the cell cycle [79, 80]. The DSBR pathway is also regulated dependent upon cell type; for example, lymphocytic stem cells create genetic diversity in immunoglobulins and T cell receptors by self-induced DSB that are repaired by NHEJ (see below) [81]. In later stages, as cells age, SVs may begin to accumulate and may become oncogenic [82]. Given the frequency of DSB and the errors that can occur, DSBR mechanisms are tightly controlled pathways; however, in

certain processes within eukaryotic cells, DSBs are actually self-induced.

Random, Recurrent, and Programmed DSB

For the most part, DSBs are caused by random events at arbitrary locations. However, there is evidence that some areas of the genome are more susceptible to DSBs, areas known as recombination hotspots. Eukaryotic cells are also capable of self-inducing DSBs to generate genetic diversity.

Recombination Hotspots

Recombination hotspots are regions of elevated occurrences of SVs, which include fragile sites, segmental duplications (SDs), and transposable elements (TEs). Fragile sites are areas of the chromosome that are prone to breaks, gaps, and constrictions in metaphase chromosomes. These sites are thought to be prone to DSBs due to replication stresses such as fork slippage, caused by repetitive sequences, or fork stalling, caused by DNA hairpin formation. SDs, also known as low copy repeats (LCR), are areas of >10 kb that share high sequence homology. Regions enriched for SDs are associated with high occurrences of SVs due in part to their shared homology, which can serve as templates for NAHR, but are also known to undergo high rates of NHEJ. Interestingly, there has been a rapid expansion of SDs in the evolution of primates, suggesting duplications of duplications, despite their role in inducing SVs [83]. TEs are sequences that have likely originated from viral integration and are capable of moving around the genome. TEs can be classified as class I "copy-and-paste" elements, requiring an RNA intermediate, and class II "cut-and-paste" elements, requiring no RNA intermediate. Due to their unique properties, TEs are involved in both NHEJ and NAHR. While being excised or reintegrating following retrotransposition, TEs can induce DSBs that follow NHEJ repair; however, due to their homology they can also serve as a template for NAHR. Although both SD and TE loci are often associated with disease, theories suggest that these hotspots have promoted rapid evolution in the human genome [84, 85•, 86, 87•].

Programmed Rearrangements

Eukaryotic cells have evolved to take advantage of what in other circumstances would be considered an error in DSBR. In germline cells, SVs are produced, which may help create population diversity within a species. In prophase I of meiosis, DSBs are also induced by *SPO11*, which are repaired through HR. HR may result in crossing over between homologous chromosomes and unique segregation of genes not possible from the chromosomal structure of the parental genomes [88]. Only a small fraction of the *SPO11*-induced breaks will result in crossovers between homologous chromosomes while most will result in pairing with sister chromatids and/or non-crossover events. Meiotic crossover hotspots are also evident, but it is unclear what controls the *SPO11* susceptibility and crossover regulation [89]. Meiotic crossovers can also be repaired improperly specifically as a result of NAHR, creating de novo SVs that can range from inert to embryonic lethal. During development, more restricted induced DSBs occur in immunological cells, creating antibody diversity that is essential for protecting against viral and bacterial infection. During development of lymphocytes, V(D)J (for variable, diverse, joining) and class-switch DNA recombination (CSR) create recombinations of regions of the immunoglobulin and T cell receptor to create diversity. In V(D)J recombination, DSBs are induced in the variable regions of both immunoglobulin heavy and light chains as well as the T cell receptor by *RAG1* and *RAG2* at conserved AT-rich heptamer sequences, and these breaks are subsequently repaired through NHEJ [90•]. In CSR, DSBs are induced in the constant region of the immunoglobulin locus by *AID* and the break is subsequently repaired by either c-NHEJ or a-NHEJ [91••]. While V(D)J and CSR are able to create an immunoglobulin repertoire in the tens of billions, this induced recombination likely arose out of viral elements similar to TE and is not without its consequences as a possible contributor to lymphoid malignancy [81]. Recent reports suggest that induced DSBs may not be restricted to meiotic and lymphocyte cells, but may also contribute to the diversity of neurons [92, 93].

Complex SVs

Next-generation sequencing of the genomic architecture of cancers and human neurodevelopmental disorders has revealed complex structural variants (also known as complex genomic rearrangements or CGRs) involving a combination of two or more deletions, duplications, insertions, or inversions over one or more chromosomes. In one study, most of the complex events detected involved one or more cryptic inversions at the rearrangement breakpoints [94•]. The most striking example of complex SVs is a phenomenon known as chromothripsis (Greek, *thripsis*: shattering), whereby chromosomes are fragmented into tens to hundreds of segments and rearranged with or without accompanying alterations in copy number. First described in chronic lymphocytic leukemia (CLL), it is now believed that 2–3 % of all cancers and up to 25 % of bone cancers harbor chromothripsis events [95–98]. Germline chromothripsis

has also been reported in cases of developmental delay or cognitive defects [78, 94•, 99]. In some cancers, this extensive genomic rearrangement likely occurs as a single event of breakage and repair rather than as the aggregation of numerous structural insults, evidenced by copy neutral states, regional clustering, and interspersed loss of heterozygous segments in the derivative chromosomes [100••]. However, the trauma and repair of chromosomal shattering may occur through multiple mechanisms, and this simultaneity does not necessarily characterize complex SV cases that are accompanied by extensive CNV [78]. While the mechanisms underlying the shattering and religation of catastrophic chromosomal events remains an area of active investigation, several theories have been proposed. Chromosomal shattering may occur by ionizing radiation or free radical oxidation to condensed mitotic figures, by breakage of dicentric chromosomes during anaphase, by recovery of a cell after an aborted apoptosis, or by formation of micronuclei containing anaphase lagging chromosomes [101, 102•, 103, 104]. Following the genomic breakage, data from Chiang et al. [94•] and others suggest that events studied to date are repaired with little to no homologous sequence at the breakpoints, indicating a predominance of error-prone NHEJ repair [94•, 98–100••]. Nonetheless, templated homology- or microhomology-based repair (e.g., MMBIR and FoSTeS), likely play a role in a subset of chromosomal shattering events, particularly those that are accompanied by copy number changes [78].

Summary

Although SVs have been visible under the microscope for several decades, only recently have some of the complexities that underlie these changes to the genomic structure become apparent. It has become evident that many SVs stem from errors in the repair of DSBs, but the molecular mechanism of repair and the cause of these errors are still being deciphered. Even more confounding, some SVs are induced by the cell itself to produce genetic diversity within the organism as well as the population. The advent of next-generation sequencing technologies has offered new perspectives on the normal and pathological architecture of the genome. These tools have allowed us to recognize that SVs are not only common, but also make up a significant proportion of human variation.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance:
•• Of major importance

1. Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993;362(6422):709–15.
2. Cathcart R, et al. Thymine glycol and thymidine glycol in human and rat urine: a possible assay for oxidative DNA damage. Proc Natl Acad Sci USA. 1984;81(18):5633–7.
3. Lindahl T, Nyberg B. Rate of depurination of native deoxyribonucleic acid. Biochemistry. 1972;11(19):3610–8.
4. Vilenchik MM, Knudson AG. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. Proc Natl Acad Sci USA. 2003;100(22):12871–6.
5. Agnez-Lima LF, et al. DNA damage by singlet oxygen and cellular protective mechanisms. Mutat Res. 2012;741(1):15–28.
6. Imlay JA, Linn S. DNA damage and oxygen radical toxicity. Science. 1988;240(4857):1302–9.
7. Prise KM, Davies S, Michael BD. Cell killing and DNA damage in Chinese hamster V79 cells treated with hydrogen peroxide. Int J Radiat Biol. 1989;55(4):583–92.
8. Kuzminov A. Collapse and repair of replication forks in *Escherichia coli*. Mol Microbiol. 1995;16(3):373–84.
9. Cox MM. Recombinational DNA repair of damaged replication forks in *Escherichia coli*: questions. Annu Rev Genet. 2001;35:53–82.
10. Kuzminov A. Single-strand interruptions in replicating chromosomes cause double-strand breaks. Proc Natl Acad Sci USA. 2001;98(15):8241–6.
11. Keeney S. Mechanism and control of meiotic recombination initiation. Curr Top Dev Biol. 2001;52:1–53.
12. Roth DB, et al. V(D)J recombination in mouse thymocytes: double-strand breaks near T cell receptor delta rearrangement signals. Cell. 1992;69(1):41–53.
13. Sutherland BM, et al. Clustered damages and total lesions induced in DNA by ionizing radiation: oxidized bases and strand breaks. Biochemistry. 2000;39(27):8026–31.
14. Povirk LF. DNA damage and mutagenesis by radiomimetic DNA-cleaving agents: bleomycin, neocarzinostatin and other enediynes. Mutat Res. 1996;355(1–2):71–89.
15. Ward JF. DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability. Prog Nucleic Acid Res Mol Biol. 1988;35:95–125.
16. Huang CH, et al. Single-strand and double-strand deoxyribonucleic acid breaks produced by several bleomycin analogues. Biochemistry. 1981;20(2):233–8.
17. Gasior SL, et al. The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol. 2006;357(5):1383–93.
18. Bennett CB, et al. Lethality induced by a single site-specific double-strand break in a dispensable yeast plasmid. Proc Natl Acad Sci USA. 1993;90(12):5613–7.
19. Rich T, Allen RL, Wyllie AH. Defying death after DNA damage. Nature. 2000;407(6805):777–83.

20. Eisen JA, Hanawalt PC. A phylogenomic study of DNA repair genes, proteins, and processes. Mutat Res. 1999;435(3): 171–213.

21. Rothkamm K, et al. Pathways of DNA double-strand break repair during the mammalian cell cycle. Mol Cell Biol. 2003;23(16):5706–15.

22. Lieber MR, et al. Nonhomologous DNA end joining (NHEJ) and chromosomal translocations in humans. Subcell Biochem. 2010;50:279–96.

23. Mao Z, et al. Comparison of nonhomologous end joining and homologous recombination in human cells. DNA Repair (Amst). 2008;7(10):1765–71.

24. •• Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. Mol Cell. 2010;40(2):179–204. *This review examines pathways, control, and error associated with DSBR.*

25. •• Thompson LH. Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: the molecular choreography. Mutat Res. 2012;751(2):158–246. *This review gives an in-depth look at the molecular mechanisms of DSBR.*

26. Neal JA, Meek K. Choosing the right path: does DNA–PK help make the decision? Mutat Res. 2011;711(1–2):73–86.

27. Roberts SA, et al. Ku is a 5′-dRP/AP lyase that excises nucleotide damage near broken ends. Nature. 2010;464(7292): 1214–7.

28. Meek K, et al. Trans autophosphorylation at DNA-dependent protein kinase's two major autophosphorylation site clusters facilitates end processing but not end joining. Mol Cell Biol. 2007;27(10):3881–90.

29. Ma Y, et al. The DNA-dependent protein kinase catalytic subunit phosphorylation sites in human *ARTEMIS*. J Biol Chem. 2005;280(40):33839–46.

30. Goodarzi AA, et al. DNA–PK autophosphorylation facilitates *ARTEMIS* endonuclease activity. EMBO J. 2006;25(16):3880–9.

31. Mahaney BL, Meek K, Lees-Miller SP. Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. Biochem J. 2009;417(3):639–50.

32. Meek K, Dang V, Lees-Miller SP. DNA–PK: the means to justify the ends? Adv Immunol. 2008;99:33–58.

33. Yu Y, et al. DNA–PK and ATM phosphorylation sites in XLF/Cernunnos are not required for repair of DNA double strand breaks. DNA Repair (Amst). 2008;7(10):1680–92.

34. Wang YG, et al. Phosphorylation and regulation of DNA ligase IV stability by DNA-dependent protein kinase. J Biol Chem. 2004;279(36):37282–90.

35. Leber R, et al. The XRCC4 gene product is a target for and interacts with the DNA-dependent protein kinase. J Biol Chem. 1998;273(3):1794–801.

36. Wang M, et al. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res. 2006;34(21):6170–82.

37. De Vos M, Schreiber V, Dantzer F. The diverse roles and clinical relevance of PARPs in DNA damage repair: current state of the art. Biochem Pharmacol. 2012;84(2):137–46.

38. Williams GJ, Lees-Miller SP, Tainer JA. Mre11-Rad50-Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks. DNA Repair (Amst). 2010;9(12):1299–306.

39. You Z, Bailis JM. DNA damage and decisions: CtIP coordinates DNA repair and cell cycle checkpoints. Trends Cell Biol. 2010;20(7):402–9.

40. McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. 2008;24(11):529–38.

41. Audebert M, Salles B, Calsou P. 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. 2004;279(53):55117–26.

42. de Jager M, et al. DNA-binding and strand-annealing activities of human Mre11: implications for its roles in DNA double-strand break repair pathways. Nucleic Acids Res. 2001;29(6): 1317–25.

43. Limbo O, et al. Ctp1 is a cell-cycle-regulated protein that functions with Mre11 complex to control double-strand break repair by homologous recombination. Mol Cell. 2007;28(1): 134–46.

44. Yu X, Chen J. DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of *BRCA1* C-terminal domains. Mol Cell Biol. 2004;24(21):9478–86.

45. Huen MS, Sy SM, Chen J. *BRCA1* and its toolbox for the maintenance of genome integrity. Nat Rev Mol Cell Biol. 2010;11(2):138–48.

46. Nimonkar AV, et al. Human exonuclease 1 and BLM helicase interact to resect DNA and initiate DNA repair. Proc Natl Acad Sci USA. 2008;105(44):16906–11.

47. Yun MH, Hiom K. CtIP-*BRCA1* modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. Nature. 2009;459(7245):460–3.

48. Wold MS. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annu Rev Biochem. 1997;66:61–92.

49. Lin FL, Sperle K, Sternberg N. Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. Mol Cell Biol. 1984; 4(6):1020–34.

50. Kass EM, Jasin M. Collaboration and competition between DNA double-strand break repair pathways. FEBS Lett. 2010;584(17):3703–8.

51. Mortensen UH, et al. DNA strand annealing is promoted by the yeast Rad52 protein. Proc Natl Acad Sci USA. 1996;93(20): 10729–34.

52. Motycka TA, et al. Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. J Biol Chem. 2004;279(14):13634–9.

53. Adair GM, et al. Role of ERCC1 in removal of long non-homologous tails during targeted homologous recombination. EMBO J. 2000;19(20):5552–61.

54. Jensen RB, Carreira A, Kowalczykowski SC. Purified human *BRCA2* stimulates *RAD51*-mediated recombination. Nature. 2010;467(7316):678–83.

55. Benson FE, Stasiak A, West SC. Purification and characterization of the human *Rad51* protein, an analogue of *E. coli* RecA. EMBO J. 1994;13(23):5764–71.

56. San Filippo J, Sung P, Klein H. Mechanism of eukaryotic homologous recombination. Annu Rev Biochem. 2008;77: 229–57.

57. West SC. Molecular views of recombination proteins and their control. Nat Rev Mol Cell Biol. 2003;4(6):435–45.

58. Barber LJ, et al. RTEL1 maintains genomic stability by suppressing homologous recombination. Cell. 2008;135(2):261–71.

59. Formosa T, Alberts BM. DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. Cell. 1986;47(5):793–806.

60. Chen JM, et al. Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet. 2007;8(10):762–75.

61. Maher RL, Branagan AM, Morrical SW. Coordination of DNA replication and recombination activities in the maintenance of genome stability. J Cell Biochem. 2011;112(10):2672–82.

62. Wu L, Hickson ID. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. Nature. 2003;426(6968):870–4.
63. Zakharyevich K, et al. Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase. Cell. 2012;149(2):334–47.
64. Llorente B, Smith CE, Symington LS. Break-induced replication: what is it and what is it for? Cell Cycle. 2008;7(7):859–64.
65. Smith CE, Llorente B, Symington LS. Template switching during break-induced replication. Nature. 2007;447(7140):102–5.
66. Hastings PJ, Ira G, Lupski JR. A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet. 2009;5(1):e1000327.
67. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell. 2007;131(7):1235–47.
68. Zhang F, Carvalho CM, Lupski JR. Complex human chromosomal and genomic rearrangements. Trends Genet. 2009;25(7):298–307.
69. Derheimer FA, Kastan MB. Multiple roles of ATM in monitoring and maintaining DNA integrity. FEBS Lett. 2010;584(17):3675–81.
70. Cimprich KA, Cortez D. ATR: an essential regulator of genome integrity. Nat Rev Mol Cell Biol. 2008;9(8):616–27.
71. Celeste A, et al. Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. Nat Cell Biol. 2003;5(7):675–9.
72. • Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006;7(2):85–97. *This review examines the types of structural variants found in the human genome and the methods used to detect them.*
73. Dugoff L. Application of genomic technology in prenatal diagnosis. N Engl J Med. 2013;367(23):2249–51.
74. Shaffer LG, McGowan-Jordan J, Schmid M, editors. ISCN 2013: An International System for Human Cytogenetic Nomenclature (2013). 1st ed. Basel: S. Karger AG; 2012. p. 140.
75. Talkowski ME, et al. Clinical diagnosis by whole-genome sequencing of a prenatal sample. N Engl J Med. 2013;367(23):2226–32.
76. Weinstock DM, Elliott B, Jasin M. A model of oncogenic rearrangements: differences between chromosomal translocation mechanisms and simple double-strand break repair. Blood. 2006;107(2):777–80.
77. Simsek D, Jasin M. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. Nat Struct Mol Biol. 2010;17(4):410–6.
78. Liu P, et al. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. Cell. 2011;146(6):889–903.
79. Chapman JR, Taylor MR, Boulton SJ. Playing the end game: DNA double-strand break repair pathway choice. Mol Cell. 2012;47(4):497–510.
80. Brandsma I, Gent DC. Pathway choice in DNA double strand break repair: observations of a balancing act. Genome Integr. 2012;3(1):9.
81. Ramsden DA, Weed BD, Reddy YV. V(D)J recombination: born to be wild. Semin Cancer Biol. 2010;20(4):254–60.
82. Zhang Y, et al. The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. Adv Immunol. 2010;106:93–133.
83. Johnson ME, et al. Recurrent duplication-driven transposition of DNA during hominoid evolution. Proc Natl Acad Sci USA. 2006;103(47):17626–31.
84. Eichler EE. Segmental duplications: what's missing, misassigned, and misassembled–and should we care? Genome Res. 2001;11(5):653–6.
85. • Kim YJ, Lee J, Han K. Transposable elements: no more 'junk DNA'. Genomics Inform. 2012;10(4):226–33. *Recent review of TE and their involvement in primate evolution.*
86. Marques-Bonet T, Eichler EE. The evolution of human segmental duplications and the core duplicon hypothesis. Cold Spring Harb Symp Quant Biol. 2009;74:355–62.
87. • Marques-Bonet T, Girirajan S, Eichler EE. The origins and impact of primate segmental duplications. Trends Genet. 2009;25(10):443–54. *This review examines SD as a driving factor in human evolution.*
88. Zickler D, Kleckner N. Meiotic chromosomes: integrating structure and function. Annu Rev Genet. 1999;33:603–754.
89. Serrentino ME, Borde V. The spatial regulation of meiotic recombination hotspots: are all DSB hotspots crossover hotspots? Exp Cell Res. 2012;318(12):1347–52.
90. • Malu S, et al. Role of non-homologous end joining in V(D)J recombination. Immunol Res. 2012;54(1–3):233–46. *Recent review of V(D)J recombination.*
91. •• Xu Z, et al. Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. Nat Rev Immunol. 2012;12(7):517–31. *This review examines the molecular mechanisms involved in CSR.*
92. Evrony GD, et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell. 2012;151(3):483–96.
93. Muotri AR, Gage FH. Generation of neuronal variability and complexity. Nature. 2006;441(7097):1087–93.
94. • Chiang C, et al. Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. Nat Genet. 2012;44(4):390–7, S1. *This report indicates that NHEJ and FoSTeS are involved in chromothripsis repair.*
95. Forment JV, Kaidi A, Jackson SP. Chromothripsis and cancer: causes and consequences of chromosome shattering. Nat Rev Cancer. 2012;12(10):663–70.
96. Magrangeas F, et al. Chromothripsis identifies a rare and aggressive entity among newly diagnosed multiple myeloma patients. Blood. 2011;118(3):675–8.
97. Molenaar JJ, et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature. 2012;483(7391):589–93.
98. Kloosterman WP, et al. Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. Genome Biol. 2011;12(10):R103.
99. Kloosterman WP, et al. Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. Hum Mol Genet. 2011;20(10):1916–24.
100. •• Stephens PJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011;144(1):27–40. *This paper suggests that chromothripsis may arise from a single event producing many DSB rather than multiple individual events producing single DSBs.*
101. Jones MJ, Jallepalli PV. Chromothripsis: chromosomes in crisis. Dev Cell. 2012;23(5):908–17.
102. • Crasta K, et al. DNA breaks and chromosome pulverization from errors in mitosis. Nature. 2012;482(7383):53–8. *This paper suggests that chromothripsis may arise from error in mitosis during chromosomal segregation.*
103. McClintock B. The stability of broken ends of chromosomes in *Zea mays*. Genetics. 1941;26(2):234–82.
104. Tubio JM, Estivill X. Cancer: when catastrophe strikes a cell. Nature. 2011;470(7335):476–7.