

Mechanisms of change in gene copy number

P. J. Hastings*, James R. Lupski*^{†§}, Susan M. Rosenberg*^{||¶#} and Grzegorz Ira*

Abstract | Deletions and duplications of chromosomal segments (copy number variants, CNVs) are a major source of variation between individual humans and are an underlying factor in human evolution and in many diseases, including mental illness, developmental disorders and cancer. CNVs form at a faster rate than other types of mutation, and seem to do so by similar mechanisms in bacteria, yeast and humans. Here we review current models of the mechanisms that cause copy number variation. Non-homologous end-joining mechanisms are well known, but recent models focus on perturbation of DNA replication and replication of non-contiguous DNA segments. For example, cellular stress might induce repair of broken replication forks to switch from high-fidelity homologous recombination to non-homologous repair, thus promoting copy number change.

Array comparative genomic hybridization

A microarray-based technique to measure the relative amount of any DNA sequence.

Paired-end mapping

A technique whereby novel linkage relationships are detected by finding short sequences linked to other short sequences in DNA fragments of uniform size.

*Department of Molecular and Human Genetics,

[†]Department of Pediatrics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

[§]Texas Children's Hospital, 1102 Bates, Houston, Texas 77030, USA.

^{||}Department of Biochemistry and Molecular Biology,

[¶]Department of Molecular Virology and Microbiology, [#]The Dan L. Duncan Cancer Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

Correspondence to P.J.H. email: hastings@bcm.edu

doi:10.1038/nrg2593

Published online 14 July 2009

Human populations show extensive polymorphism — both additions and deletions — in the number of copies of chromosomal segments, and the number of genes in those segments^{1–6}. This is known as copy number variation. A high proportion of the genome, currently estimated at up to 12%, is subject to copy number variation⁴. Copy number variants (CNVs) can arise both meiotically and somatically, as shown by the finding that identical twins can have different CNVs⁷ and that repeated sequences in different organs and tissues from the same individual can vary in copy number⁸. Copy number variation seems to be at least as important as SNPs in determining the differences between individual humans⁹ and seems to be a major driving force in evolution, especially in the rapid evolution that has occurred, and continues to occur, within the human and great ape lineage^{10–14}. Changes in copy number might change the expression levels of genes included in the regions of variable copy number, allowing transcription levels to be higher or lower than those that can be achieved by control of transcription of a single copy per haploid genome. Possible adaptive advantages of copy number variation are discussed in BOX 1. Additional copies of genes also provide redundancy that allows some copies to evolve new or modified functions or expression patterns while other copies maintain the original function^{15,16}. The non-homologous recombination events that underlie changes in copy number also allow generation of new combinations of exons between different genes by translocation, insertion or deletion^{17,18}, so that proteins might acquire new domains, and hence new or modified activities.

However, much of the variation in copy number is disadvantageous. Change in copy number is involved in cancer formation and progression^{19,20}, and contributes to cancer proneness²¹. In many situations, a change in copy number of any one of many specific genes is not well tolerated, and leads to a group of pathological conditions known as genomic disorders²². Because particular gene imbalances are associated with specific clinical syndromes, data on rare clinical cases of change in copy number are available and have facilitated the study of the chromosomal changes underlying copy number variation. Further examples have come from studies of complete genomes, and from genome-wide surveys of CNVs using techniques such as array comparative genomic hybridization⁴, comparison of expression levels²³ or paired-end mapping³.

Mechanisms of chromosomal structural change have been studied in model organisms, notably *Saccharomyces cerevisiae*, *Escherichia coli* and *Drosophila melanogaster*. By bringing together the findings from model organisms with the characteristics of copy number variation in human and primate genomes, we can begin to work towards an understanding of the processes that lead to chromosomal structural change, and thus gain insights into a major driving force of human evolution²⁴. Extrapolation from one organism to another is not always reliable, but it has proved very successful in the study of processes acting on DNA; almost all DNA repair mechanisms acting in humans were first described in model organisms, particularly bacteria²⁵.

Non-allelic homologous recombination
Homologous recombination between lengths of homology in different genomic positions.

In this Review, we describe the properties of CNVs and the mechanisms that lead to change in copy number, including homologous recombination (HR) and non-homologous repair mechanisms. We suggest that replicative mechanisms might be particularly important, and discuss a potential relationship between cellular stress and copy number change, which could have implications for understanding genome evolution and human disease.

Characteristics of CNVs

A change in copy number requires a change in chromosome structure, joining two formerly separated DNA sequences. These junctions give important insights into how the structural change has arisen. Many structural changes show recurrent end-points; that is, most events at a given locus have their junctions confined to a few genomic positions. The junctions of these recurrent

CNVs are found to be in low copy repeats (LCRs) that provide extensive homology. LCRs, also called segmental duplications, are sequences that occur twice or a few times in a haploid genome. For practical purposes, the definition is limited by degree of identity (commonly >95%) and length (usually >1 kb). It is likely that recurrent CNVs arose by HR between repeated sequences. This process is called non-allelic homologous recombination (NAHR; also known as ectopic HR), and is discussed below.

Other structural changes show non-recurrent end-points. Most non-recurrent CNVs occur at sites of limited homology of 2–15 bp (microhomology) (for examples, see REFS 26–28) — much too short to have occurred by HR, as discussed below. A second characteristic of non-recurrent events is that chromosomal structural changes can be complex: they can have short sequences from elsewhere inserted at the junctions; they can include a mixture of duplications, triplications, inversions and deletions; or they can be interspersed with lengths of unchanged sequence^{29–32}. An example of a complex rearrangement with microhomology junctions is shown in FIG. 1. A third characteristic is that, although the non-recurrent junctions do not coincide with LCRs, they tend to occur in the vicinity of regions that are rich in LCRs — either direct or inverted repeats — resulting in complex regional genomic architecture^{33–35}. The origin of LCRs and of regions in which LCRs are prevalent is presumably the same as the origin of the non-recurrent events that we witness today; hence the mechanisms of non-recurrent copy number change are the mechanisms of evolution of genomes.

Box 1 | How much copy number variation is adaptive?

Initially, copy number variation seemed to be advantageous because the set of genes that were found to vary in copy number is enriched for genes involved in olfaction, immunity and secreted proteins — that is, genes relevant to the immediate environment¹²⁰. These genes were reported to be under recent selection because they contain higher than average frequencies of non-synonymous mutations¹²⁰. Now, alternative explanations have been offered for these features¹³⁸. The genomic regions where copy number variants (CNVs) are found might be those where copy number is less important than it is in other genomic regions. In other words, most copy number variation is rapidly purged from the population, but purifying selection is weaker in some regions and these regions are where CNVs are retained. Relatively weak purifying selection would also explain why a higher mutation frequency is observed in CNV-rich regions than in the genome as a whole, because even loss-of-function alleles in these regions would be purged from the population more slowly¹³⁸.

Evidence that a variable copy number of specific genes offers a selective advantage is sparse. The salivary amylase gene, *AMY1*, shows copy number variation in human populations¹³⁹, and the amount of salivary amylase is directly proportional to the copy number of *AMY1*. The average number of copies of *AMY1* is higher in cultures that consume a high level of starch than in cultures that consume little starch¹³⁹, suggesting that a high copy number of *AMY1* is advantageous in cultures with a high starch intake and neutral in cultures with low starch intake¹³⁹. A second possible example is the correlation between the copy number of the chemokine gene *CCL3L1* and susceptibility to HIV/AIDS¹⁴⁰. An example of a reduction in copy number being beneficial has been suggested for the α -globin locus, because the disadvantages of an α -globin gene deletion in homozygotes might be balanced by resistance to malaria for heterozygotes (reviewed in REF. 141). Other examples are certain to be found, but most of the extensive copy number variation in humans seems to be non-adaptive or disadvantageous, and is present only because it has not yet been purged. Approximately 75% of copy number variation is found at a frequency of less than 3% in human populations⁴, suggesting a stochastic origin and maintenance of most of this variation¹⁴² (similar results in REFS 143–145). Furthermore, many private CNVs (those specific to a family) have been described, and there is limited overlap in lists of CNVs found in different genome-wide searches². Many CNVs have gone to fixation in the human and other primate lines, and are now seen as low copy repeats^{10,12,13,146}. The low frequency in populations of almost every CNV polymorphism that has been described suggests that few of the CNVs that we see now are tending towards fixation.

However, the rarity of specific CNVs described above has been challenged by a study showing that 80% of CNVs are found in more than 5% of individuals¹⁴⁷. Such discrepancies might arise owing to uncertainty in the ascertainment of copy number variation, caused by differing criteria for the length and degree of homology that define a CNV and by the use of different techniques. There is also a problem concerning the reference genome, because of widespread CNV polymorphism and the fact that the reference was established as a haploid genome rather than the natural diploid state.

Mechanisms of structural change

Change in copy number involves change in the structure of the chromosomes such that previously separated chromosomal regions are now juxtaposed. Because the mechanisms of all structural changes are the same as those that cause copy number variation, we discuss them here for the understanding that they provide of the mechanisms of copy number change.

Changes in the structure of chromosomes occur by two general mechanisms: HR and non-homologous recombination. HR requires extensive DNA sequence identity (~50 bp in *E. coli*³⁶ and up to 300 bp in mammalian cells and humans^{37,38}) and most HR mechanisms also require a strand exchange protein, RecA in prokaryotes and its orthologue Rad51 in eukaryotes. The reason for this dual requirement is that an early step in most HR pathways is the RecA/Rad51-catalysed invasion of homologous duplex sequence by the 3' end of ssDNA; that is, the 3' end replaces the equivalent strand of the duplex. By contrast, non-homologous recombination mechanisms use only microhomology or no homology.

HR is the basis of several mechanisms of accurate DNA repair that use another identical sequence to repair damaged sequence. Chromosomal structural change can occur by HR — not because the mechanism is inaccurate, but because many genomes have tracts of LCRs. There will be no change in structure if a damaged sequence is repaired using homologous sequence in the same

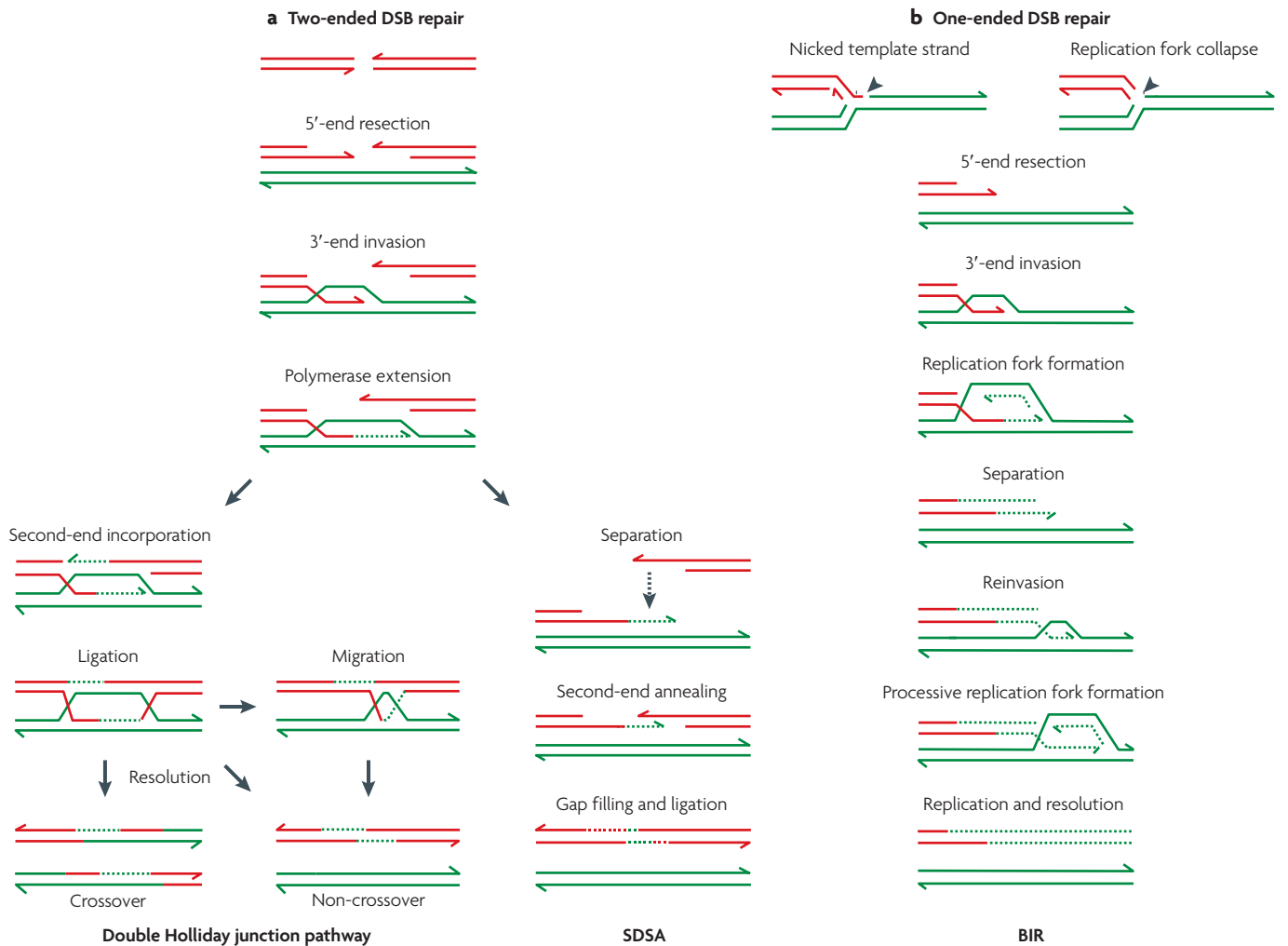


Figure 2 | Mechanisms of homologous recombination. In all parts of the figure, each line shows a single nucleotide chain. Polarity is indicated by half arrows on 3' ends. New synthesis is shown by dotted lines. The broken DNA molecule is shown in red, a homologue or sister molecule is shown in green. Proteins are not shown. **a** | Two-ended double-stranded break (DSB) repair can result in the formation of double Holliday junctions, or in synthesis-dependent strand annealing (SDSA). In the double Holliday junction pathway, the 5' ends of a DSB are resected to leave 3' overhanging tails. These are coated with RecA in prokaryotes or its orthologue Rad51 in eukaryotes, which catalyses invasion by one or both 3' ends into homologous sequence, forming a D loop. The 3' end then primes DNA synthesis (dotted green line), which extends past the position of the original break. In the left resolution pathway, the second end is incorporated into the D-loop by annealing, and is also extended. Following ligation, which forms a double Holliday junction, the junctions are resolved by an endonuclease. The overall effect will be either a non-crossover or a crossover, depending on whether the two junctions are resolved in the same or different orientations, respectively. An alternative resolution pathway is mediated by a helicase and a topoisomerase, which migrate and resolve the double Holliday junction generating only a non-crossover outcome⁴⁸. SDSA begins in the same way as the double Holliday junction pathway, but after the polymerase extension step the invading end, together with the newly synthesized DNA, is separated from the template by a helicase. The invading end now encounters the second end from the DSB, and anneals with it by complementary base pairing (dotted arrow). The second end is extended by DNA synthesis and ligated, thus completing repair. **b** | One-ended DSB repair can be achieved by the break-induced replication (BIR) pathway. The BIR pathway begins at collapsed (broken) replication forks that occur when the replicative helicase at a replication fork encounters a nick in a template strand (solid arrowhead). BIR can be understood as a modification of SDSA. As in SDSA, a 3' tail invades a homologue, usually the sister from which it broke, and is extended after replication fork formation by low processivity polymerization that includes both leading and lagging strands¹⁰⁴. However, the separated extended 3' end fails to find a complementary second end to which to anneal. This 3' end then reinvasades, and is extended further by a low processivity replication fork. This process of invasion, extension and separation might be repeated several times until a more processive replication fork is formed. The fork can now complete replication to the end of the molecule⁵⁰. In the reinvasion and processive replication fork formation steps, we show the Holliday junction following the replication fork, giving conservative segregation of old and new DNA. It is also possible that the Holliday junction is cleaved by an endonuclease, in which case segregation will be semiconservative. Part **b** is modified from REF. 24.

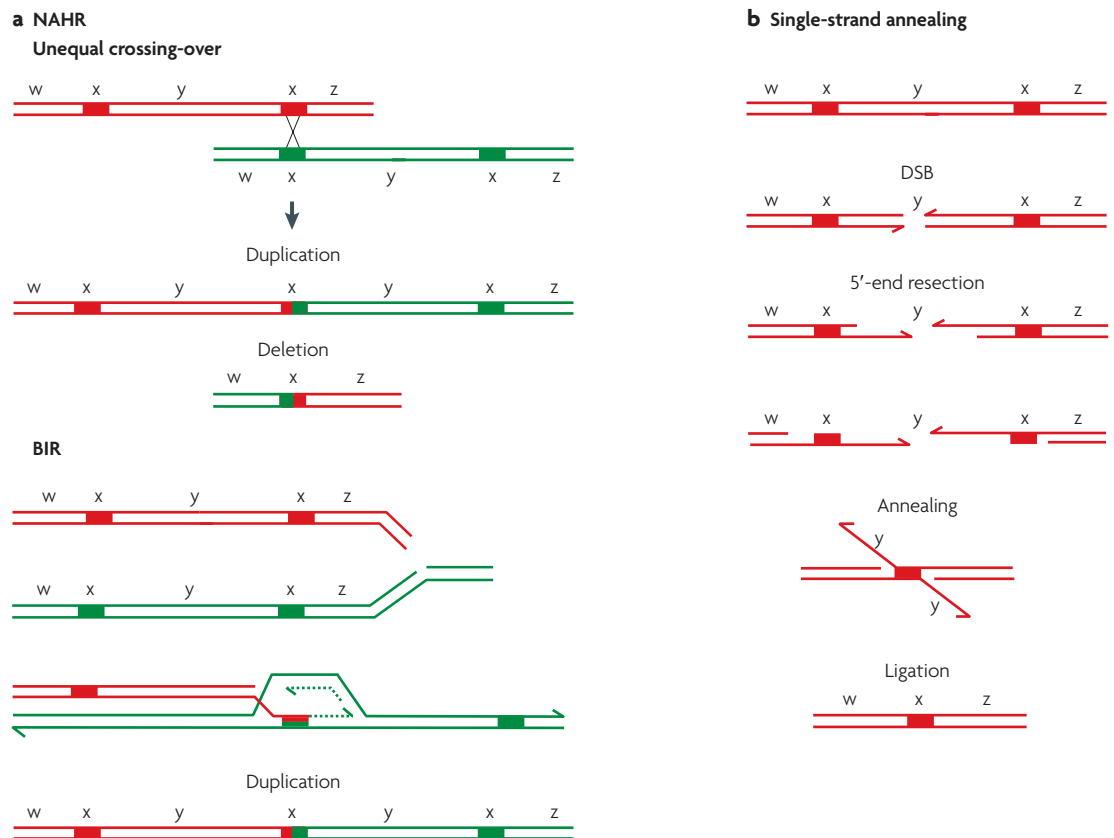


Figure 3 | Change in copy number by homologous recombination. Each line represents a single DNA strand, polarity is indicated by half arrows on 3' ends, and specific sequences are identified by letters w–z. **a** | Non-allelic homologous recombination (NAHR) will occur by unequal crossing over if a recombination repair event uses a direct repeat (x) as homology (upper panel). In this situation, a crossover outcome leads to products that are reciprocally duplicated and deleted for the sequence between the repeats (y). These might segregate from each other at the next cell division, thus changing the copy number in both daughter cells. NAHR can also occur by BIR when the broken molecule uses ectopic homology to restart the replication fork (lower panel). BIR will form duplications and deletions in separate events. **b** | Single-strand annealing. When 5'-end resection on either side of a double-stranded break (DSB) does not lead to invasion of homologous sequence, resection continues. If this resection reveals complementary single-stranded sequence (x) shown by the filled regions, these can anneal. Removal of flaps, gap filling and ligation complete repair of the DSB with deletion of the sequence between the repeats (y) and of one of the repeats.

Loss of heterozygosity
 Loss of an allelic difference between two chromosomes in a diploid cell.

Helicase
 An enzyme that separates the two nucleic acid strands of a double helix, resulting in the formation of regions of ssDNA or ssRNA.

Topoisomerase
 An enzyme that can remove (or create) supercoiling and concatenation (interlocking) in duplex DNA by creating transitory breaks in one (type I topoisomerase) or both (type II topoisomerase) strands of the sugar–phosphate backbone.

generate crossovers. SDSA seems to be a mechanism for avoiding crossing over and loss of heterozygosity (LOH), although it is still capable of producing changes in copy number when the DNA template contains direct repeats (reviewed in REF. 41).

Crossing over between homologous chromosomes can lead to LOH if the chromatids carrying the same alleles segregate together at mitosis. If a crossover forms when the interacting homologies are in non-allelic positions on the same chromosome (NAHR) this will result in duplication and deletion of sequence between the repeats owing to unequal crossing over (FIG. 3a). Crossing over during intrachromosomal recombination between direct or inverted repeats leads to deletion or inversion, respectively. In all organisms tested, including humans, there is a bias in vegetative cells towards the non-crossover outcome (for example, see REFS 42,43). The differences in crossing over frequency can be explained if HR often occurs by the double Holliday

junction model in meiotic cells and by the SDSA model in mitotic cells (FIG. 2a). Several different DNA helicases and topoisomerases can channel DSB repair into a non-crossover pathway either by unwinding the D loop after DNA synthesis^{44–47} (which has been primed by an invading strand) or by resolving a double Holliday junction into a non-crossover event⁴⁸ (FIG. 2a). Another factor that restricts crossing over is repeat length — crossovers are unlikely to form during HR between short repeats, probably owing to the decreased ability to form an intermediate of crossing over, that is, the double Holliday junction⁴⁹.

HR is used not only to repair two-ended DSBs (FIG. 2a), but also to repair collapsed or broken replication forks (FIG. 2b). This process is called break-induced replication (BIR). BIR is normally faithful and leaves no trace, except that it can lead to LOH if the broken end invades a homologue instead of a sister molecule. If the repair process involves homologous sequence in

Single-strand annealing

A double-stranded break repair mechanism that deletes sequence between repeats.

Alu

A family of short interspersed nuclear elements that are common in human and primate genomes.

Mismatch repair

A DNA repair system that corrects a mismatched base pair in duplex DNA by excision of a length of one strand followed by synthesis of the sequence complementary to the remaining strand.

a different chromosomal position, then translocation⁵⁰, duplication or deletion can result — constituting an alternative mechanism for NAHR (FIG. 3a). Several authors have suggested that BIR is a mechanism that results in chromosomal structural change^{31,51–55}. In addition to its HR mechanism, in a later section we discuss its possible involvement in a microhomology-mediated mechanism of copy number change.

Small deletions can occur by a mechanism of break repair that acts at directly repeated sequences. This mechanism, known as single-strand annealing (SSA), was first described in mammalian and amphibian cells^{56,57}. SSA happens when neither of the ends at a two-ended DSB invades homologous sequence. In this case, erosion of the 5' ends (resection) continues, exposing substantial lengths of single-stranded 3' ends (FIG. 3b). If this process exposes complementary sequences in two single strands, annealing can occur. Removal of the flaps followed by ligation completes the repair process, but sequence between the two repeated sequences and one of the repeats have been deleted. Because there is no invasion step SSA does not require RecA/Rad51, but does require the annealing protein Rad52. In yeast, SSA has been found to be limited, in most situations, to

deletions of up to a few tens of kilobase pairs (reviewed in REFS 41,58). In humans, DSB-induced SSA has been observed between identical *Alu* repeats that are separated by few hundred base pairs⁵⁹. The longer the sequence separating the repeats, the less likely it is that resection reaches both repeats and the less likely it is that the break is repaired by SSA. This length restriction means that SSA is likely to be only a minor player in the formation of CNVs.

Correct choice of recombination partner prevents chromosomal structural change. Many of the pathways of chromosomal structural change described here result from a choice of a non-allelic partner for repair. Cells regulate the choice of partner for repair in several different ways. First, mismatch repair provides a barrier to the choice of homeologous sequence (similar sequences that share less than about 97% identity) for repair. In *E. coli* this is presumably because the mismatch repair system — which includes MutS and MutL — can undo base-paired DNA molecules that are imperfectly matched⁶⁰. Mismatch repair also prevents the use of very short lengths of homology as the partner for repair. Second, a sister chromatid is the preferred partner for recombinational repair. The proteins that hold two sister chromatids together are called cohesins. Cohesins are assembled at DSBs in both yeast and humans^{61–63} and facilitate DSB repair⁶⁴; they restrict the opportunity to utilize either intrachromosomal or interchromosomal NAHR templates. In yeast, cohesins regulate the copy number of tandem ribosomal RNA gene repeats (rDNA)⁶⁵, which are susceptible to deletions and insertions. Transcription of rDNA has been suggested to cause local disruption of cohesin binding, thus leading to the choice of nonallelic repeats for repair in rDNA and consequent copy number change⁶⁶. However, cells respond to such changes by regulating recombination to bring the number of the repeats back to the normal level. It seems likely that loss of cohesion between chromatids might cause copy number change at other loci.

In addition to holding sister chromatids together following DNA damage, yeast and human cells also keep the two ends of a single DSB together^{67,68}. In yeast, Sgs1 (an orthologue of human BLM helicase) is one of the proteins that coordinate the choice of template for the repair of the two ends of a DSB^{69,70}. However, multiple reports show that the two ends of a single DSB can engage in recombination with different homologous templates — this copying of different sequences from different templates by the two ends of a single DSB will lead to rearrangements.

Although HR provides vital repair mechanisms, it is also hazardous, as revealed by the many ways in which it can lead to chromosomal structural change, including copy number variation. HR repair mechanisms minimize this by avoiding crossing over, by regulating partner choice and by requiring substantial lengths of perfect homology. However, meiosis requires crossing over, and we see a possible effect of this requirement in the elevated frequency of CNVs arising in meiosis (BOX 2).

Box 2 | When and how frequently do changes in copy number occur?

Much copy number variation in humans occurs as inherited polymorphism, but variation also arises *de novo* at a significant rate, both in the germ line and in somatic cells. A study of four hot spots at which copy number variation occurs by non-allelic homologous recombination (NAHR)¹⁴⁸ found a frequency of copy number change of between 10^{-6} and 5×10^{-5} per gamete, as determined from sperm cell analysis. A study using similar methods to analyse blood and sperm from two individuals for NAHR-mediated deletions at the α -globin locus reported a frequency of over 10^{-6} in blood cells, and over 10^{-5} in sperm cells¹⁴⁹. Similar results were obtained for duplications at the α -globin locus¹⁵⁰. A study that analysed three specific chromosomal inversions arising by NAHR between inverted low copy repeats (LCRs) in blood found very high frequencies ranging from 10^{-4} to 10^{-1} . Newborns carried much less NAHR product than older individuals, suggesting that somatic structural changes accumulate during life¹⁵¹. There are few comparable data for copy number variation caused by non-recurrent changes. Extensive somatically generated copy number variation between different embryonic stem cell lines derived from the same inbred laboratory strains of mice has also been reported¹⁵²; most, but not all, of the variants were associated with LCRs, suggesting NAHR. Thus, the rate of CNV formation is several orders of magnitude higher than that of point mutations, and is especially high during meiosis.

Data on the occurrence of sporadic genomic disorders have recently been reviewed¹⁵³, and frequencies have been reported in the range of 10^{-6} to 10^{-4} copy number changes per gamete, including non-recurrent rearrangements at the dystrophin locus¹⁵⁴. This frequency is two to four orders of magnitude more than for point mutation¹⁵³, and is in good agreement with the estimate from sperm cells¹⁴⁸ (discussed above). There is, however, a notable difference in that deletions were approximately twice as common as duplications in sperm cells¹⁴⁸, whereas in healthy individuals deletions and duplications are approximately equally common⁴. This suggests that there is less selection against duplications than against deletions, because sperm cells are not subject to the selective pressures during development.

It is interesting to speculate that a high rate of generation of copy number variation had a role in the rapid evolution of the primate lineage¹⁷. Rats and mice have fewer LCRs than humans¹⁵⁵, and laboratory strains of mice show copy number variation that co-localizes with LCRs, suggesting a NAHR¹⁵⁶ mechanism. Therefore, the low occurrence of LCRs in mice makes it likely that *de novo* CNV formation will be substantially less frequent than that in humans. By contrast, copy number variation and LCRs in chimpanzees seem to be evolving at a rate that is comparable to that in humans^{10–13}.

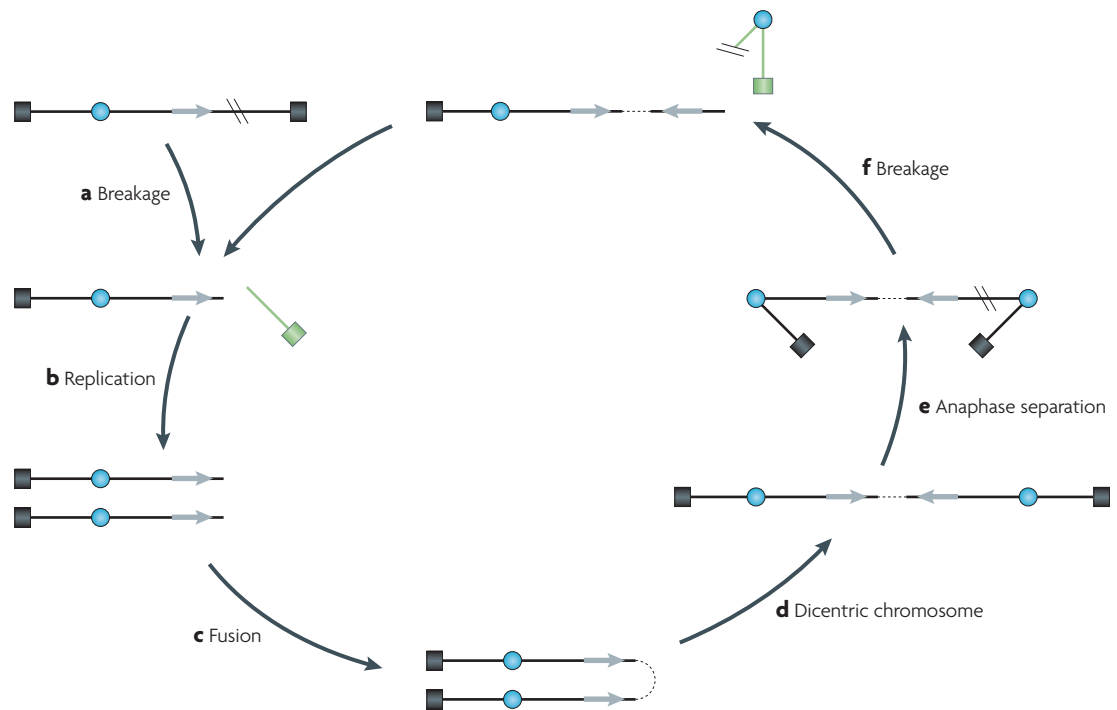


Figure 4 | The breakage–fusion–bridge cycle. Centromeres are indicated by a blue circle, telomeres by a black block and genomic sequence as grey arrows showing orientation. Breakage points are shown as double black lines, and fragments that are lost are in green. **a** | A doubled-stranded break (DSB) occurs in an unreplicated chromosome, causing it to lose a telomere. **b** | After replication, both sister chromatids lack telomeres. **c** | These two ends are thought to fuse. **d** | The fusion in part **c** forms a dicentric chromosome. **e** | At anaphase, the two centromeres of the dicentric chromosome are pulled apart, initially forming a bridge between the telophase nuclei. **f** | Eventually, the bridge is broken in a random position. This inevitably leads to the formation of a large inverted duplication. The chromosome once again has an unprotected end, and after replication will form two sister chromatids that can fuse to form a new a dicentric chromosome, and so the process is repeated until the end acquires a telomere from another source. Amplification of the large inverted duplication can occur by random breakage in later cycles (not shown).

Retrotransposon

A transposon (mobile element) that is copied from the host genome by transcription as RNA, and is later reverse-transcribed into DNA and reintegrated into the host genome.

Endonuclease

An enzyme that breaks the sugar–phosphate backbone of a DNA or RNA molecule where there is no free end.

Telomere

A structure at the ends of linear chromosomes that avoids shortening of chromosomes after replication, and that protects the end from homologous and non-homologous recombination.

Dicentric chromosome

A chromosome with two centromeres. These are pulled to opposite poles during mitosis but are unable to separate without chromosome breakage.

Non-replicative non-homologous repair

In addition to HR pathways, there are mechanisms of DNA repair that use very limited or no homology. When homology is not used to ensure that molecules are rejoined in the correct positions, there is some probability that genetic change such as copy number variation will result. These mechanisms that do not use HR can be divided into non-replicative and replicative mechanisms, discussed in this and the next section respectively.

Non-homologous end joining. There are two pathways of DSB repair that either do not require homology or need very short microhomologies for repair: non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ). These pathways have recently been described in detail elsewhere^{71–73}. NHEJ either rejoins DSB ends accurately or leads to small (1–4 bp) deletions, and in some cases to insertion of free DNA, often from mitochondria or retrotransposons^{74,75}. In MMEJ, 5–25 bp homologies anneal at the ends of DSBs and, like SSA, MMEJ leads to deletions of sequences between annealed microhomologies. A second distinction between these pathways is that they require different proteins: for example, the DNA end-binding proteins Ku70 and Ku80 are required for NHEJ but not for MMEJ. Also,

the strand-annealing protein Rad52 is not required for MMEJ but is required for SSA, thereby distinguishing these two pathways.

It is likely that NHEJ and MMEJ contribute to some chromosomal rearrangements by joining non-homologous sequences. This might occur during repair of two-ended DSBs (such as endonuclease-induced breaks), through damage by exogenous agents (including chemotherapeutic agents), or when two converging replication forks encounter a nick in the DNA. Programmed two-ended DSBs occur in the immune system, and their repair might relate to the formation of some translocations seen in patients with leukaemia (reviewed in REF. 76). Programmed two-ended DSBs also occur in cells undergoing meiosis — in this case they initiate HR. As discussed below, single-ended DSBs are likely to be a more frequent spontaneous lesion and are likely to undergo replicative repair.

Breakage–fusion–bridge cycle. After replication of a chromosome that has lost its telomere because of a DSB, there will be two sister chromatids that lack telomeres. McClintock⁷⁷ proposed that sister chromatids that lack telomeres will fuse, creating a dicentric chromosome (FIG. 4). During anaphase, the two centromeres will be

pulled to separate nuclei, causing eventual breakage of the dicentric chromosome. The break will lead, after replication, to new ends that lack telomeres — these new ends will fuse and form a new dicentric chromosome, and a cycle is established. Random breakage causes large inverted duplications, and repeated cycles could lead to amplification of the inverted repeat. The cycle will cease when the chromosome acquires a telomere. This process, the breakage–fusion–bridge cycle, has been linked to the formation of amplification in mammalian cells (reviewed in REF. 78), and it is believed to play a major part in amplification in cancer. The random breakage of the dicentric chromosome formed by fusion of the ends of sister chromatids provides an explanation for the occurrence of large inverted repeats in human cancer cells⁷⁹. The breakage–fusion–bridge cycle can be induced by enzymatic breakage of chromosomes and by inhibition of DNA synthesis⁸⁰, and the bridges that are formed can be observed microscopically.

Some events that change chromosome structure that have been attributed to the breakage–fusion–bridge cycle could also be caused by any repeated non-homologous recombination process without repeated breakage and fusion⁵⁵. Clearly, when a translocation forms in inverted orientation, thereby creating a dicentric chromosome, a second event will be required to restore stability to the genome. This second event might be part of the first translocation event, as described in replicative models below, rather than being a result of anaphase bridge formation.

Replicative non-homologous repair

The presence of microhomology at a site of non-homologous recombination has been regarded as the signature of NHEJ^{34,81}. However, evidence that the formation of microhomology junctions is in some cases linked to DNA replication has accumulated, and replicative mechanisms, particularly BIR, are increasingly thought to underlie the formation of structural changes with microhomology junctions^{24,31,51,52,54,55}. The evidence for the involvement of replication in at least some chromosomal structural change has recently been reviewed²⁴.

There is also growing evidence that replicative stress might underlie copy number change. Aphidicolin, an inhibitor of replicative DNA polymerases, induces copy number variation at chromosomal fragile sites and throughout the genome^{80,82–85}. Double-stranded ends are known to result from replication inhibition⁸⁶. In one study, aphidicolin-induced CNVs were found to have microhomology (65% of CNVs) or no homology at their end-points, showing that they did not arise by HR⁸². Therefore, although other mechanisms could also be involved in repairing DSBs, these studies suggest that non-homologous replicative mechanisms might be important in copy number change. In the following sections we explore the replicative mechanisms that have been proposed as the origin of copy number variation.

Replication slippage or template switching. When short lengths of DNA sequence identity occur in the part of the genome that is expected to be single-stranded during

replication — that is, the length of an Okazaki fragment (1 or 2 kb in *E. coli*, shorter in humans) — the sequence between the homologous regions is often deleted or duplicated. This deletion or duplication has been attributed to a mechanism of replication slippage along the exposed template during DNA replication^{87,88} (FIG. 5A). In *E. coli*, replication slippage can occur in the absence of RecA^{87–90}, and is strongly dependent on the length of homology^{87,91,92} and the distance between the repeat units^{91,93,94}. The frequency of these events is increased by mutations in genes encoding components of the replicative DNA polymerase holoenzyme, presumably because perturbation of replication promotes the slippage^{95–97}. When the repeat sequences are not identical (that is, they contain mismatches), the frequency of replication slippage is higher in mutants carrying a mutation in the mismatch repair system⁹⁸. These findings, together with the failure to find genetic requirements for short homology deletion⁹⁷ (suggesting that essential functions are involved, so that most mutations in these genes would render a cell inviable) and the absence of any requirement for HR functions⁹⁷, suggest a replicative mechanism of structural change. Because of the strong distance limitation, the replication slippage mechanism is proposed to operate within a replication fork and so cannot account for most of the events that change copy number in humans, in which distances of tens of kilobases to megabases are involved.

Fork stalling and template switching. Study of stress-induced amplification of the *lac* genes, using the *E. coli* Lac system of Cairns and Foster⁹⁹, led Slack *et al.*¹⁰⁰ to propose that template switching is not confined to single replication forks, but can also occur between different replication forks. This model, now called fork stalling and template switching (FoSTeS)²⁶ (FIG. 5B), proposes that when replication forks stall, the 3' primer end of a DNA strand can change templates to a ssDNA template in a nearby replication fork. This hypothesis was generated because the mean length of amplified units (amplicons) in that study was ~20 kb¹⁰⁰, which is too long to occur within a replication fork.

There are several lines of evidence that FoSTeS is replicative. First, the junctions between amplicons showed only microhomology (4–15 bp)^{100,101}, indicating that HR is not involved. Second, there was a requirement for DNA polymerase I, specifically for its 5' flap endonuclease domain. This suggests that lagging strands at replication forks are involved in template switching because the excision repair functions of DNA polymerase I were not involved¹⁰⁰. Third, overproduction of the main 3' ssDNA exonuclease, ExoI, decreased the frequency of rearrangements — this implies that 3' DNA ends promote the amplification of the *lac* genes mentioned above and suggests that DNA synthesis is primed from the 3' ends during amplification¹⁰⁰. The reciprocal result was seen when the gene for ExoI was deleted in short-range deletion events^{102,103}. The physical properties of the amplicons, microhomology at the boundaries and complexity in the structure of amplicons in *E. coli*^{100,101} have also been found to be properties of human duplications

Amplification

Also called gene amplification. The formation of more than two repetitions of a chromosomal segment in tandem, dispersed or as autonomous circular molecules.

Fragile site

A position on a chromosome where spontaneous breaks occur frequently.

Okazaki fragment

The discontinuous length of DNA that is synthesized as one piece on the lagging strand template during DNA replication.

Amplicon

The repeat unit, or unit length of genome, that is amplified.

Exonuclease

An enzyme that degrades DNA or RNA from an end.

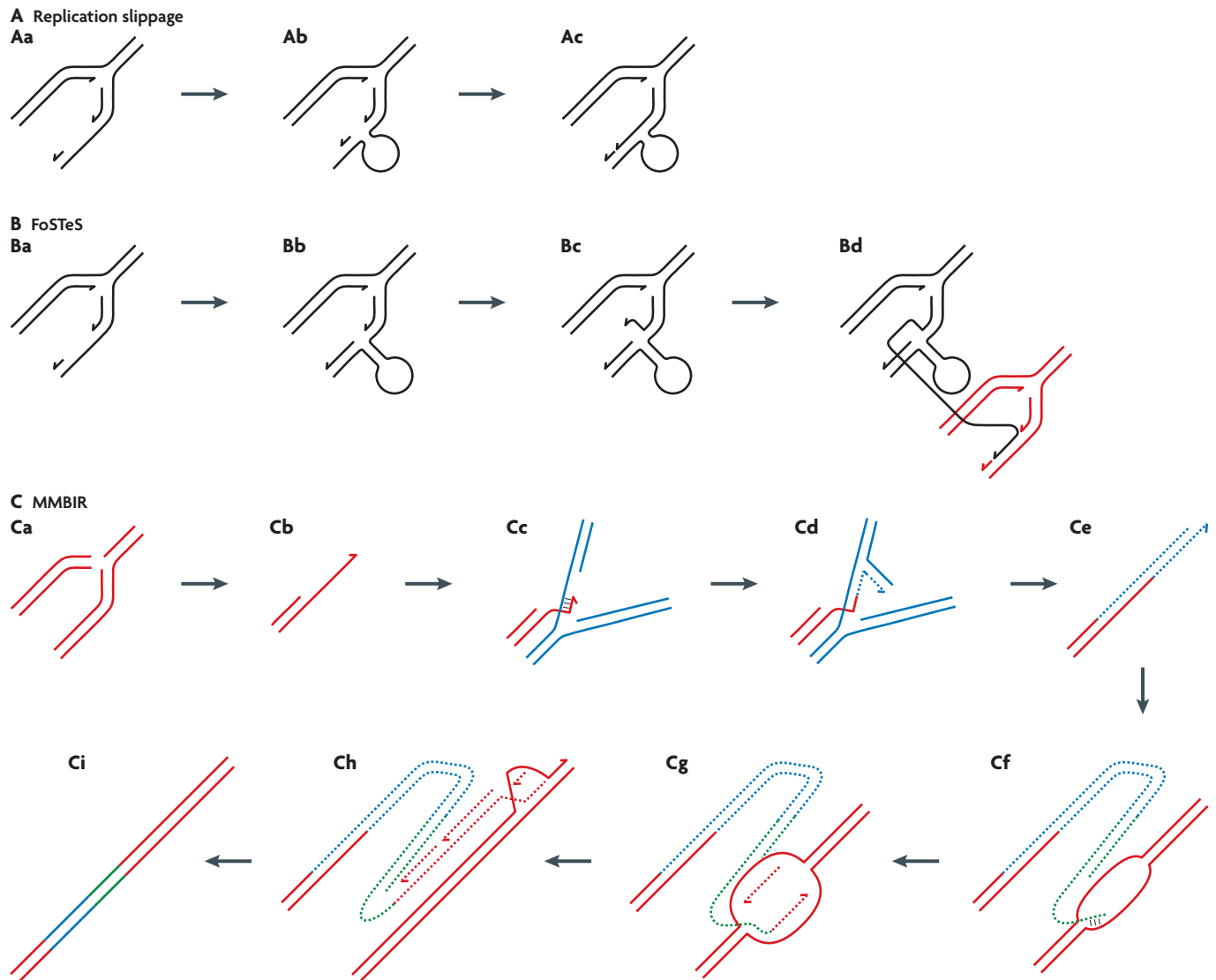


Figure 5 | Replicative mechanisms for non-homologous structural change. Each line represents a single DNA strand, polarity is indicated by half arrows on 3' ends, and arrowheads show the position of nicks and breaks. Microhomology junctions are indicated by short black lines between single DNA strands. **A** | Replication slippage. During replication, a length of lagging-strand template becomes exposed as a single strand (**Aa**). The 3' primer end can move to another sequence showing a short length of homology on the exposed template (**Ab**); this move might occur owing to the formation of secondary structures in the lagging-strand template. Lagging strand synthesis can continue after having failed to copy part of the template (**Ac**). As shown, this will produce a deletion. Several variations on this mechanism can also produce a duplication of a length of DNA sequence with or without sister chromatid exchange (reviewed in REF. 157). Events occurring by this mechanism are confined to the length of genome found in a single replication fork. **B** | Fork stalling and template switching (FoSTeS)^{26,100}. An exposed single-stranded lagging strand template (**Ba**) might acquire a secondary structure (**Bb**), which can block the progress of the replication fork. The 3' end then becomes free from its template (**Bc**), and might then alight on another exposed single-stranded template sequence on another replication fork that shares microhomology (**Bd**), thus causing duplication, deletion, inversion or translocation, depending on the relative position of the other replication fork. Fork stalling can be caused by other situations, such as lesions in the template strand or shortage of deoxynucleotide triphosphates. **C** | Microhomology-mediated break-induced replication (MMBIR). Replication fork collapse (**Ca**) in which

one arm breaks off a replication fork can occur because the fork encounters a nick on a template strand, or can be caused by endonuclease. The 5' end of the broken molecule (**Cb**) will be recessed from the break, exposing a 3' tail. When insufficient RecA or Rad51 is present to allow invasion of homologous duplex as shown in FIG. 2, the 3' tail will anneal to any exposed ssDNA that shares microhomology. The 3' tail can anneal to the lagging strand template of another replication fork (blue) (**Cc**). A replication fork can then be established with both leading and lagging strand synthesis from the microhomology junction (**Cd**). The replication is of low processivity, and the broken end, which was extended by a length of a different sequence (blue), is separated from the template and again processed to a 3' tail, which will then anneal to another single-stranded microhomology sequence (**Ce**). The extended broken end, now carrying both the sequence identified in blue and a length of different sequence identified in green (formed by the same process), anneals with single-stranded sequence back onto the red molecule (**Cf**). In this case, the single-stranded sequence is shown as a locally melted length of DNA. Another short-processivity fork is established (**Cg**), but this one becomes a fully processive replication fork (**Ch**) that can continue to the end of the chromosome or replicon. The molecule that is produced carries short sequences from other genomic locations (**Ci**). Whether a length of red sequence is duplicated or deleted depends on the position at which synthesis returns to the red chromosome relative to where the initial fork collapse occurred. If the last sequence (red) is a homologous chromosome instead of the sister chromatid, there will be extensive loss of heterozygosity downstream from the event.

and deletions, as discussed above^{18,26,28}. This led to the proposal^{26,100} that FoSTeS is involved in the formation of some human chromosomal rearrangements and CNVs.

Microhomology-mediated BIR. Other authors have proposed that BIR can be mediated by microhomology. Notably, Payen *et al.*⁵⁴ demonstrated the involvement of BIR in microhomology-mediated non-homologous recombination by showing a requirement for Pol32 — a non-essential DNA polymerase previously shown to be required for BIR in yeast¹⁰⁴. Bauters *et al.*⁵¹ invoked a mechanism of BIR mediated by microhomology to explain non-recurrent copy number changes in humans. Both these teams proposed that there was a microhomology-mediated invasion of dsDNA. We, however, propose that invasion will not occur without extensive homology, and that a mechanism other than invasion is involved when only microhomology is present.

As the FoSTeS model does not propose mechanistic molecular detail, does not involve DNA double-stranded ends and is not readily testable, we suggest that it should be superseded by a new model — microhomology-mediated break-induced replication (MMBIR). MMBIR is based on the mechanism of BIR repair of single double-stranded ends²⁴ (FIG. 5C) (see REF. 24 for a more extensive discussion of the evidence). We propose that when a single double-stranded end results from replication fork collapse in a cell under stress, classical BIR repair of the double-stranded end cannot occur because RecA/Rad51 is downregulated as part of the stress response; BIR is strongly RecA/Rad51-dependent because it includes an invasion by a 3' DNA end into dsDNA of the repair partner. However, BIR is known to occur at a low rate in the absence of Rad51 (REFS 105,106). MMBIR postulates that, because strand invasion is limited or not possible when RecA/Rad51 is downregulated, the 3' end from the collapsed fork will anneal to any single-stranded template with which it shares microhomology and that is present in physical proximity to the 3' DNA end. This single-stranded template could be ssDNA that occurs in the lagging strand template of other replication forks, or ssDNA at excision repair tracts, at sites of transcription and at secondary structures in DNA. This annealing initiates DNA synthesis and a low-processivity replication fork. The annealing reaction does not require RecA/Rad51 and requires very little homology, so annealing will occur with the sister molecule either in front of or behind the position of replication fork collapse, leading to deletion or duplication, respectively, and in either orientation, giving the opportunity for inversion. Microhomology might also be found in a different chromosome, leading to translocation. Annealing with the homologous chromosome instead of the sister chromosome could be a cause of extensive LOH. The repeated extension of and separation from the template strand that are characteristic of BIR⁵⁰ might cause several of these changes to occur in the same repair event, which would lead to the complex junctions that have been observed. The ability of MMBIR to explain the complexity of multiple junctions in close proximity (FIG. 1) is an attractive feature of this model.

Supporting the idea that chromosomal structural change can result from an insufficiency of RecA/Rad51 is the observation that deletion of one copy of the *RAD51* homologue in *D. melanogaster* gives a mixture of homologous and non-homologous junctions from DSB repair¹⁰⁷. There are two lines of evidence to support the idea that RecA/Rad51 is downregulated when cells are under stress and that this leads to MMBIR. First, hypoxic stress in human cancer cell lines leads to repression of *RAD51* and to reduced HR (reviewed in REFS 108,109). This has been interpreted as a stress-induced switch from high-fidelity HR to lower-fidelity NHEJ¹¹⁰. However, in the case of a collapsed replication fork, NHEJ is not possible because there is only one end. We suggest that such a switch in DSB repair could therefore lead to a BIR-based mechanism, such as MMBIR. Hypoxia is known to induce gene amplification in cancer cell lines by activating fragile sites that lead to DSBs¹¹¹, presumably producing single double-stranded ends (fragile sites are also activated by DNA synthesis inhibition⁸⁵). The second line of evidence is that, in *E. coli*, amplification that involves the formation of microhomology junctions^{100,101} (discussed above) is induced by starvation stress; amplification does not begin to appear until the cells are starved¹¹² and it requires induction of the general and starvation stress response by the RpoS transcriptional activator¹¹³. However, it has not been shown that this stress response leads to downregulation of the *recA* gene. Another switch from high-fidelity to error-prone DSB repair that is seen in starved *E. coli* depends on the expression of the cells' major general stress response¹¹⁴ genes; even artificially inducing the stress response in the absence of stress causes this switch¹¹⁴.

Because either homologue can be copied during BIR, the MMBIR model predicts that structural change will often be accompanied by extensive LOH and, in some cases, by loss of imprinting²⁴. Many instances of deletions associated with extensive LOH have been reported in cells from patients with acute lymphoblastic leukaemia¹¹⁵. We expect that NAHR will often be associated with LOH, because crossovers readily lead to LOH. However, we know of no reason to expect end-joining mechanisms to show this association, and finding microhomology junctions associated with LOH provides strong support for an alternative mechanism.

As described above, end-joining and MMBIR mechanisms could lead to microhomology junctions and insertion of other sequences at the junction, but there is currently no way to determine which mechanism led to a particular microhomology junction after it has formed. However, the presence of complexity is more characteristic of MMBIR than of NHEJ, and although MMEJ leads to microhomology junctions, they are associated with deletions but not with insertions. We think that mechanisms such as MMBIR are more likely than end-joining mechanisms to be responsible for the generation of most non-recurrent copy number variation. We suggest this because most insertions at end-points are insertions of nearby sequence; there is extensive and increasing evidence that replication has a role in the generation of non-recurrent copy number variation; it

is expected that single-ended DSBs will be much more common than two-ended DSBs; and explaining duplications, triplications and complex rearrangements by end-joining mechanisms would involve an intricate series of events, but they are easily explained by a BIR-based mechanism.

Effects of chromosome architecture

CNVs are not randomly distributed in the human genome, but tend to be clustered in regions of complex genomic architecture, which consist of complex patterns of direct and inverted LCRs. Some clustering might result from the absence of dosage-sensitive genes in particular regions, but there is ample evidence that specific features of chromosomal architecture are also involved.

The most obvious effect of architecture is that changes mediated by NAHR occur where there are pre-existing LCRs that provide the homology needed for recombination (discussed above). More subtle influences include: the preferential occurrence of copy number variation in regions of heterochromatin near telomeres^{116,117} and centromeres^{118–120}; the association of copy number variation with sequence-specific structures such as replication origins, replication terminators⁵⁴ and scaffold attachment sequences^{27,121}; the occurrence of non-recurrent changes in regions carrying multiple LCRs^{26,33}, including inverted repeats and palindromic sequences (reviewed in REF. 78); and the role of repetitive sequences and of long and short interspersed nuclear elements (LINEs and SINEs) in generating structural change¹²². The ability of DNA sequences to adopt a non-B conformation, such as a cruciform shape, affects chromosomal structural change in a way that depends on the structures rather than the specific sequences that generate them^{123–125}. Finally, there are reports of specific consensus sequences associated with copy number variation^{30,100,121,126}.

The preferential occurrence of non-recurrent structural changes close to LCRs has been recognized for some time^{33,34,127}. This has been explained as a tendency for secondary structures in DNA, which can form in LCRs, to cause replication fork stalling²⁶ and also to provide single-stranded regions that can facilitate the formation of microhomology junctions²⁴. A large proportion of the human genome consists of SINE retrotransposons, predominantly *Alu* sequences, and LINE retrotransposons. In addition to causing mutation by insertion into coding sequences, these elements cause copy number variation by NAHR and provide a focus for non-recurrent changes in copy number¹²⁸ in a way that is not understood — one study found *Alu* elements at 13 of 40 microhomology deletion end-points¹²². The association of LINEs and SINEs with CNVs could be caused either by DNA breakage being frequent in these areas (because of an active transposase, for example), which could initiate non-homologous recombination, or by persistent single-strandedness in these regions (owing to extensive transcription, secondary structures or replication pausing), which could make them preferred sites for annealing by ssDNA ends (the step in repair mechanisms that is proposed to lead to microhomology at junctions).

Thus, we conclude that copy number changes are not randomly distributed, but that multiple genomic features can affect the probability of their occurrence. Detailed mechanistic explanations for the impact of these architectural characteristics on CNV formation await further work.

Conclusions and ramifications

There are at least two main mechanisms for change in copy number: NAHR and microhomology-mediated events. NAHR can occur either by HR-mediated DSB repair through a double Holliday junction, or from BIR, which restarts broken replication forks by HR. However, the LCRs that mediate NAHR were presumably formed in the past by the same mechanisms that are forming non-recurrent copy number changes now. Thus, microhomology-mediated mechanisms seem to underlie most copy number change.

Based on the evidence favouring replicative mechanisms, on the enzymes known to be involved in DNA transactions in model organisms, and on the evidence presented above concerning the potential involvement of stress responses in altering the availability of DNA repair proteins, we suggest that a mechanism such as MMBIR is currently our best working hypothesis for many copy number change events. It is also likely that end-joining mechanisms, including NHEJ, MMEJ and SSA, play a part, especially in cells of the immune system. The breakage–fusion–bridge cycle has been shown to operate in experimental systems and seems to be important in amplification in some cancers. However, we need not assume that only one mechanism acts in any given event. Microhomology-mediated events might trigger the breakage–fusion–bridge cycle by forming a dicentric chromosome, which must eventually be resolved to a more stable genotype. This could also happen when NAHR causes formation of a dicentric chromosome. Similarly, end-joining mechanisms might have a role in repairing free ends that result from other events. Notably, the fusion step of the breakage–fusion–bridge cycle might be mediated by any of these end-joining mechanisms.

The molecular events proposed in MMBIR have not been shown experimentally. However, because of its molecular detail, several aspects of the model are testable. Testing the suggested involvement of stress responses will be important. The potential for extensive LOH downstream from the initiating event has already been seen in some systems, and further testing of this correlation should be available by studying genome-wide SNP data. This LOH might extend as far as the next replication fork travelling in the opposite direction, or it might process to the telomere.

If it can be substantiated that copy number variation stems from stress response, this has interesting implications for physiology, evolution and disease. First, stress-inducible chromosomal structural variation suggests that cells have an inducible ability to evolve (this property is termed *evolubility*). If the mechanism can be induced by stress, then cells and organisms will be predisposed to genome rearrangement when they

Heterochromatin

A highly condensed form of chromatin (the eukaryotic complex of DNA with proteins) that shows reduced gene expression and is replicated late in S phase.

LINE

Long interspersed nuclear elements. A class of transposable element lacking long terminal repeats.

SINE

Short interspersed nuclear elements. A class of short (<500 bp) transposable elements.

Evolubility

The capacity of an organism to evolve.

activate stress responses; this will occur particularly when they are maladapted to their environments. Thereby, if stress fuels CNV formation, generation of the genetic diversity upon which natural selection acts will be increased when a population will benefit most from such diversity, potentially fuelling evolution at that time. The idea of mutation as a stress response has been developed to explain how mutagenesis affects the evolution of bacterial populations (reviewed in REFS 129,130), including the evolution of antibiotic resistance following antibiotic exposure^{131,132}. If CNV formation can be induced by stress, then CNVs may not only be important promoters of evolutionary divergence¹² but, in addition, their formation may be a mechanism to enhance evolvability. Although human cells show stress-inducible genetic change^{109,133,134}, it is not yet known in any organism with a differentiated germ line whether such stress-inducible genome instability mechanisms can contribute to genetic variation in the germ line and therefore to evolution. This is an important topic for future study.

Second, the prediction that LOH would be a result of CNV formation is likely to be important in human cancer, in which mutations and LOH drive tumour progression and resistance to therapies. The same molecular repair mechanisms that cause copy number variation could produce translocations when the repair mechanism uses microhomology in a different chromosome²⁴, leading to further structural changes that could drive cancer progression. The problem of stress-inducible cancer progression and resistance mechanisms are discussed elsewhere^{108,129}.

Finally, given the large and ever increasing number of developmental, neurological and psychiatric syndromes that have been linked to copy number variation, we wish to propose an analogy, and a corollary from cancer biology, for these other CNV-based diseases. It is well known that mutator mutations, which increase the mutation rate, promote cancer predisposition because cancer is a genetic disease fuelled by mutagenesis and genome instability^{135–137}. Mutator mutations are common in microorganisms and many are known in a wide variety of cancer-predisposing and ageing syndromes²⁵. We propose that, in addition to a direct involvement of CNVs in various syndromes, there will be human CNVs with increased rates of copy number change owing to mutations in any one of many genes that affect DNA damage processing and repair (among other processes). We expect that such alleles will predispose families and individuals to the mental, developmental, neurological and other syndromes caused by CNVs. Future work should seek to identify such modifier locus mutations that affect human developmental and other disorders — we predict that some of these mutations will do so by elevating CNV formation rates. Individuals with such mutations might also be cancer prone. In theory, in cases of disease that are caused by a CNV that is present in only a subset of cells (mosaicism), therapies to reduce CNV genesis might reduce the penetrance of such a disease; as yet, drugs that do this are neither known nor even being considered. Identification of the many genes, proteins and pathways that affect CNV formation rates and understanding their mechanisms of action are prerequisites to considering such potential therapeutic approaches.

1. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
2. Kidd, J. M. *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56–64 (2008).
3. Korb, J. O. *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426 (2007).
4. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
5. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
6. Wong, K. K. *et al.* A comprehensive analysis of common copy-number variations in the human genome. *Am. J. Hum. Genet.* **80**, 91–104 (2007).
7. Bruder, C. E. G. *et al.* Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am. J. Hum. Genet.* **82**, 1–9 (2008).
8. Piotrowski, A. *et al.* Somatic mosaicism for copy number variation in differentiated human tissues. *Hum. Mutat.* **29**, 1118–1124 (2008).
9. Beckmann, J. S., Estivill, X. & Antonarakis, S. E. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nature Rev. Genet.* **8**, 639–646 (2007).
10. Dumas, L. *et al.* Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res.* **17**, 1266–1277 (2007). **This paper traces the history of copy number variation through the evolution of the primate lineage.**
11. Nahon, J. L. Birth of 'human-specific' genes during primate evolution. *Genetica* **118**, 193–208 (2003).
12. Bailey, J. A. & Eichler, E. E. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nature Rev. Genet.* **7**, 552–564 (2006).
13. Stankiewicz, P., Shaw, C. J., Withers, M., Inoue, K. & Lupski, J. R. Serial segmental duplications during primate evolution result in complex human genome architecture. *Genome Res.* **14**, 2209–2220 (2004).
14. Marques-Bonet, T. *et al.* A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457**, 877–881 (2009).
15. Inoue, K. & Lupski, J. R. Molecular mechanisms for genomic disorders. *Annu. Rev. Genomics Hum. Genet.* **3**, 199–242 (2002).
16. Ohno, S. *Evolution by Gene Duplication* (Springer, Berlin, New York, 1970).
17. Rotger, M. *et al.* Partial deletion of CYP2B6 owing to unequal crossover with CYP2B7. *Pharmacogenet. Genomics* **17**, 885–890 (2007).
18. Zhang, F. *et al.* The DNA replication FoSTeS/MMBIR mechanism can generate human genomic, genic, and exon shuffling rearrangements. *Nature Genet.* **41**, 849–853 (2009).
19. Volik, S. *et al.* Decoding the fine-scale structure of a breast cancer genome and transcriptome. *Genome Res.* **16**, 394–404 (2006).
20. Brodeur, G. M. & Hogarty, M. D. In *The Genetic Basis of Human Cancer* (eds Vogelstein, B. & Kinzler, K. W.) 161–172 (McGraw-Hill, New York, 1998).
21. Frank, B. *et al.* Copy number variant in the candidate tumor suppressor gene *MTUS1* and familial breast cancer risk. *Carcinogenesis* **28**, 1442–1445 (2007).
22. Lupski, J. R. Genomic disorders: structural features of the genome can lead to DNA rearrangement and human disease traits. *Trends Genet.* **14**, 417–422 (1998).
23. Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
24. Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* **5**, e1000327 (2009). **This review presents the MMBIR model for chromosomal rearrangement with detail of the evidence on which it is based.**
25. Friedberg, E. C. *et al.* *DNA Repair and Mutagenesis* (ASM, Washington DC, 2005).
26. Lee, J. A., Carvalho, C. M. & Lupski, J. R. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**, 1235–1247 (2007). **Description of the complex structure and microhomology of non-recurrent duplications seen in patients with a genomic disorder.**
27. Nobile, C. *et al.* Analysis of 12 deletion breakpoints in dystrophin intron 49. *Hum. Genet.* **110**, 418–421 (2002).
28. Carvalho, C. M. *et al.* Complex rearrangements in patients with duplications of *MECP2* can occur by Fork Stalling and Template Switching. *Hum. Mol. Genet.* **18**, 2188–2203 (2009).
29. Chen, J. M., Chuzhanova, N., Stenson, P. D., Fèrec, C. & Cooper, D. N. Intrachromosomal serial replication slippage in *trans* gives rise to diverse genomic rearrangements involving inversions. *Hum. Mutat.* **26**, 362–373 (2005).
30. Gajicka, M. *et al.* Unexpected complexity at breakpoint junctions in phenotypically normal individuals and mechanisms involved in generating balanced translocations t(1;22)(p36;q13). *Genome Res.* **18**, 1733–1742 (2008).
31. Sheen, C. R. *et al.* Double complex mutations involving *F8* and *FUNDC2* caused by distinct break-induced replication. *Hum. Mutat.* **28**, 1198–2006 (2007).

32. Vissers, L. E. *et al.* Complex chromosome 17p rearrangements associated with low-copy repeats in two patients with congenital anomalies. *Hum. Genet.* **121**, 697–709 (2007).
33. Stankiewicz, P. *et al.* Genome architecture catalyzes nonrecurrent chromosomal rearrangements. *Am. J. Hum. Genet.* **72**, 1101–1116 (2003).
34. Lee, J. A. *et al.* Role of genomic architecture in *PLP1* duplication causing Pelizaeus–Merzbacher disease. *Hum. Mol. Genet.* **15**, 2250–2265 (2006).
35. Lee, J. A. *et al.* Spastic paraplegia type 2 associated with axonal neuropathy and apparent *PLP1* position effect. *Ann. Neurol.* **59**, 398–403 (2006).
36. Lovett, S. T., Hurley, R. L., Sutura, V. A. Jr, Aubuchon, R. H. & Lebedeva, M. A. Crossing over between regions of limited homology in *Escherichia coli*. RecA-dependent and RecA-independent pathways. *Genetics* **160**, 851–859 (2002).
37. Liskay, R. M., Letsou, A. & Stachek, J. L. Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. *Genetics* **115**, 161–167 (1987).
38. Reiter, L. T. *et al.* Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. *Am. J. Hum. Genet.* **62**, 1023–1033 (1998).
39. Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
40. Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. *Annu. Rev. Genet.* **38**, 233–271 (2004).
41. Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).
42. Esposito, M. S. Evidence that spontaneous mitotic recombination occurs at the two-strand stage. *Proc. Natl Acad. Sci. USA* **75**, 4436–4440 (1978).
43. Stark, J. M. & Jasin, M. Extensive loss of heterozygosity is suppressed during homologous repair of chromosomal breaks. *Mol. Cell. Biol.* **23**, 733–743 (2003).
44. Dupaigne, P. *et al.* The Srs2 helicase activity is stimulated by Rad51 filaments on dsDNA: implications for crossover incidence during mitotic recombination. *Mol. Cell.* **29**, 243–254 (2008).
45. Sun, W. *et al.* The FANCM ortholog Fml1 promotes recombination at stalled replication forks and limits crossing over during DNA double-strand break repair. *Mol. Cell* **32**, 118–128 (2008).
46. Ira, G., Malkova, A., Liberi, G., Foiani, M. & Haber, J. E. Srs2 and Sgs1–Top3 suppress crossovers during double-strand break repair in yeast. *Cell* **115**, 401–411 (2003).
47. Prakash, R. *et al.* Yeast Mph1 helicase dissociates Rad51-made D-loops: implications for crossover control in mitotic recombination. *Genes Dev.* **23**, 67–79 (2009).
48. Wu, L. & Hickson, I. D. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* **426**, 870–874 (2003).
49. Prado, F. & Aguilera, A. Control of cross-over by single-strand DNA resection. *Trends Genet.* **19**, 428–431 (2003).
50. Smith, C. E., Llorente, B. & Symington, L. S. Template switching during break-induced replication. *Nature* **447**, 102–105 (2007). **This paper shows experimental evidence from yeast on the nature of BIR. Specifically, template switching between homologous chromosomes (or sometimes non-homologous chromosomes, which causes translocation). It showed replication out to the telomere after switching.**
51. Bauters, M. *et al.* Nonrecurrent *MECP2* duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair. *Genome Res.* **18**, 847–858 (2008).
52. Deem, A. *et al.* Defective break-induced replication leads to half-crossovers in *Saccharomyces cerevisiae*. *Genetics* **179**, 1845–1860 (2008).
53. Narayanan, V. & Lobachev, K. S. Intrachromosomal gene amplification triggered by hairpin-capped breaks requires homologous recombination and is independent of nonhomologous end-joining. *Cell Cycle* **6**, 1814–1818 (2007).
54. Payen, C., Koszul, R., Dujon, B. & Fischer, G. Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. *PLoS Genet.* **4**, e1000175 (2008).
- Evidence from yeast that LCRs arise by a replicative mechanism, specifically one involving BIR.**
55. Schmidt, K. H., Wu, J. & Kolodner, R. D. Control of translocations between highly diverged genes by Sgs1, the *Saccharomyces cerevisiae* homolog of the Bloom's syndrome protein. *Mol. Cell. Biol.* **26**, 5406–5420 (2006).
56. Lin, F. L., 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.* **4**, 1020–1034 (1984).
57. Sweigert, S. E. & Carroll, D. Repair and recombination of X-irradiated plasmids in *Xenopus laevis* oocytes. *Mol. Cell. Biol.* **10**, 5849–5856 (1990).
58. Haber, J. E. Exploring the pathways of homologous recombination. *Curr. Opin. Cell Biol.* **4**, 401–412 (1992).
59. Elliott, B., Richardson, C. & Jasin, M. Chromosomal translocation mechanisms at intronic Alu elements in mammalian cells. *Mol. Cell* **17**, 885–894 (2005).
60. Ruysschaert, C., Thaler, D. S. & Radman, M. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch repair mutants. *Nature* **342**, 396–401 (1989).
61. Unal, E. *et al.* DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain. *Mol. Cell* **16**, 991–1002 (2004).
62. Strom, L., Lindroos, H. B., Shirahige, K. & Sjogren, C. Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. *Mol. Cell* **16**, 1003–1015 (2004).
63. Kim, J. S., Krasieva, T. B., LaMorte, V., Taylor, A. M. & Yokomori, K. Specific recruitment of human cohesin to laser-induced DNA damage. *J. Biol. Chem.* **277**, 45149–45153 (2002).
64. Sjogren, C. & Nasmyth, K. Sister chromatid cohesion is required for postreplicative double-strand break repair in *Saccharomyces cerevisiae*. *Curr. Biol.* **11**, 991–995 (2001).
65. Kobayashi, T., Horiuchi, T., Tongaonkar, P., Vu, L. & Nomura, M. SIR2 regulates recombination between different rDNA repeats, but not recombination within individual rDNA genes in yeast. *Cell* **117**, 441–453 (2004).
66. Kobayashi, T. & Ganley, A. R. Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. *Science* **309**, 1581–1584 (2005).
67. Kaye, J. A. *et al.* DNA breaks promote genomic instability by impeding proper chromosome segregation. *Curr. Biol.* **14**, 2096–2106 (2004).
68. Soutoglou, E. *et al.* Positional stability of single double-strand breaks in mammalian cells. *Nature Cell Biol.* **9**, 675–682 (2007).
69. Oh, S. D. *et al.* BLM ortholog, Sgs1, prevents aberrant crossing-over by suppressing formation of multichromatid joint molecules. *Cell* **130**, 259–272 (2007).
70. Jain, S. *et al.* A recombination execution checkpoint regulates the choice of homologous recombination pathway during DNA double-strand break repair. *Genes Dev.* **23**, 291–303 (2009).
71. McVey, M. & Lee, S. E. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* **24**, 529–538 (2008).
72. Lieber, M. R. The mechanism of human nonhomologous DNA end joining. *J. Biol. Chem.* **283**, 1–5 (2008).
73. Daley, J. M., Palmos, P. L., Wu, D. & Wilson, T. E. Nonhomologous end joining in yeast. *Annu. Rev. Genet.* **39**, 431–451 (2005).
74. Haviv-Chesner, A., Kobayashi, Y., Gabriel, A. & Kupiec, M. Capture of linear fragments at a double-strand break in yeast. *Nucleic Acids Res.* **35**, 5192–5202 (2007).
75. Yu, X. & Gabriel, A. Ku-dependent and Ku-independent end-joining pathways lead to chromosomal rearrangements during double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* **163**, 843–856 (2003).
76. Nickoloff, J. A., De Haro, L. P., Wray, J. & Hromas, R. Mechanisms of leukemic translocations. *Curr. Opin. Hematol.* **15**, 338–345 (2008).
77. McClintock, B. Chromosome organization and genic expression. *Cold Spring Harb. Symp. Quant. Biol.* **16**, 13–47 (1951).
78. Tanaka, H. & Yao, M. C. Palindromic gene amplification — an evolutionarily conserved role for DNA inverted repeats in the genome. *Nature Rev. Cancer* **9**, 216–224 (2009).
79. Tanaka, H., Bergstrom, D. A., Yao, M. C. & Tapscoft, S. J. Large DNA palindromes as a common form of structural chromosome aberrations in human cancers. *Hum. Cell* **19**, 17–23 (2006).
80. Coquelle, A., Pipiras, E., Toledo, F., Buttin, G. & Debatisse, M. Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell* **89**, 215–225 (1997).
81. Shaw, C. J. & Lupski, J. R. Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. *Hum. Genet.* **116**, 1–7 (2005).
82. Arlt, M. F. *et al.* Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *Am. J. Hum. Genet.* **84**, 339–350 (2009).
83. Durkin, S. G. *et al.* Replication stress induces tumor-like microdeletions in FHIT/FRA3B. *Proc. Natl Acad. Sci. USA* **105**, 246–251 (2008).
84. Kuo, M. T., Vyas, R. C., Jiang, L. X. & Hittelman, W. N. Chromosome breakage at a major fragile site associated with P-glycoprotein gene amplification in multidrug-resistant CHO cells. *Mol. Cell Biol.* **14**, 5202–5211 (1994).
85. Coquelle, A., Rozier, L., Dutrillaux, B. & Debatisse, M. Induction of multiple double-strand breaks within an hsr by meganuclease-SceI expression or fragile site activation leads to formation of double minutes and other chromosomal rearrangements. *Oncogene* **21**, 7671–7679 (2002).
86. Michel, B., Ehrlich, S. D. & Uzest, M. DNA double-strand breaks caused by replication arrest. *EMBO J.* **16**, 430–438 (1997).
87. Albertini, A. M., Hofer, M., Calos, M. P. & Miller, J. H. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell* **29**, 319–328 (1982).
88. Farabaugh, P. J., Schmeissner, U., Hofer, M. & Miller, J. H. Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *J. Mol. Biol.* **126**, 847–857 (1978).
89. Ikeda, H., Shimizu, H., Ukita, T. & Kumagai, M. A novel assay for illegitimate recombination in *Escherichia coli*: stimulation of lambda bio transducing phage formation by ultra-violet light and its independence from RecA function. *Adv. Biophys.* **31**, 197–208 (1995).
90. Shimizu, H. *et al.* Short-homology-independent illegitimate recombination in *Escherichia coli*: distinct mechanism from short-homology-dependent illegitimate recombination. *J. Mol. Biol.* **266**, 297–305 (1997).
91. Bi, X. & Liu, L. F. *recA*-independent and *recA*-dependent intramolecular plasmid recombination: differential homology and requirement and distance effect. *J. Mol. Biol.* **235**, 414–423 (1994).
92. Mazin, A. V., Kuzminov, A. V., Dianov, G. L. & Salganik, R. I. Mechanisms of deletion formation in *Escherichia coli* plasmids. II. Deletions mediated by short direct repeats. *Mol. Gen. Genet.* **228**, 209–214 (1991).
93. Chedin, F., Dervyn, E., Dervyn, R., Ehrlich, S. D. & Noiret, P. Frequency of deletion formation decreases exponentially with distance between short direct repeats. *Mol. Microbiol.* **12**, 561–569 (1994).
94. Lovett, S. T., Gluckman, T. J., Simon, P. J., Sutura Jr, V. A. & Drapkin, P. T. Recombination between repeats in *Escherichia coli* by a *recA*-independent, proximity-sensitive mechanism. *Mol. Gen. Genet.* **254**, 294–300 (1994).
95. Biernie, H., Vilette, D., Ehrlich, S. D. & Michel, B. Isolation of a *dnaE* mutation which enhances RecA-independent homologous recombination in the *Escherichia coli* chromosome. *Mol. Microbiol.* **24**, 1225–1234 (1997).
96. Saveson, C. J. & Lovett, S. T. Enhanced deletion formation by aberrant DNA replication in *Escherichia coli*. *Genetics* **146**, 457–470 (1997).
97. Bzymek, M. & Lovett, S. T. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc. Natl Acad. Sci. USA* **98**, 8319–8325 (2001).
98. Lovett, S. T. & Feshenko, V. V. Stabilization of diverged tandem repeats by mismatch repair: evidence for deletion formation via a misaligned replication intermediate. *Proc. Natl Acad. Sci. USA* **93**, 7120–7124 (1996).

99. Cairns, J. & Foster, P. L. Adaptive reversion of a frameshift mutation in *Escherichia coli*. *Genetics* **128**, 695–701 (1991).
100. Slack, A., Thornton, P. C., Magner, D. B., Rosenberg, S. M. & Hastings, P. J. On the mechanism of gene amplification induced under stress in *Escherichia coli*. *PLoS Genet.* **2**, e48 (2006).
Experimental evidence from *E. coli* suggesting that chromosomal structural change occurs by a replicative mechanism, and also revealing that the characteristics of amplification in *E. coli* are similar to those of non-recurrent changes seen in human genomic disorders. Proposes CNV formation by template switching between forks.
101. Kugelberg, E., Kofoid, E., Reams AB, Andersson, D. I. & Roth, J. R. Multiple pathways of selected gene amplification during adaptive mutation. *Proc. Natl Acad. Sci. USA* **103**, 17319–17324 (2006).
102. Allgood, N. D. & Silhavy, T. *Escherichia coli xonA (sbcB)* mutants enhance illegitimate recombination. *Genetics* **127**, 671–680 (1991).
103. Bzymek, M., Saveson, C. J., Feschenko, V. V. & Lovett, S. T. Slipped misalignment mechanisms of deletion formation: *in vivo* susceptibility to nucleases. *J. Bacteriol.* **181**, 477–482 (1999).
104. Lydeard, J. R., Jain, S., Yamaguchi, M. & Haber, J. E. Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* **448**, 820–823 (2007).
105. VanHulle, K. *et al.* Inverted DNA repeats channel repair of distant double-strand breaks into chromatid fusions and chromosomal rearrangements. *Mol. Cell Biol.* **27**, 2601–2614 (2007).
106. Davis, A. P. & Symington, L. S. RAD51-dependent break-induced replication in yeast. *Mol. Cell Biol.* **24**, 2344–2351 (2004).
107. McVey, M., Adams, M., Staeve-Vieira, E. & Sekelsky, J. J. Evidence for multiple cycles of strand invasion during repair of double-strand gaps in *Drosophila*. *Genetics* **167**, 699–705 (2004).
108. Bindra, R. S., Crosby, M. E. & Glazer, P. M. Regulation of DNA repair in hypoxic cancer cells. *Cancer Metastasis Rev.* **26**, 249–260 (2007).
109. Huang, L. E., Bindra, R. S., Glazer, P. M. & Harris, A. L. Hypoxia-induced genetic instability — a calculated mechanism underlying tumor progression. *J. Mol. Med.* **85**, 139–148 (2007).
110. Bindra, R. S. & Glazer, P. M. Repression of RAD51 gene expression by E2F4/p130 complexes in hypoxia. *Oncogene* **26**, 2048–2057 (2007).
111. Coquelle, A., Toledo, F., Stern, S., Bieth, A. & Debatisse, M. A new role for hypoxia in tumor progression: induction of fragile site triggering genomic rearrangements and formation of complex DMs and HSRs. *Mol. Cell* **2**, 259–265 (1998).
112. Hastings, P. J., Bull, H. J., Klump, J. R. & Rosenberg, S. M. Adaptive amplification: an inducible chromosomal instability mechanism. *Cell* **103**, 723–731 (2000).
113. Lombardo, M.-J., Aponyi, I. & Rosenberg, S. M. General stress response regulator RpoS in adaptive mutation and amplification in *Escherichia coli*. *Genetics* **166**, 669–680 (2004).
114. Ponder, R. G., Fonville, N. C. & Rosenberg, S. M. A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Mol. Cell* **19**, 791–804 (2005).
115. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).
116. Shao, L. *et al.* Identification of chromosome abnormalities in subtelomeric regions by microarray analysis: a study of 5,380 cases. *Am. J. Med. Genet. A* **146A**, 2242–2251 (2008).
117. Yatsenko, S. A. *et al.* Molecular mechanisms for subtelomeric rearrangements associated with the 9q34.3 microdeletion syndrome. *Hum. Mol. Genet.* **18**, 1924–1936 (2009).
118. Zhang, L., Lu, H. H., Chung, W. Y., Yang, J. & Li, W. H. Patterns of segmental duplication in the human genome. *Mol. Biol. Evol.* **22**, 135–141 (2005).
119. She, X. *et al.* The structure and evolution of centromeric transition regions within the human genome. *Nature* **430**, 857–864 (2004).
120. Nguyen, D. Q., Webber, C. & Ponting, C. P. Bias of selection on human copy-number variants. *PLoS Genet.* **2**, e20 (2006).
121. Visser, R. *et al.* Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion. *Am. J. Hum. Genet.* **76**, 52–67 (2005).
122. de Smith, A. J. *et al.* Small deletion variants have stable breakpoints commonly associated with *Alu* elements. *PLoS ONE* **3**, e3104 (2008).
123. Bacolla, A. *et al.* Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc. Natl Acad. Sci. USA* **101**, 14162–14167 (2004).
124. Bacolla, A., Wojciechowska, M., Kosmider, B., Larson, J. E. & Wells, R. D. The involvement of non-B DNA structures in gross chromosomal rearrangements. *DNA Repair (Amst.)* **5**, 1161–1170 (2006).
125. Inagaki, H. *et al.* Chromosomal instability mediated by non-B DNA: cruciform conformation and not DNA sequence is responsible for recurrent translocation in humans. *Genome Res.* **19**, 191–198 (2009).
126. Myers, S., Freeman, C., Auton, A., Donnelly, P. & McVean, G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature Genet.* **40**, 1124–1129 (2008).
127. Shaw, C. J. & Lupski, J. R. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum. Mol. Genet.* **13**, R57–R64 (2004).
128. Bi, W. *et al.* Increased LIS1 expression affects human and mouse brain development. *Nature Genet.* **41**, 168–177 (2009).
129. Galhardo, R. S., Hastings, P. J. & Rosenberg, S. M. Mutation as a stress response and the regulation of evolvability. *Crit. Rev. Biochem. Mol. Biol.* **42**, 399–435 (2007).
A broad review of stress-induced mutation and its meaning for evolution.
130. Rosenberg, S. M. Evolving responsively: adaptive mutation. *Nature Rev. Genet.* **2**, 504–515 (2001).
131. Cirz, R. T. *et al.* Inhibition of mutation and combating the evolution of antibiotic resistance. *PLoS Biol.* **3**, e176 (2005).
132. Riesenfeld, C., Everett, M., Piddock, L. J. & Hall, B. G. Adaptive mutations produce resistance to ciprofloxacin. *Antimicrob. Agents Chemother.* **41**, 2059–2060 (1997).
133. Bindra, R. S. S. *et al.* Down-regulation of Rad51 and decreased homologous recombination in hypoxic cancer cells. *Mol. Cell Biol.* **24**, 8504–8518 (2004).
This paper shows that HR enzymes are downregulated by stress in human cells, and that this is accompanied by reduction in HR.
134. Mihaylova, V. T. *et al.* Decreased expression of the DNA mismatch repair gene *Mlh1* under hypoxic stress in mammalian cells. *Mol. Cell Biol.* **23**, 3265–3273 (2003).
135. Kolodner, R. D. *et al.* Germ-line *msh6* mutations in colorectal cancer families. *Cancer Res.* **59**, 5068–5074 (1999).
136. Loeb, L. A. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* **51**, 3075–3079 (1991).
137. Modrich, P. Mismatch repair, genetic stability and tumour avoidance. *Phil. Trans. R. Soc. Lond. B* **347**, 89–95 (1995).
138. Nguyen, D. Q. *et al.* Reduced purifying selection prevails over positive selection in human copy number variant evolution. *Genome Res.* **18**, 1711–1723 (2008).
139. Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nature Genet.* **39**, 1256–1260 (2007).
140. Gonzalez, E. *et al.* The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434–1440 (2005).
141. Higgs, D. R. *et al.* A review of the molecular genetics of the human alpha-globin gene cluster. *Blood* **73**, 1081–1104 (1989).
142. Nozawa, M., Kawahara, Y. & Nei, M. Genomic drift and copy number variation of sensory receptor genes in humans. *Proc. Natl Acad. Sci. USA* **104**, 20421–20426 (2007).
143. Jakobsson, M. *et al.* Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998–1003 (2008).
144. Locke, D. P. *et al.* Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am. J. Hum. Genet.* **79**, 275–290 (2006).
145. Sharp, A. J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
146. Fortna, A. *et al.* Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol.* **2**, E207 (2004).
147. McCarrroll, S. A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genet.* **40**, 1166–1174 (2008).
148. Turner, D. J. *et al.* Germline rates of *de novo* meiotic deletions and duplications causing several genomic disorders. *Nature Genet.* **40**, 90–95 (2008).
149. Lam, K. W. & Jeffreys, A. J. Processes of copy-number change in human DNA: the dynamics of α -globin gene deletion. *Proc. Natl Acad. Sci. USA* **103**, 8921–8927 (2006).
150. Lam, K. W. & Jeffreys, A. J. Processes of *de novo* duplication of human α -globin genes. *Proc. Natl Acad. Sci. USA* **104**, 10950–10955 (2007).
151. Flores, M. *et al.* Recurrent DNA inversion rearrangements in the human genome. *Proc. Natl Acad. Sci. USA* **104**, 6099–6106 (2007).
152. Liang, Q., Conte, N., Skarnes, W. C. & Bradley, A. Extensive genomic copy number variation in embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 17453–17456 (2008).
153. Lupski, J. R. Genomic rearrangements and sporadic disease. *Nature Genet.* **39**, S43–47 (2007).
154. van Ommen, G. J. Frequency of new copy number variation in humans. *Nature Genet.* **37**, 333–334 (2005).
155. Tuzun, E., Bailey, J. A. & Eichler, E. E. Recent segmental duplications in the working draft assembly of the brown Norway rat. *Genome Res.* **14**, 493–506 (2004).
156. Graubert, T. A. *et al.* A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet.* **3**, e3 (2007).
157. Lovett, S. T. Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* **52**, 1243–1253 (2004).

Acknowledgements

This work was supported by grants from the National Institutes of Health, R01 GM64022 to P.J.H., R01 NS59529 to J.R.L., R01 GM53158 and R01 CA85777 to S.M.R., and R01 GM80600 to G.I.

FURTHER INFORMATION

Philip J. Hastings' homepage:
<http://www.bcm.edu/genetics/faculty/hastings.html>
James R. Lupski's homepage:
<http://www.bcm.edu/genetics/faculty/lupski.html>
Susan M. Rosenberg's homepage:
<http://www.bcm.edu/genetics/faculty/rosenberg.html>
Grzegorz Ira's homepage:
<http://www.bcm.edu/genetics/faculty/ira.html>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF