Genome instability: a mechanistic view of its causes and consequences

Andrés Aguilera and Belén Gómez-González

Abstract | Genomic instability in the form of mutations and chromosome rearrangements is usually associated with pathological disorders, and yet it is also crucial for evolution. Two types of elements have a key role in instability leading to rearrangements: those that act *in trans* to prevent instability — among them are replication, repair and S-phase checkpoint factors — and those that act *in cis* — chromosomal hotspots of instability such as fragile sites and highly transcribed DNA sequences. Taking these elements as a guide, we review the causes and consequences of instability with the aim of providing a mechanistic perspective on the origin of genomic instability.

Spindle mitotic checkpoint
A quality-control mechanism that blocks anaphase entry, arresting cell growth until all chromosomes are properly attached to the mitotic spindle to achieve their accurate segregation.

Centro Andaluz de Biologia Molecular y Medicina Regenerativa CABIMER, Universidad de Sevilla-CSIC, Avd. Américo Vespucio s/n, 41092 Sevilla, Spain
Correspondence to A.A.
e-mail: aguilo@us.es
doi:10.1038/nrg2268
Published online
29 January 2008

Cell proliferation involves numerous processes that need to be tightly coordinated to ensure the preservation of genome integrity and to promote faithful genome propagation. Efficient and error-free DNA replication is key for the faithful duplication of chromosomes before their segregation. Coordination of DNA replication with DNA-damage sensing and repair and cell-cycle progression ensures, with a high probability, genome integrity during cell divisions, thus preventing mutations and DNA rearrangements. Although such events can be harmful for the cell and the organism, they also drive evolution at the molecular level and generate genetic variation. Genetic instability can also have a specialized role in the generation of variability in developmentally regulated processes, such as immunoglobulin (Ig) diversification[1]. Nevertheless, genetic instability is usually associated with pathological disorders, and in humans it is often associated with premature ageing, predisposition to various types of cancer and with inherited diseases (TABLE 1). Notably, the well-established principle that cancer risk increases with age has a parallel in yeast, in which ageing is associated with increased genetic instability[2].

Genetic instability refers to a range of genetic alterations from point mutations to chromosome rearrangements and can be divided into classes according to the type of event stimulated. Chromosomal instability (CIN) refers to changes in chromosome number that lead to chromosome gain or loss[3]. CIN is caused by failures in either mitotic chromosome transmission or the spindle mitotic checkpoint. Micro- and minisatellite instability (MIN) leads to repetitive-DNA expansions and contractions and can occur by replication slippage, by mismatch repair (MMR) impairment or by homologous recombination (HR). Instability leading to mutations, including base substitutions, micro-insertions and micro-deletions, is mainly associated with replication errors, impairment of base excision repair (BER) and MMR, or error-prone translesion synthesis. Instability leading to rearrangements refers to events that involve changes in the genetic linkage of two DNA fragments. Increases in HR-mediated events — such as unequal sister-chromatid exchange (SCE) and ectopic HR between non-allelic repeated DNA fragments — or in end-joining between non-homologous DNA fragments can result in gross chromosomal rearrangements (GCRs) such as translocations, duplications, inversions or deletions. What these instability events leading to rearrangements have in common is that they are generated by DNA breaks. As the term GCR is primarily used to refer to the events that occur between non-homologous or divergent DNA sequences, in this Review we refer to HR and GCR events separately to emphasize the mechanistic differences between them. Despite the broad spectrum of proteins and breakpoints that are associated with rearrangements, the common feature is their association with replication stress. Indeed, replication failures seem to be the primary cause of cancer; the checkpoint that senses DNA damage and replication failures is a potential anticancer barrier[4].

Genomic instability that leads to mutations has been reviewed extensively elsewhere (for example, REF. 5). Although we briefly discuss MIN, the recent advancements in our understanding of the mechanisms of replication fork progression and checkpoints — together with the burst of information about *cis* and *trans* elements that control instability leading to rearrangements — provide an excellent opportunity to bring all these important

Replisome
A complex of proteins involved in DNA replication elongation that moves along the DNA as the nascent strands are synthesized.

discoveries together to get a broad overview of genome instability (GIN). Here we review the causes and consequences of instability that lead to rearrangements, with the aim of providing an integral view of the mechanisms that are responsible for the preservation of chromosome integrity. For simplicity, we will use the term GIN, even though we only refer to instability leading to rearrangements. We begin with an overview of the basic aspects of replication-fork (RF) progression and how defects in this process can become a source of DNA breaks.

Replication as a source of DNA breaks

Under natural conditions, DNA is most vulnerable during the S phase of the cell cycle. During replication the replisome must overcome obstacles (such as DNA adducts, secondary structures or tightly bound proteins) that can cause RF stalling, which can compromise genome integrity if not properly processed (FIG. 1). Eukaryotic cells have developed checkpoint functions that are constantly monitoring the integrity of the DNA and serve to coordinate replication with

Table 1 | **A selection of eukaryotic genes with a role in the maintenance of genome integrity (part 1)**

Saccharomyces cerevisiae	Mammals	Function	Human disease	GCR	HR
Replication					
CAC1, *2* and *3*	*CHAF1A* and *B*, *p48*	Chromatin assembly	Not known	High	High
ASF1	*ASF1A* and *B*	Chromatin assembly	Not known	High	High
TOP1	*TOP1*	Topoisomerase I	Not known	Not known	High
TOP2	*TOP2A* and *B*	Topoisomerase II	Not known	Not known	High
MCM4	*MCM4*	Replicative helicase subunit	Cancer predisposition	Not known	High
ORC3–ORC5	*ORC3–6L*	Origin replication complex	Not known	High	Not known
CDC6	*CDC6*	Replication initiation	Not known	Not known	High
CDC9	*LIG1*	Ligase I	Not known	Not known	High
POL1–PRI1 and *2*	*POLA1– PRIM1* and *2*	Polymerase α–Primase complex	Not known	High	High
DPB11	*TOPBP11*	Polymerase ε subunit, checkpoint mediator	Not known	High	Not known
POL3 (also called *CDC2*)	*POLD1*	Polymerase δ	Not known	Not known	High
POL30	*PCNA*	PCNA	Not known	Normal	High
RAD27	*FEN1*	Flap endonuclease	Not known	High	High
RFA1, *2* and *3*	*RPA70*, *32* and *14*	Replication factor A, checkpoint signalling	Cancer predisposition	High	High
PIF1	*PIF1*	RNA–DNA helicase	Not known	High	Not known
RRM3	Not known	DNA Helicase	Not known	Normal	High
RFC1–5	*RFC1–5*	Clamp loader, checkpoint sensor	Not known	High	High
Checkpoint					
PDS1	*PTTG1*	Mitotic arrest	Not known	High	Not known
RAD9	Not known	Damage-checkpoint mediator	Not known	High	High
MEC1	*ATR*	Transducer kinase	Seckel syndrome	High	High
TEL1	*ATM*	Transducer kinase	Ataxia telangiectasia; cancer predisposition	High	High
CHK1	*CHEK1*	Effector kinase	Rare tumours; cancer predisposition	High	Not known
RAD53	*CHEK2*	Effector kinase	Li-Fraumeni syndrome variant; cancer predisposition	High	Not known
DUN1	Not known	Effector kinase	Not known	High	High
DDC1–RAD17–MEC3	*RAD9–RAD1–HUS1*	PCNA-like complex	Not known	High	Not known
RAD24	*RAD17*	RFC-like, S-phase checkpoint	Not known	High	High
CTF18	*CHTF18*	RFC-like, cohesion	Not known	Not known	High
ELG1	Not known	RFC-like	Not known	High	High
DDC2	*ATRIP*	Signalling	Not known	High	High

The table lists genes in which mutations have been shown to increase homologous recombination (HR), gross chromosomal rearrangements (GCR) or both, without specifying whether the increase is strong or weak. See Supplementary information S1 (table) for the references from which the information was derived. PCNA, proliferating cell nuclear antigen; RFC, replication factor C complex.

Table 1 | **A selection of eukaryotic genes with a role in the maintenance of genome integrity (part 2)**

Saccharomyces cerevisiae	Mammals	Function	Human disease	GCR	HR
DSB repair					
MRE11	*MRE11*	HR and NHEJ	Ataxia telangiectasia-like disease	High	High
XRS2	*NBS1*	HR and NHEJ	Nijmegen breakage syndrome, cancer predisposition	High	Not known
RAD50	*RAD50*	HR and NHEJ	Not known	High	High
RAD52	*RAD52*	HR	Not known	High	Not known
RAD51, *54*, *57* and *59*	*RAD51*, *51B-D*, *XRCC2*, *3* and *RAD54*	HR	Not known	High	Not known
Not known	*BRCA1*	Damage checkpoint mediator	Familial breast, ovarian cancer	High	High
Not known	*BRCA2* (also known as *FANC-D1*)	Repair of crosslinks, HR	Familial breast cancer	High	Not known
YKU70–YKU80	*KU70–KU80*	NHEJ	Not known	High	Not known
DNL4	*LIG4*	NHEJ	Lig4 syndrome	High	Not known
H2A	*H2AX*	Chromatin decondensation	Not known	High	Not known
SRS2	*FBH1*	HR	Not known	Normal	High
TOP3	*TOP3A* and *B*	HR	Not known	High	High
SGS1	*BLM*	RecQ helicase	Bloom syndrome; cancer predisposition	High	High
	WRN	RecQ helicase	Werner syndrome; cancer predisposition	High	Not known
Other repair pathways					
RAD5	Not known	Post-replicative repair	Not known	High	High
RAD18	*RAD18*	Post-replicative repair	Not known	High	High
MSH2	*MSH2*	MMR	HNPCC; cancer predisposition	High	Not known
Not known	*FANCA–G*, *D1* (also known as *BRCA2*), *D2* and *L*	Crosslinkage repair	Fanconi Anemia; cancer predisposition	High	High
mRNP biogenesis					
HPR1–THO2–MFT1–THP2	*THOC1*, *2* and *THOC5–7*	THO complex, mRNP biogenesis	Not known	Not known	High
SUB2–YRA1	*UAP56–ALY*	mRNP biogenesis to export	Not known	Not known	High
Not known	*ASF/SF2*	Pre-mRNA splicing	Not known	High	Not known
THP1–SAC3	*SHD1* and/or *GANP*	mRNA export	Not known	Not known	High
RRP6	*EXOSC10*	Nuclear exosome	Not known	Not known	High
RNA14	*CSTF3*	mRNA 3′-end processing	Not known	Not known	High
NAB2	Not known	hnRNP	Not known	Not known	High
Others					
TSA1	*PRDX2*	Thioredoxin peroxidase	Not known	High	Not known
CDC5	*CDC5L*	Cell-cycle regulator kinase	Not known	Not known	High
CDC13	*POT1*	Telomere capping	Not known	Not known	High
CDC14	*CDC14A* and *B*	Cell-cycle phosphatase, mitotic exit	Not known	Not known	High
YCS4	*NCAPD2*	Mitotic condensation	Not known	High	Not known
SMC5–SMC6	*SMC5–SMC6*	Cohesion-related and/or repair	Not known	High	Not known
SIC1	Not known	G1–S transition	Not known	High	Not known

The table lists genes in which mutations have been shown to increase homologous recombination (HR), gross chromosomal rearrangements (GCR) or both, without specifying whether the increase is strong or weak. See Supplementary information S1 (table) for the references from which the information was derived. HNPCC, hereditary non-polyposis colorectal cancer; hnRNP, heterogeneous nuclear ribonucleoprotein; MMR, mismatch repair; mRNP, messenger ribonucleoprotein; NHEJ, non-homologous end joining.

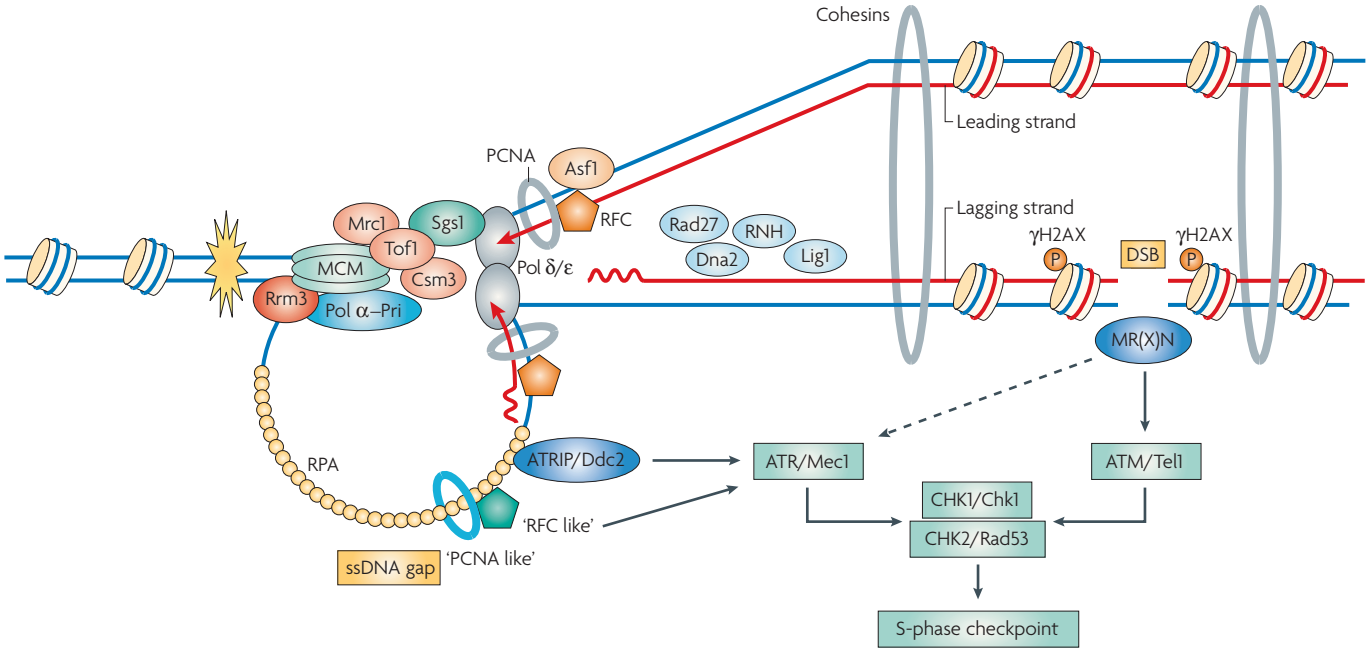

Figure 1 | **Replication fork progression and stalling.** Replication is catalysed by the replisome multisubunit complex, which contains replication elongation factors. The eukaryotic MCM helicase unwinds the parental duplex to allow access to the DNA polymerase α primase (Pol α–Pri). Processive elongation is catalysed by the replicative polymerases δ and ε (Pol δ/ε). This polymerase switch is mediated by the replication processivity clamp (PCNA) that is loaded by the replication factor C complex (RFC) and stabilizes DNA polymerases. During DNA synthesis in the discontinuous lagging strand, primer RNA displacement results in a flap structure that is cleaved by the Rad27 nuclease (FEN1 in humans); the resulting intermediate is processed by the DNA2 helicase, RNase H (RNH), pol δ and DNA ligase I (Lig1). Cohesins hold the two sister chromatids together until anaphase. Encountering an obstacle can cause replication-fork (RF) stalling, leading to ssDNA gaps and double-stranded breaks (DSBs). Several factors associate with the RF to prevent its collapse, including the *Saccharomyces cerevisiae* Rrm3 helicase, the Mrc1 checkpoint mediator in association with Tof1 and Csm3, or the nucleosome assembly factor Asf1. ssDNA gaps and DSBs are sensed by the S-phase checkpoint which is activated through Tel1 (ATM in humans) and Mec1 (ATR in humans) (see BOX 1). In the case of a DSB, the checkpoint signalling spreads around the DSB site by histone H2AX phosphorylation (γH2AX) in humans (H2A in yeast). ATRIP/Ddc2, ATR/Mec1 interacting protein; CHK1/Chk1 and CHK2/Rad53, serine/threonine-protein kinases; MCM, replicative helicase; MR(X)N, a nuclease complex; RPA, replication protein A; Sgs1, ATP-dependant helicase.

repair, chromosome segregation and cell-cycle progression. The importance of the checkpoint response as a DNA surveillance mechanism is reflected by the fact that most checkpoint factors are evolutionarily conserved and many are tumour suppressors (TABLE 1). Among the cellular checkpoints, S-phase checkpoints are crucial for maintaining genome integrity — they respond to RF stalling and intra-S-phase damage by preventing RF collapse and breakdown (BOX 1). Consistently, yeast S-phase checkpoint mutants, such as *rfc5*, show a strong increase in genome rearrangements, whereas mutations in genes involved in G1 and G2 damage checkpoints, such as *RAD9*, have little effect on rearrangements[6,7].

When a stalled RF remains associated with a functional replisome, DNA synthesis can resume after the obstacle has been removed. RFs can also be delayed to allow coordination of fork processing and repair with replication resumption[8]. Nevertheless, under replication stress, which is caused by a replication inhibition and/or S-phase checkpoint inactivation (BOX 1), RFs can collapse causing replisome disassembly, ssDNA gaps and DNA breaks[9,10] (FIG. 1).

Replication-associated DNA breaks can be generated in several ways. First, an RF can encounter a ssDNA nick, which results in discontinued synthesis of the nascent strand and leads to a double-stranded break (DSB)[11]. If the nick is on the leading strand, the DSB would be one-ended and could promote the restart of synthesis by break-induced replication (BIR) (FIG. 2a). Second, RF progression or leading-strand synthesis can be blocked; in this case, leading-strand and lagging-strand synthesis are uncoupled[12–14]. This can be followed by fork reversal leading to the formation of a Holliday junction (HJ) or a 'chicken-foot' structure if the replisome is destabilized[10,15,16]. An RF can restart after reversal of the HJ back to a fork, or following HJ-cleavage and BIR (reviewed in REF. 17) (FIG. 2b). Third, lesions could block the synthesis of only one DNA strand without impeding fork progression. A lesion that blocks lagging-strand synthesis creates a ssDNA gap (or a DSB if the lesion is a nick) between two flanking Okazaki fragments. Lesions other than nicks that block leading-strand synthesis can be bypassed by the RF, which restarts downstream of the lesion leaving ssDNA gaps behind[13,18] (FIG. 2c). ssDNA gaps are then repaired by error-prone translesion synthesis (TLS) or

DNA adduct
A DNA sequence that is covalently-bound to a chemical residue such as cisplatin or benzopyrene.

Loss-of-heterozygosity (LOH). The loss of one of the alleles at a given locus as a result of a genomic change, such as mitotic deletion, gene conversion or chromosome missegregration.

preferentially by error-free HR using the sister chromatid as a template[19]. In the current view, HR is initiated by DSBs, but the possibility that ssDNA gaps could initiate HR repair without being converted into DSBs first cannot yet be excluded[20]. RF stalling and collapse can therefore elicit DNA breaks that could be a source of GIN (FIG. 3).

Elements contributing to genome instability

The development of genetic assays in *Saccharomyces cerevisiae* in the early 1980s was crucial to the study of genetic regulation of GIN in eukaryotes. The observation that artificially constructed DNA repeats could be used as active HR substrates[21,22] opened up the possibility of using genetic assays for the identification of mutations that increased HR between repeated DNA fragments[23] (FIG. 3). HR between artificial repeats can serve as a model for GIN that involves interspersed repeat sequences (LINES and SINES) in mammals. HR-mediated deletions between ectopic repeats are one source of loss-of-heterozygosity (LOH). However, many rearrangements associated with cancer or genetic diseases in fact occur between non-homologous or divergent DNA regions, implying that they can occur by mechanisms other than HR (FIG. 3). Our understanding of the origin of such events was greatly enhanced by the development in the late 1990s of yeast genetic assays that focus on GCRs involving one chromosome arm, between a centromere-linked locus and a telomere-proximal marker[24,25].

Studies in the past three decades, primarily in yeast and mammals, have led to the identification of three key contributors to GIN: suppressors — proteins that act *in trans* to prevent GIN; fragile sites — DNA regions that are frequently found at the breakpoints of GIN events; and transcription — highly transcribed DNA regions show high recombination frequencies.

Suppressors of genome instability

In this Review we use the term suppressor to define a gene or protein that acts *in trans* to preserve genomic integrity. The large and increasing number of eukaryotic functions with a suppressor activity makes it impossible to evaluate them here individually. For this reason, information on functionally relevant suppressors can be found in TABLE 1 (full references can be found in the Supplementary information S1 (table)); this section focuses on suppressors from the perspective of the different biological processes in which they function. Suppressors are classified into groups according to their functional relationships. It is worth noting that in the term suppressor we also include replication factors; even though their primary function is genome replication, we want to include all proteins that, through their proper action, prevent the formation of intermediates that could lead to instability.

Classic genetic studies in *Escherichia coli* and *S. cerevisiae* have shown that mutations in DNA ligase I, DNA polymerase I (α in *S. cerevisiae*), DNA polymerase III (δ in *S. cerevisiae*), Replication factor A, DNA thymidylate kinase or DNA adenosine methylase strongly increases the levels of spontaneous chromosomal exchanges[23,26–28] (FIG. 4a). These findings have provided evidence that impairment of replication can be a source of recombinogenic DNA breaks. Mutations in many loci can result in increased recombination; among them are those that encode the Rad27 Flap endonuclease, which is involved in the removal of RNA primers from Okazaki fragments, the Sgs1 helicase and the nucleosome assembly factors Cac1 and Asf1 that are required during replication[29–32] (TABLE 1). This list of mutations indicates that hyper-recombination is predominantly associated with substrates that arise from replication failures. A number of replication mutants, such as *rfa1* and *rad27*, are synthetic lethal with *rad52* and other mutations that affect HR[24,29], implying that DSBs are the most frequent outcome of such replication failures. Breaks can result from defective sealing of Okazaki fragments or from RF stalling. Two lines of evidence support the latter: HR is induced by the natural RF block near the *MAT* locus in *Schizosaccharomyces pombe*[33] and transcription causes RF stalling in *S. cerevisiae*[34,35].

GCRs are stimulated in yeast by mutations that affect replication proteins such as Rfa1, Rad27, the chromatin assembly factors Cac1 and Asf1, or the cyclin dependent kinase (CDK) inhibitor Sic1 that cause premature entry into S phase and, in several cases, accumulation

Box 1 | Replication and the S-phase checkpoint

During replication, the MCM helicase unwinds the parental duplex to allow access to the DNA polymerase α primase that synthesizes RNA primers, which are elongated by the replicative polymerases δ and ε (FIG. 1). Replication through obstacles can make the replisome pause or stall. The S-phase checkpoint responds to replication fork (RF) stalling and to intra-S-phase damage (mainly ssDNA gaps and double-stranded breaks (DSBs)), preventing the firing of late replication origins and entry into mitosis. In this way the checkpoint contributes to the maintenance of functional forks by preventing their collapse[10,141,142]. Several factors associate with the RF to prevent stalling or collapse. In *Saccharomyces cerevisiae*, these include the Rrm3 helicase[143], required for RF progression through natural impediments, Mrc1, which forms a complex with Tof1 and Csm3 and functions in RF maintenance together with the Sgs1 helicase, and the Asf1 chromatin assembly factor[9,144–146] (FIG. 1).

The mammalian transducer kinases ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and RAD3 related (ATR) are the key players in triggering the S-phase checkpoint response. ATR acts in response to stalled RFs and other types of damage that lead to the accumulation of ssDNA, such as UV-induced damage or resected DSBs, whereas ATM responds directly to DSBs to which it is recruited through the MRN complex (MRE11–RAD50–NBS1) (reviewed in REF. 147). ATR is recruited by its cofactor ATRIP, which recognizes RPA-coated ssDNA, but requires further activation by the RAD9–RAD1–HUS1 (9–1–1) replication processivity clamp (PCNA)-like complex, which is loaded onto stalled forks by the RAD17 'RFC-like' complex. ATR and ATM kinases phosphorylate the effector kinases CHK1 and CHK2 to trigger the checkpoint response (reviewed in REF. 147). In this sequence of events, MCM is phosphorylated, which contributes to its association with active forks; in yeast, the ATR orthologue Mec1 phosphorylates Sgs1 and Mrc1 to prevent replisome disassembly and collapse[9,144] (FIG. 1).

Restarting of the RF is mediated by the S-phase checkpoint to prevent unscheduled recombination[148]. This was deduced from the observation that yeast S-phase checkpoint mutations (such as *rad53*) and mutations in checkpoint targets (such as *Sgs1*) cause fork collapse in the presence of replication inhibitors, and accumulation of Holliday junction structures[9,10,33,149]. If DSBs are generated, histone H2AX is phosphorylated (γH2AX) at its C-terminal tail[150] as one of the earlier events occurring at the break site[151]. γH2AX spreads around the break, thus amplifying the initial damage signal and resulting in large, megabase-long chromatin domains that are responsible for the stable accumulation of damage-response and cohesion factors that favour repair by sister-chromatid exchange[152,153]. If this entire process is disrupted by replication stress or S-phase checkpoint inactivation, breaks accumulate that could trigger genomic instability.

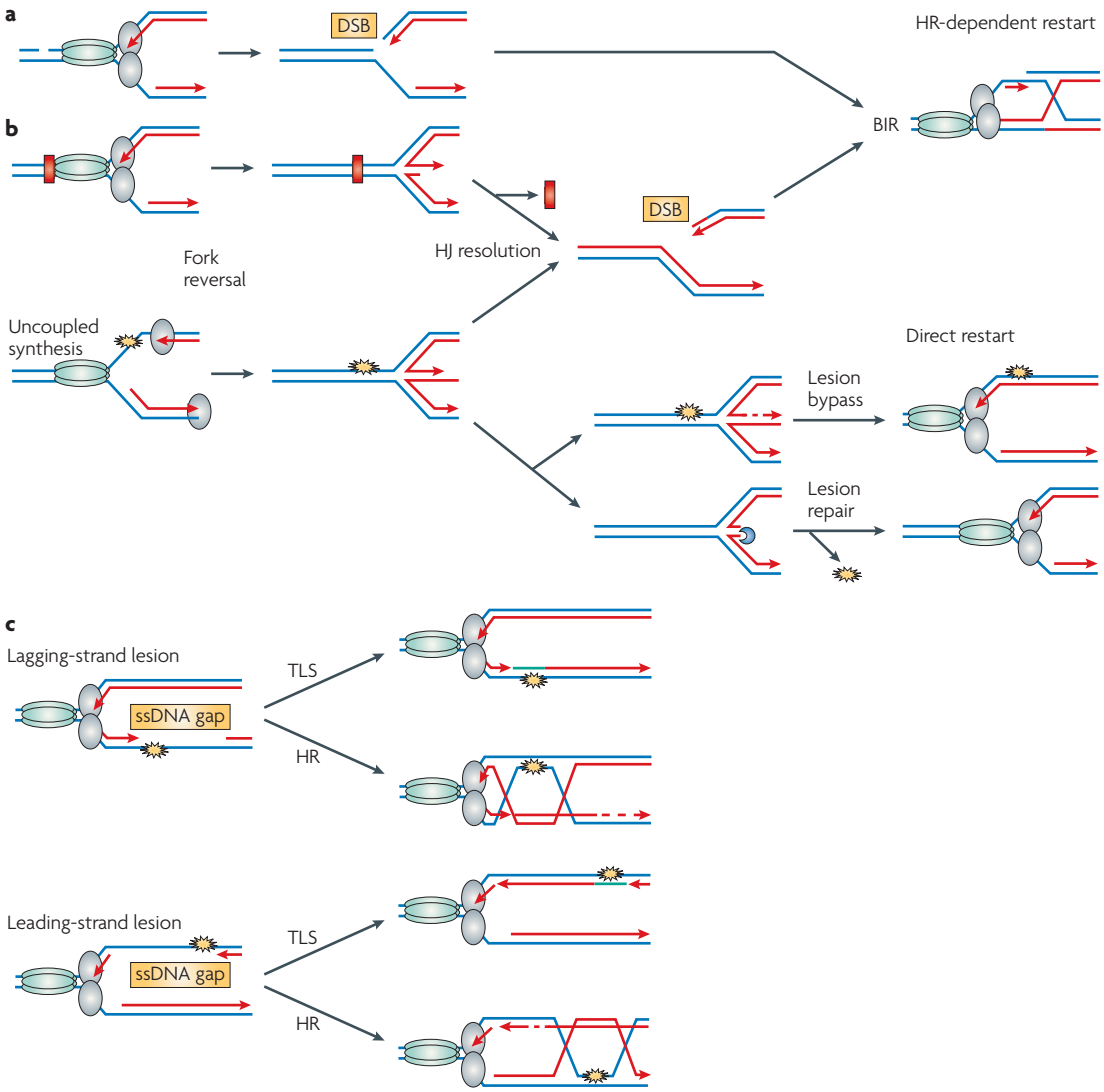

Figure 2 | **Formation of double-stranded breaks and ssDNA gaps during replication. a** | Encountering a ssDNA nick in the leading strand template will directly lead to replication-fork (RF) collapse and a one-ended double-stranded break (DSB) that could promote restart by break-induced replication (BIR). **b** | DNA adducts or tightly bound proteins (indicated by a red rectangle), such as the transcription machinery, can block RF progression or leading-strand synthesis. In the latter case, leading-strand and lagging-strand synthesis would be uncoupled. Subsequently, fork reversal can take place leading to the formation of a Holliday junction (HJ) or a 'chicken-foot' structure. Restart can occur by HJ-cleavage followed by BIR but the RF can also undergo direct restart by HJ reversal to a fork after lesion bypass through template switching or lesion repair. **c** | Lesions (indicated by a star) can block the synthesis of only one DNA strand without impeding fork progression. A lesion that blocks lagging-strand synthesis will not cause RF stalling but will create a ssDNA gap between two flanking Okazaki fragments (top), whereas a lesion that blocks leading-strand synthesis could be bypassed by the RF, which will restart downstream of the lesion leaving a ssDNA gap behind (bottom). In both cases, ssDNA gaps can be repaired by error-prone translesion synthesis (TLS) or by error-free homologous recombination (HR) using the sister chromatid as a template (template switching). Whether template switching requires conversion of the ssDNA gap to a DSB before repair remains unclear.

of breaks[24,25,31,36–39]. GCRs are also strongly increased by mutations that affect S-phase checkpoint proteins such as Rfc5, Rad24 or Mec1, which contribute to the stabilization and restarting of stalled RFs[6,7,40] (TABLE 1). Thus, it seems that GCRs might be associated with breaks generated by stalling and/or collapse of RFs. This conclusion has now been extended to mammals based on the findings that frequent chromosome breaks are seen in mouse cell lines with defective alleles of *MCM4* or *RPA1* (REFS 41,42). The identification of S-phase checkpoint mutants, such as *dbp11* (TABLE 1), that increase GCRs but not HR supports the notion that S-phase checkpoints promote restart of collapsed RFs by HR. Checkpoint inactivation could lead to non-homologous mechanisms of repair such as non-homologous end joining (NHEJ) or breakage–bridge fusion, causing GCRs.

Minisatellite
A class of repetitive sequences, 7–100 nucleotides each, that span 0.5–20 kb and are located throughout the genome, especially towards chromosome ends.

The fact that replication stress increases both HR and GCRs indicates that both types of event might be induced by similar intermediates (FIG. 3). Therefore, it is likely that in addition to the checkpoint, the ability of the different repair mechanisms to process replication-dependent DNA breaks can influence the outcome and rate of GIN. Indeed, abolition of HR by *rad52Δ* in yeast strongly increases GCRs, implying a role for HR in the suppression of GCR. However, in other HR repair genes such as *RAD51,* or in NHEJ genes such as *YKU70*, *YKU80* and *DNL4*, mutations only slightly affect GCR, opening up the possibility that BIR (FIG. 3) or a non-standard NHEJ pathway could also play a part in GCR suppression[25,36].

Nevertheless, not all GCRs arise as a result of breaks that are associated with replication defects. For example, some translocations are generated by telomere fusions, presumably followed by breakage–bridge fusion cycles[36,40]. GCRs can also result from a channelling of repair events from standard DSB-repair pathways to other type of events, such as *de novo* telomere addition, which is normally suppressed by the Pif1 DNA–RNA helicase that inhibits telomerase[36,43], or chromosome rearrangements that occur when equal SCE promoted by cohesion factors fails[44,45].

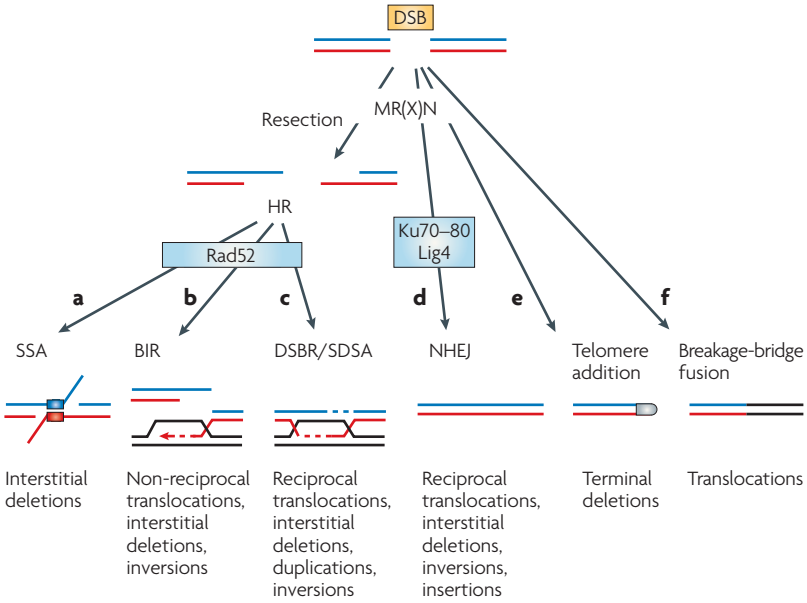

Figure 3 | **Mechanisms of double-stranded break repair leading to different chromosome rearrangements. a** | A double-stranded break (DSB) flanked by two homologous DNA repeats can be repaired by single-strand annealing (SSA), which depends on Rad52 but not on the strand-exchange functions of Rad51 or Rad54. **b** | One-ended DSBs can be repaired by break-induced replication (BIR). This pathway normally uses Rad51-mediated strand exchange, but can also occur at low efficiency in the absence of Rad51. **c** | Standard homologous recombination (HR) events that occur by either the standard DSB-repair (DSBR) model, requiring resolution of two Holliday junction structures, or the synthesis-dependent strand-annealing (SDSA) pathway. SDSA will only cause rearrangements by unequal gene conversion between non-allelic regions (not shown). **d** | In the absence of sequence homology, the ends of two different DSBs can be joined together by non-homologous end joining (NHEJ). **e** | A broken chromosome can be stabilized by telomere addition. **f** | Breakage–bridge fusion can generate translocations and other types of rearrangements. Ku70–Ku80, a NHEJ protein complex; Lig4, DNA ligase 4; MR(X)N, a nuclease complex.

Fragile sites

Fragile sites are DNA sequences that show gaps and breaks following partial inhibition of DNA synthesis[46]. They are frequently associated with hotspots for translocations, gene amplifications, integration of exogenous DNA and other rearrangements[47]. There are two types of fragile sites, common and rare. Common fragile sites account for more than 95% of all known fragile sites and are naturally present in mammalian genomes. Rare fragile sites, also known as dynamic mutations, account for less than 5% of the cases and arise as a consequence of DNA-repeat expansion. Rare fragile sites are associated with genetic diseases, such as fragile X mental retardation syndrome, Friedrich's ataxia, Huntington disease, myotonic dystrophy or several spinocerebellar ataxias, which are caused by the effect of the DNA expansion at the RNA or protein level of the locus affected (reviewed in REF. 47). Although most of the rare fragile sites are destabilized by altering DNA metabolism under folate stress, common fragile sites are destabilized if exposed to replication stress by aphidicolin, an inhibitor of eukaryotic replicative DNA polymerases α and ε. They are composed of AT-rich sequences and are found at the breakpoint of chromosome rearrangements that can be seen in tumours[48].

DNA secondary structures at DNA repeats. Fragile sites are usually associated with trinucleotide repeats (TNRs) of the type CGG–CCG, CAG–CTG, GAA–TTC and GCN–NGC, with specific G-rich tetra- to dodecanucleotide repeats or with long AT-rich repeats, such as the 33 or 42 minisatellites of the FRA16B and FRA10B common fragile sites[49]. Fragile sites are conserved in mammals[50] and are also present in yeast[51–53]. This, together with the observation that mammalian fragile-site sequences are unstable in yeast and bacteria[54–57], suggests that fragile sites are ubiquitous and that the molecular basis of their instability is inherent in their DNA structure and might be related to DNA-repeat instability (FIG. 5). Thus, instability of TNRs and AT-rich minisatellites seems to be associated with their capability to adopt unusual secondary structures, such as hairpins or DNA triplexes[58–61]. This feature is common to different types of repeated DNA. Accordingly, repeat instability is dependent on MMR in mice and yeast, consistent with the observation that sequences at repetitive DNA sites form short hairpins that are targeted by the Msh2–Msh6 MMR system[61,62]. Furthermore, palindromic sequences are associated with genetic instability in a SbcCD-nuclease-dependent and MRX-nuclease-dependent manner in *E. coli* and yeast, respectively[62,63]. sbcCD and MRX can cleave hairpins that are formed during lagging-strand synthesis, yielding a DSB[64,65] (FIG. 5).

The observation that expansions and contractions depend on repeat orientation relative to replication origins[54,66–68] suggests that repeat instability is associated with replication perturbation of the lagging strand. It is possible that hairpins, stem-loops or triplexes that are preferentially formed on the lagging strand during its transient single-stranded conformation would perturb

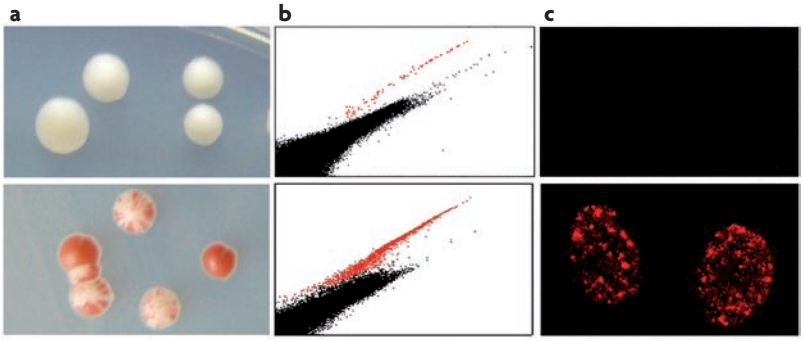

Figure 4 | **Detection of genome instability and DNA breaks. a** | Detection of genome instability (GIN) caused by replicative stress using a colour-sectoring assay. GIN is revealed by hyper-recombination between direct repeats in *Saccharomyces cerevisiae* that are carrying two copies of a long DNA fragment flanking an *ADE2* marker between them. A high level of deletion of the intervening *ADE2* marker are observed as a red sectoring of the colonies in a DNA polymerase δ mutant, whereas deletions are below detection levels in wild-type cells. **b** | Detection of activation-induced cytidine deaminase (AID)-mediated transcription-associated recombination by fluorescence-activated cell sorting (FACS) analysis. Homologous recombination between two truncated forms of the GFP gene re-establishes a wild-type GFP copy encoding active GFP. The high levels of recombinants in yeast cells in which messenger ribonucleoprotein biogenesis is defective and which express human AID (bottom), as compared with cells not expressing AID (top), can be directly visualized by FACS analysis. The recombinant population emitting green fluorescence is shown in red, the non-recombinant in black. **c** | Detection of double-stranded break (DSB) accumulation by phosphorylated histone H2AX (γH2AX) foci. HeLa cells treated with neocarzinostatin (a genotoxic anticarcinogenic agent that blocks proliferation) accumulate DSBs that are signalled by phosphorylation of H2AX. High levels of DSBs can be inferred from the high level of anti-γH2AX foci observed in treated cells (bottom panel) versus non-treated cells (top panel). Part **c** courtesy of M. Domínguez-Sánchez, University of Seville, Spain.

DNA synthesis thereby causing slippage or blockage, which would be enhanced under replication stress. This is supported by the high instability of repetitive DNA in yeast mutants that are defective in replication factors such as DNA polymerase α and δ, DNA ligase I or PCNA[56,69] and in mutants that are defective in S-phase checkpoint functions such as Mec1, Rad53, Rad17 or Rad24 (REF. 70). Hairpin structures formed at the 5′ end of a displaced Okazaki fragment could promote tract expansions. This is supported by the fact that yeast mutants lacking the Rad27 Flap endonuclease show length-dependent destabilization of CAG–CTG tracts and a substantially higher expansion frequency[55] (FIG. 5). The fact that secondary structures formed at repetitive DNA regions perturb replication is further supported by the occurrence of expansions in yeast TNRs that are capable of forming hairpins and the lack of expansions in yeast TNRs that cannot[71].

Replication impairment at fragile sites. As is the case with repeat instability, perturbation of replication leading to DNA breaks or gaps is crucial for fragile-site instability[72]. This instability is detected in cultured cells under replication inhibition; additionally, the stability of mammalian fragile sites depends on the Mec1 and other S-phase checkpoints[73]. Notably, this is also the case in yeast, in which fragile sites are seen in mutants in which the replication machinery or the Mec1 S-phase checkpoint is affected directly[51–53].

Reduced expression of the catalytic subunit of DNA polymerase α in yeast stimulates chromosome loss and translocations, with breakpoints associated with a head-to-head pair of Ty elements that might form hairpins[52]. Fragile sites are also seen in DNA regions containing multiple tRNA genes in yeast cells that are treated with the replication inhibitor hydroxyurea or in yeast lacking the Rrm3 helicase, which is required for replication progression through obstacles that are created by protein complexes or specific DNA structures[37,74]. Two-dimensional-gel electrophoresis reveals that replication in *E. coli*[54,67] and yeast[75,76] is stalled at CCG and CAG–CTG repeats found in mammalian fragile sites, supporting the idea that fragile sites are linked to replication impairment through DNA secondary structures that lead to DSBs (FIG. 5). This has been confirmed recently by showing that FRA16D, an AT-rich common fragile site in yeast, stalls replication and accumulates DSBs[57]. Formation of DSBs is supported by the fact that inhibiting DSB repair by downregulating the HR repair protein Rad51 and NHEJ repair factors DNA–PKCs and ligase IV causes the accumulation of DSB-specific phosphorylated histone H2AX (γH2AX) foci at common fragile sites in the presence of aphidicolin[77]. Therefore, fragile sites appear when replication progression is impaired, mainly at a DNA stem-loop or triplex that impedes RF progression, presumably leading to RF stalling and formation of DNA breaks that are responsible for rearrangements (FIG. 5).

Transcription-associated instability

Transcription of a DNA sequence increases its frequency of recombination, a phenomenon referred to as transcription-associated recombination (TAR)[78]. Approximately 30 years ago it was reported that site-specific *recA*-independent recombination of λ DNA in the *E. coli* genome is enhanced by transcription — probably owing to enhanced access of the integrase to its target sequence[79]. Later, the stimulation of HR by transcription was shown in *S. cerevisiae* (both for RNA polymerase (RNAP) I[80] and RNAPII-driven transcription[81]), in *S. pombe*[82], in general transduction in bacteria[83] and in mammals[84]. In addition to stimulating recombination, transcription stimulates mutations, a phenomenon referred to as transcription-associated mutation (TAM). TAM was first demonstrated almost four decades ago in *E. coli*[85,86]. Like TAR, it also occurs in all organisms and has been shown to occur spontaneously in bacteria and yeast[78,87,88]. A comparative analysis of a 1.5-Mb fragment of a human chromosome containing seven genes, with orthologous regions in eight other mammalian species, revealed a mutational strand asymmetry that is consistent with an evolutionary shaping of each genome by transcription that has produced an excess of G and T over A and C on the coding strands in most genes[89]. Despite their products being different, TAR and TAM might be two different outcomes of the same initial intermediate. Below, we mainly focus on TAR as a mechanism of GIN leading to rearrangements, with some references to TAM when appropriate.

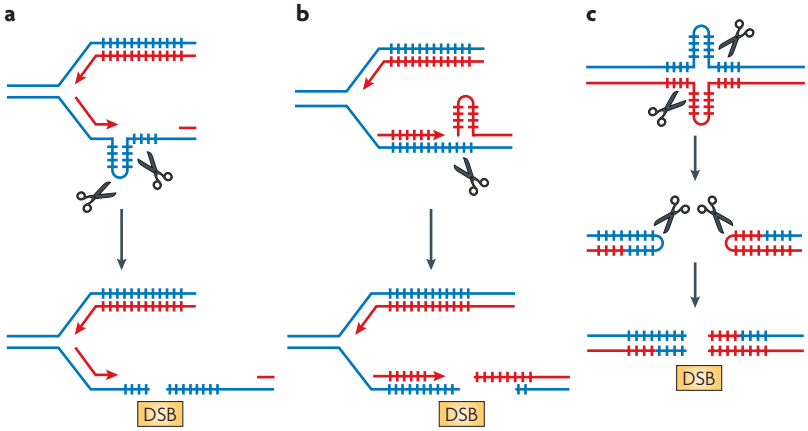

Figure 5 | **Models of double-stranded break formation at fragile sites. a** | Hairpins, stem-loops, triplexes and other secondary DNA structures can be formed at fragile sites, preferentially at the lagging-strand template. Such structures can be substrates for the MR(X)N nuclease complex or other nucleases, the action of which (represented by scissors) creates a nick that can be converted into a double-stranded break (DSB) following replication. For simplicity, only a hairpin is shown as a representative secondary structure. **b** | Expansions that are caused by the formation of secondary structures such as stem-loops, triplexes or hairpin-like structures in the 5′ end of an Okazaki fragment might reduce the ability of the Rad27 Flap endonuclease (FEN1 in humans) to recognize the Flap structure, facilitating breakage of the template strand and the consequent formation of a DSB. For simplicity, only a palindrome is shown as a representative secondary structure. **c** | Formation of stem-loops or hairpins at both the template and daughter strands, preferentially at the lagging strand, can lead to a Holliday junction structure that can be resolved into two hairpin-ended molecules that will be processed by MR(X)N into a DSB.

Transcription-associated recombination. Key to the understanding of TAR and TAM is the fact that ssDNA is chemically more unstable than dsDNA[90]. For example, spontaneous deamination of cytosine to uracil by hydrolysis occurs 140-fold more efficiently on ssDNA than on dsDNA *in vivo*[91]. Transient ssDNA regions are formed during transcription as a consequence of DNA-strand opening, which is caused by the transient accumulation of localized negatively supercoiled DNA behind the advancing RNAP (FIG. 6a). Consistently, mutation rates correlate with the strength of transcription and superhelical stress in bacteria and yeast[38,92]. Genotoxic agents such as 4-nitroquinoline 1-oxide or methylmethanesulphonate increase recombination efficiency more than 100-fold when a gene is transcribed[93].

Several observations in *E. coli* and humans indicate that the non-transcribed strand (NTS) is more susceptible to damage than the transcribed strand (TS)[87,94,95]. Strand opening by negative supercoiling is not sufficient to explain the strand asymmetry of sensitivity to mutations. A more likely explanation is that the NTS but not the TS is single stranded, as is the case in R loops in which the TS forms an RNA–DNA hybrid. Although during transcription a short 8–9 nucleotide RNA–DNA hybrid forms at the transcription bubble inside the catalytic pocket of RNAP, it is unlikely that this short bubble creates the conditions necessary to increase damage in the NTS. Instead, the nascent mRNA extruded from RNAP might hybridize with the TS, creating R loops that would be facilitated by the local negative supercoiling that accumulates behind RNAP (FIG. 6b). Accordingly, R loops form in highly transcribed DNA sequences in *topA* mutants in *E. coli*[96,97], although it remains to be seen whether R loops form in natural conditions. However, R loops are linked to transcription-associated GIN in mutants that are defective in the biogenesis and processing of messenger ribonucleoprotein (mRNP) particles in yeast, chicken and human cells[98,99].

Transcription is coupled to numerous mRNA metabolic processes, including 5′- and 3′-end mRNA processing and mRNP assembly in eukaryotes[100,101], or to translation in prokaryotes. In eukaryotes, as the nascent mRNA is extruded from RNAPII it is coated by proteins, some of which contribute to the assembly of an export-competent mRNP. In *S. cerevisiae*, mutations in genes with a role at the interface between transcription and RNA export or RNA processing increase TAR[102–105], these mutations include the THO–TREX complex, the Sub2–Yra1 heterodimer, Thp1–Sac3 complexes, the nuclear exosome or 3′-end RNA processing. It is not known whether the molecular intermediate triggering TAR is the same in all cases, but the nascent RNA molecule might play a part. The most-studied case is that of THO-complex mutants. There is evidence that R loops are formed in these mutants, implying that an improper formation of the mRNP molecule allows the nascent RNA to hybridize with its DNA template[98]. In THO mutants, R loops are formed in the same high-GC and promoter-distant regions in which TAR is increased[98]. Consistently, the human activation-induced cytidine deaminase (AID) preferentially mutates the NTS of these regions[106] (FIG. 4b).

R loops also form in chicken DT40 cells and human HeLa cells depleted of the ASF (also known as SF2) splicing factor[99], which opens the possibility that ASF also has a role in mRNP formation. In this case, R loops are linked to GIN and are associated with DSB formation[99]. As will be discussed below, class-switch recombination (CSR) of Ig genes, a developmentally regulated TAR process, also involves R-loop formation[107]. It can therefore be concluded that R loops are a widespread structure found in organisms from bacteria to mammals with an important role in TAR.

Transient ssDNA accumulation might also be a major substrate for recombinogenic DNA adducts in highly transcribed genes, although by itself this might not be sufficient to explain TAR, as such adducts also seem to be intimately associated with replication impairment. Consistent with ssDNA having a potential to form secondary DNA structures at particular repeat segments — such as stem-loops that can interfere with DNA replication — simple repetitive poly-GT tracts are destabilized in a transcription-dependent manner by errors that are introduced by the replicative DNA polymerase[108]. In addition, both transcription and replication are halted in *E. coli* when the poly-C strand of a poly-GC tract is on the TS[109]. This is reminiscent of the probability of such tracts forming ssDNA, and hence secondary structures that obstruct the RF, as a consequence of transcription-dependent negative supercoiling. In this sense, it is particularly intriguing that, when transcribed, Friedrich's ataxia GAA triplets form R loops in bacteria and *in vitro*[110].

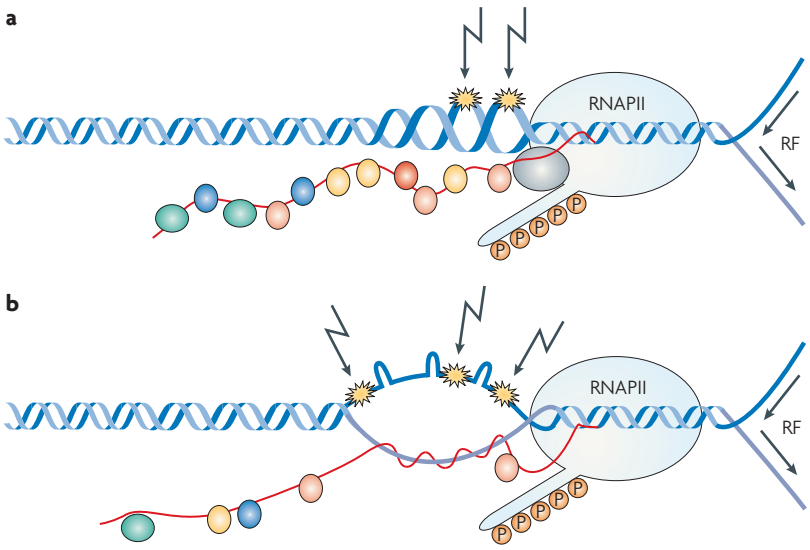

Figure 6 | **DNA intermediates that are potentially associated with transcription-associated recombination. a** | Local negative supercoiling during transcription elongation might facilitate transient ssDNA formation (represented by stars) behind an elongating RNA polymerase (RNAP) that would be susceptible to damage and could therefore compromise replication fork (RF) progression. **b** | When co-transcriptional formation of an optimal mRNA–particle complex is impaired, the RNA can hybridize with its template DNA strand forming a transient R loop, which is distinct from the short 8–9 nucleotide-long DNA–RNA hybrid formed inside the active pocket of the RNAP. The single-stranded non-translated strand region can either be more susceptible to damage or to the formation of DNA secondary structures that can compromise RF progression. This can occur in THO-complex mutants in yeast and in ASF (also known as SF2)-depleted DT40 and HeLa cells, and also at natural loci such as in the switch (S) regions of immunoglobulin (Ig) genes. It is possible that under natural conditions, specific mRNA segments have an intrinsically reduced ability to assemble messenger ribonucleoproteins (mRNPs) properly, leading to a localized suboptimal mRNP conformation that could facilitate R-loop formation. This could be the case for the G-rich mRNA fragment of the S region of Ig genes, but also for regions that contain specific repeats or sequences susceptible to forming secondary structures.

Although RF impairment can be caused by DNA adducts that are formed during transcription, several reports suggest that this might not be the only cause of RF progression impairment. In *S. cerevisiae* TAR is detected in S phase but not in G1-transcribed constructs, suggesting that when transcription is convergent to replication, a collision between RNAPII and the RF can trigger TAR[34]. Similarly, head-on collisions of transcription machinery with RFs that are initiated at the unidirectional ColE1 replication origin inserted in the *E. coli* chromosome upstream and downstream of the *rrnB* operon cause RNAPs to dislodge — this is not the case for co-directional transcription and replication[111]. Although topological constraints between converging RNA and DNA polymerases could contribute to fork stalling, RF blocks are seen in *E. coli* and *S. cerevisiae* inside the transcribed region regardless of the distance between transcription and replication initiation sites[34,112]. RF stalling and collisions are seen in RNAPI-transcribed genes at the rDNA region[35], in genes that are highly transcribed by RNAPII in both wild type and THO mutants[34,113], and in RNAPIII-transcribed tRNA genes[114]. Stalling is stronger in the absence of the Rrm3 helicase, indicating that the RF must overcome non-nucleosomal obstacles in transcribed regions[34,37]. The observation that the Rrm3 helicase is found at highly transcribed genes in the yeast genome[115] supports the view that transcription can be an obstacle to replication, and that Rrm3 would be more often required during replication of such highly transcribed genes.

The ability of RNAP to interfere with RF progression has also been suggested in bacteria. The ability of *E. coli* cells to survive UV damage correlates with their capacity to synthesize ppGpp, the stringent response regulator that helps to destabilize RNAP open complexes in the absence of the HJ resolvase RuvABC. This is explained by a reduction in the incidence of stalled RNAP complexes at damaged sites, facilitating nucleotide excision repair and allowing RF progression[116]. Consistently, lack of the GreA elongation factor is required for transcription to resume, whereas lack of the Mfd transcription-coupled repair factor, which helps remove RNAP stalled at damaged sites, reduces cell viability when repair is compromised[117].

Therefore, it seems that TAR is linked to the ability of RNAP transcription to interfere with RF progression, either by physically obstructing the RF or by promoting lesions that block DNA synthesis, whether or not it is mediated by R loops. This is probably the reason why, in *S. cerevisiae,* specific protein–DNA complexes act as RF-block sites, preventing replication from proceeding through highly transcribed rRNA genes, thereby avoiding RF–RNAPI collisions[118].

AID-mediated class switching and translocations. Following V(D)J recombination, catalysed by the site-specific RAG1 and RAG2 recombinases, Ig genes can undergo a second wave of mutations and rearrangements in B cells through somatic hyper-mutation (SHM) and CSR in vertebrates, and hyper-gene conversion in birds. Such events are triggered by the B-cell-specific enzyme AID, which deaminates dCs into dUs in actively transcribed DNA (reviewed in REF. 119), representing unique and specialized cases of transcription-associated GIN. It is probably the best example of a phenomenon that in principle is harmful but that has acquired a key developmentally controlled role in the generation of genetic diversity.

Although it is not clear how transcription of switch (S) and variable (V) regions of Ig genes acts as a signal to target AID action, there is evidence that S regions, consisting of highly G-rich sequences, form R loops[107,120,121] that facilitate the ssDNA configuration of NTS, which, as shown *in vitro*, would be stabilized in the form of G loops[122]. As ssDNA is the natural substrate of AID[123,124], the NTS of S regions would be its preferential target (FIG. 6b). However, some data indicate that AID can also access the TS[125]. The observation that AID co-immunoprecipitates with RNAPII in activated murine splenic cells[99,126] or interacts *in vitro* with the RNAPII elongation complex[127], opens up the possibility that the transcriptional machinery could directly recruit AID. This possibility could explain the specific pattern of SHM found in V segments, the mutation

G loops
Structures that are formed during transcription that contain a stable mRNA–DNA hybrid on the transcribed strand and a G-quartet DNA on the G-rich non-transcribed strand.

frequency of which peaks at nucleotide 200 of the promoter and disappears after 1.5 kb as AID cannot access the DNA[128]. Because the ssDNA-binding protein RPA stimulates AID action *in vitro*, it is also possible that RPA plays a part in AID targeting[129], although the binding of RPA to transiently formed ssDNA regions or R loops during transcription remains to be demonstrated. Indeed, if this were the case, it would be necessary to explain why AID does not act on any region undergoing replication, in which RPA binds to the transiently formed single-stranded lagging strand. Another possibility would be that a natural but putatively suboptimal mRNP packaging of the G-rich S region mRNA segment could facilitate the segment's hybridization with its DNA template to form an R loop[106] (FIG. 6b). Whatever the mechanism of AID targeting, AID-dependent rearrangements provide a unique system to understand the causes and mechanisms of GIN.

The mechanism by which AID-promoted dUs lead to SHM is beginning to be clarified[119], but little is known about the mechanism by which dUs ensure CSR. dUs can be targeted by the BER machinery: dUs are recognized by the uracil-glycosylase UNG, which creates an abasic site that, by the subsequent action of apurinic–apirimidinic (AP) or MRN endonucleases, will give rise to ssDNA breaks, which can become DSBs during replication (FIG. 5). Alternatively, the Msh2–Msh6 dimer could recognize the dG–dU mismatch leading to the removal of the dU-containing strand, resulting in either mutations or DNA nicks. CSR seems to occur by NHEJ[130,131] and AID activation leads to γH2AX foci[132], implying that DSBs are a pre-requisite for CSR. Nicks could be converted into DSBs by the action of endonucleases or topoisomerases, or by replication, although the recent observation that DSBs are detected in G1 rather than S–G2 phases[133] suggests that replication is unlikely to be required for CSR. The most widely accepted view is that staggered nicks that are induced in both strands by BER and MMR could constitute a DSB.

Despite the role of AID in the formation of DNA breaks, steps that follow deamination might share common intermediates and mechanisms with GIN that are not mediated by AID. This is of great interest because a large proportion of adenomas and lymphomas of B and T cells are caused by translocations between Ig segments and proto-oncogenes, and Ig diversification is mediated by breaks that share the potential to induce oncogenic translocations. For example, cleavage initiated by RAG proteins is linked to translocations found in certain types of tumours[134]. However, most human lymphomas do not express RAG proteins but are associated with Ig translocations. One of the best-studied examples is Burkitt's lymphoma, which involves translocations between S regions and *c-Myc*. These translocations occur in mice by an AID-mediated mechanism, with the breakpoints mapping to a G-rich segment of *c-Myc* that shares structural properties of the S regions[135,136]. Other non-Ig genes, such as *BCL6, FAS, BCL2, BCR, ABL, TEL1* or *AML1* undergo hyper-mutation or rearrangements in B cells from healthy individuals, which suggests that their ORFs could also be targets of AID (reviewed in REF. 137).

AID-dependent translocations are detected in murine KU80-depleted B cells, indicating that these translocations might not occur by a standard NHEJ pathway[138]. Indeed, CSR and translocations occur in XRCC4- or LIG4-deficient mouse B cells, consistent with a role of alternative end-joining in both types of rearrangements[139]. However, whether other mechanisms (such as BIR) generate translocations has not been sufficiently investigated. Whatever mechanisms are responsible for AID-dependent translocations, they seem to be triggered by the same substrates that trigger CSR; they should therefore help us to obtain an integrated view of GIN mechanisms.

Conclusions and perspectives

The identification and analysis of an increasing number of genetic disorders associated with GIN indicates that the key elements in the origin of instability are the suppressor genes, which function *in trans*, and the chromosomal sites such as fragile sites or highly-transcribed DNA sequences, which act *in cis* as hotspots of GIN. The identification of the molecular function of suppressors, the primary DNA sequence of fragile sites and a better understanding of TAR has allowed us to draw a common scenario for the origin of GIN. A replication obstacle might impede DNA synthesis of one strand or might lead to RF stalling and/or collapse resulting in DNA breaks or ssDNA gaps, the S-phase checkpoint being required for prevention of RF collapse and for RF restart. The growing list of cellular functions that prevent GIN suggests that, ultimately, rearrangements can primarily be the result of replication impairment. This impairment can be caused locally by particular structures, as in fragile sites or highly transcribed DNA, or by malfunction of the replication and S-phase checkpoint machinery.

Which DNA-break repair pathway is used in each case determines which GIN event takes place in different genetic disorders. For example, if HR or SCE are impaired, the predominant GIN event is GCR (TABLE 1). However, why replication-generated breaks in some cases induce HR and in other cases induce GCRs is not clear. Many genetic disorders involve both HR and GCRs, suggesting that they both derive from a similar initiation event. In order to know the causes of GIN, we need to identify the mechanism by which specific DNA breaks or ssDNA gaps arise. Thus, the GIN outcome that is initiated by a DSB could be different depending on whether it occurs in the lagging or the leading strand, or as a consequence of RF collapse or reversal (FIG. 2). In yeast, DSBs that are generated by the RF encountering a ssDNA nick can be repaired by SCE using HR functions[11]. However, studies using *C. elegans* RFS-1 (a Rad51 paralogue) mutants suggest that stalled RFs do not cause DSBs like those made with HO and I-*Sce*I endonucleases. Instead, RFS-1 has a specialized role in the loading of RAD-51 onto the breaks that presumably form at stalled RFs[140]. It is possible that even though HR repairs all types of DSBs, they might require different initial processing steps depending on how they are generated (by replication or enzymatically). In addition, we need to know how mutations in genes that encode specific replication

cofactors or S-phase checkpoints determine the nature of the DNA break or gap generated and how they contribute to the choice of the DSB repair pathway.

Although many lines of evidence indicate that replication impairment is the main cause of GIN, many specific questions remain to be answered: does transcription contribute to fragile-site activation? Do all types of fragile sites affect RF progression in a similar manner? Does the S-phase checkpoint respond differently to ssDNA gaps versus DSBs generated by RF stalling or collapse? How does an RF bypass different types of physical barriers? How do cells decide whether to use HR or NHEJ to repair a DSB? Are the breakpoints of rearrangements caused by replication inhibitors the same as those caused by checkpoint inactivation? How does checkpoint inactivation affect TAR? Is an R loop involved in all TAR events? Is chromatin structure a modulator of GIN? A combination of biochemistry, genetics, and molecular and cell biology is required to answer these questions, but the structural and functional genomics developed in different eukaryotic organisms will strongly contribute to our understanding of the different mechanisms and factors governing GIN.

1. Maizels, N. Immunoglobulin gene diversification. *Annu. Rev. Genet.* **39**, 23–46 (2005).
2. McMurray, M. A. & Gottschling, D. E. An age-induced switch to a hyper-recombinational state. *Science* **301**, 1908–1911 (2003).
This paper describes a strong increase in genomic instability that is linked to ageing in the budding yeast, *S. cerevisiae*. This constitutes a rational explanation for the association of high cancer risk with age in mammals.
3. Draviam, V. M., Xie, S. & Sorger, P. K. Chromosome segregation and genomic stability. *Curr. Opin. Genet. Dev.* **14**, 120–125 (2004).
4. Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–913 (2005).
This paper shows that the DNA-damage response is activated in a panel of human lung hyperplasias, whereas its inactivation leads to malignant transformation, pointing to a role of checkpoints as a natural anticancer barrier.
5. Friedberg, E. C. *et al. DNA Repair and Mutagenesis* 461–750 (American Society For Microbiology Press, Washington DC, 2006).
6. Myung, K., Datta, A. & Kolodner, R. D. Suppression of spontaneous chromosomal rearrangements by S phase checkpoint functions in *Saccharomyces cerevisiae*. *Cell* **104**, 397–408 (2001).
7. Myung, K. & Kolodner, R. D. Suppression of genome instability by redundant S-phase checkpoint pathways in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **99**, 4500–4507 (2002).
8. Rudolph, C. J., Upton, A. L. & Lloyd, R. G. Replication fork stalling and cell cycle arrest in UV-irradiated *Escherichia coli*. *Genes Dev.* **21**, 668–681 (2007).
9. Cobb, J. A. *et al.* Replisome instability, fork collapse, and gross chromosomal rearrangements arise synergistically from Mec1 kinase and RecQ helicase mutations. *Genes Dev.* **19**, 3055–3069 (2005).
10. Sogo, J. M., Lopes, M. & Foiani, M. Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science* **297**, 599–602 (2002).
This paper presents electron microscopy visualization of RF intermediates in *S. cerevisiae rad53* checkpoint mutants under replication stress induced by hydroxyurea, revealing long ssDNA accumulation and hemi-replicated molecules. HJ structures are observed as a consequence of fork reversal.
11. Cortes-Ledesma, F. & Aguilera, A. Double-strand breaks arising by replication through a nick are repaired by cohesin-dependent sister-chromatid exchange. *EMBO Rep.* **7**, 919–926 (2006).
12. Pages, V. & Fuchs, R. P. Uncoupling of leading- and lagging-strand DNA replication during lesion bypass *in vivo*. *Science* **300**, 1300–1303 (2003).
13. Lopes, M., Foiani, M. & Sogo, J. M. Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol. Cell* **21**, 15–27 (2006).
14. Cordeiro-Stone, M., Makhov, A. M., Zaritskaya, L. S. & Griffith, J. D. Analysis of DNA replication forks encountering a pyrimidine dimer in the template to the leading strand. *J. Mol. Biol.* **289**, 1207–1218 (1999).
15. Higgins, N. P., Kato, K. & Strauss, B. A model for replication repair in mammalian cells. *J. Mol. Biol.* **101**, 417–425 (1976).
16. Seigneur, M., Bidnenko, V., Ehrlich, S. D. & Michel, B. RuvAB acts at arrested replication forks. *Cell* **95**, 419–430 (1998).
In this paper, the presence of linear DNA as evidence of DSBs in *E. coli* together with genetic data are used as a demonstration of the occurrence of fork reversal to a HJ structure in arrested RFs.
17. Heller, R. C. & Marians, K. J. Replisome assembly and the direct restart of stalled replication forks. *Nature Rev. Mol. Cell Biol.* **7**, 932–943 (2006).
18. Heller, R. C. & Marians, K. J. Replication fork reactivation downstream of a blocked nascent leading strand. *Nature* **439**, 557–562 (2006).
19. Berdichevsky, A., Izhar, L. & Livneh, Z. Error-free recombinational repair predominates over mutagenic translesion replication in *E. coli*. *Mol. Cell* **10**, 917–924 (2002).
20. Fabre, F., Chan, A., Heyer, W. D. & Gangloff, S. Alternate pathways involving Sgs1/Top3, Mus81/Mms4 and Srs2 prevent formation of toxic recombination intermediates from single-stranded gaps created by DNA replication. *Proc. Natl Acad. Sci. USA* **99**, 16887–16892 (2002).
21. Klein, H. L. & Petes, T. D. Intrachromosomal gene conversion in yeast. *Nature* **289**, 144–148 (1981).
22. Jackson, J. A. & Fink, G. R. Gene conversion between duplicated genetic elements in yeast. *Nature* **292**, 306–311 (1981).
23. Aguilera, A. & Klein, H. L. Genetic control of intrachromosomal recombination in *Saccharomyces cerevisiae*. I. Isolation and genetic characterization of hyper-recombination mutations. *Genetics* **119**, 779–790 (1988).
24. Chen, C., Umezu, K. & Kolodner, R. D. Chromosomal rearrangements occur in *S. cerevisiae rfa1* mutator mutants due to mutagenic lesions processed by double-strand-break repair. *Mol. Cell* **2**, 9–22 (1998).
25. Chen, C. & Kolodner, R. D. Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nature Genet.* **23**, 81–85 (1999).
26. Marinus, M. G. & Konrad, E. B. Hyper-recombination in *dam* mutants of *Escherichia coli* K-12. *Mol. Gen. Genet.* **149**, 273–277 (1976).
27. Hartwell, L. H. & Smith, D. Altered fidelity of mitotic chromosome transmission in cell cycle mutants of *S. cerevisiae*. *Genetics* **110**, 381–395 (1985).
28. Zieg, J., Maples, V. F. & Kushner, S. R. Recombinant levels of *Escherichia coli* K-12 mutants deficient in various replication, recombination or repair genes. *J. Bacteriol.* **134**, 958–966 (1978).
29. Tishkoff, D. X., Filosi, N., Gaida, G. M. & Kolodner, R. D. A novel mutation avoidance mechanism dependent on *S. cerevisiae* Rad27 is distinct from DNA mismatch repair. *Cell* **88**, 253–263 (1997).
30. Prado, F., Cortes-Ledesma, F. & Aguilera, A. The absence of the yeast chromatin assembly factor Asf1 increases genomic instability and sister chromatid exchange. *EMBO Rep.* **5**, 497–502 (2004).
31. Myung, K., Pennaneach, V., Kats, E. S. & Kolodner, R. D. *Saccharomyces cerevisiae* chromatin-assembly factors that act during DNA replication function in the maintenance of genome stability. *Proc. Natl Acad. Sci. USA* **100**, 6640–6645 (2003).
32. Gangloff, S., McDonald, J. P., Bendixen, C., Arthur, L. & Rothstein, R. The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. *Mol. Cell Biol.* **14**, 8391–8398 (1994).
33. Lambert, S., Watson, A., Sheedy, D. M., Martin, B. & Carr, A. M. Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell* **121**, 689–702 (2005).
34. Prado, F. & Aguilera, A. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J.* **24**, 1267–1276 (2005).
35. Takeuchi, Y., Horiuchi, T. & Kobayashi, T. Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes Dev.* **17**, 1497–1506 (2003).
36. Myung, K., Chen, C. & Kolodner, R. D. Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*. *Nature* **411**, 1073–1076 (2001).
In this paper, genetic analyses serve to define three groups of suppressors of the formation of GCRs in *S. cerevisiae*: S-phase checkpoints, recombination factors and proteins that are involved in *de novo* telomere addition.
37. Ivessa, A. S. *et al.* The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Mol. Cell* **12**, 1525–1536 (2003).
38. Schmidt, K. H. & Kolodner, R. D. Suppression of spontaneous genome rearrangements in yeast DNA helicase mutants. *Proc. Natl Acad. Sci. USA* **103**, 18196–18201 (2006).
39. Lengronne, A. & Schwob, E. The yeast CDK inhibitor Sic1 prevents genomic instability by promoting replication origin licensing in late G1. *Mol. Cell* **9**, 1067–1078 (2002).
40. Mieczkowski, P. A., Mieczkowska, J. O., Dominska, M. & Petes, T. D. Genetic regulation of telomere–telomere fusions in the yeast *Saccharomyces cerevisae*. *Proc. Natl Acad. Sci. USA* **100**, 10854–10859 (2003).
41. Wang, Y. *et al.* Mutation in *Rpa1* results in defective DNA double-strand break repair, chromosomal instability and cancer in mice. *Nature Genet.* **37**, 750–755 (2005).
42. Shima, N. *et al.* A viable allele of *Mcm4* causes chromosome instability and mammary adenocarcinomas in mice. *Nature Genet.* **39**, 93–98 (2007).
Together with reference 41, this paper demonstrates the existence of GIN in replication mutants (*Rpa1* and *Mcm4*) of mice, indicating that replication impairment is a source of GIN in mammals.
43. Schulz, V. P. & Zakian, V. A. The *Saccharomyces* PIF1 DNA helicase inhibits telomere elongation and *de novo* telomere formation. *Cell* **76**, 145–155 (1994).
44. Kobayashi, T. & Ganley, A. R. Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. *Science* **309**, 1581–1584 (2005).
45. De Piccoli, G. *et al.* Smc5–Smc6 mediate DNA double-strand-break repair by promoting sister-chromatid recombination. *Nature Cell Biol.* **8**, 1032–1034 (2006).
46. Sutherland, G. R. Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium. *Science* **197**, 265–266 (1977).
47. Durkin, S. G. & Glover, T. W. Chromosome fragile sites. *Annu. Rev. Genet.* (2007).
48. Yunis, J. J. & Soreng, A. L. Constitutive fragile sites and cancer. *Science* **226**, 1199–1204 (1984).
49. Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **447**, 932–940 (2007).
50. Shiraishi, T. *et al.* Sequence conservation at human and mouse orthologous common fragile regions, FRA3B/FHIT and Fra14A2/Fhit. *Proc. Natl Acad. Sci. USA* **98**, 5722–5727 (2001).
51. Cha, R. S. & Kleckner, N. ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. *Science* **297**, 602–606 (2002).

52. Lemoine, F. J., Degtyareva, N. P., Lobachev, K. & Petes, T. D. Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell* **120**, 587–598 (2005).
In this paper, the elimination of Mec1 in budding yeast leads to RF collapse in specific regions of the genome, named replication slow zones. Reduced levels of the replicative polymerase α result in translocations and chromosome loss in a region containing two inverted copies of Ty. References 51 and 52 provide evidence of the existence of fragile sites in yeast with similar features to those of mammals.
53. Raveendranathan, M. *et al.* Genome-wide replication profiles of S-phase checkpoint mutants reveal fragile sites in yeast. *EMBO J.* **25**, 3627–3639 (2006).
54. Samadashwily, G. M., Raca, G. & Mirkin, S. M. Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.* **17**, 298–304 (1997).
In this paper, electrophoretic analyses of RF progression in an *E. coli* plasmid that contains trinucleotide repeats show that RF stalling depends on repeat length and orientation with respect to replication origins, suggesting that the formation of unusual DNA structures in the lagging-strand template is responsible for replication blockage and repeat expansions.
55. Freudenreich, C. H., Kantrow, S. M. & Zakian, V. A. Expansion and length-dependent fragility of CTG repeats in yeast. *Science* **279**, 853–856 (1998).
56. Callahan, J. L., Andrews, K. J., Zakian, V. A. & Freudenreich, C. H. Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. *Mol. Cell Biol.* **23**, 7849–7860 (2003).
57. Zhang, H. & Freudenreich, C. H. An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in *S. cerevisiae*. *Mol. Cell* **27**, 367–379 (2007).
In this paper, analysis of the replication intermediates formed in the human common fragile site FRA16D placed into a yeast artificial chromosome reveals that RF stalling and DSBs are responsible for the fragility of these sites.
58. Gacy, A. M., Goellner, G., Juranic, N., Macura, S. & McMurray, C. T. Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell* **81**, 533–540 (1995).
59. Mitas, M., Yu, A., Dill, J. & Haworth, I. S. The trinucleotide repeat sequence d(CGG)$_{15}$ forms a heat-stable hairpin containing Gsyn.Ganti base pairs. *Biochemistry* **34**, 12803–12811 (1995).
60. Wells, R. D. Molecular basis of genetic instability of triplet repeats. *J. Biol. Chem.* **271**, 2875–2878 (1996).
61. Moore, H., Greenwell, P. W., Liu, C. P., Arnheim, N. & Petes, T. D. Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc. Natl Acad. Sci. USA* **96**, 1504–1509 (1999).
62. Manley, K., Shirley, T. L., Flaherty, L. & Messer, A. *Msh2* deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature Genet.* **23**, 471–473 (1999).
63. Richard, G. F., Goellner, G. M., McMurray, C. T. & Haber, J. E. Recombination-induced CAG trinucleotide repeat expansions in yeast involve the Mre11–Rad50–Xrs2 complex. *EMBO J.* **19**, 2381–2390 (2000).
64. Leach, D. R., Okely, E. A. & Pinder, D. J. Repair by recombination of DNA containing a palindromic sequence. *Mol. Microbiol.* **26**, 597–606 (1997).
65. Lobachev, K. S., Gordenin, D. A. & Resnick, M. A. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* **108**, 183–193 (2002).
66. Kang, S., Jaworski, A., Ohshima, K. & Wells, R. D. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nature Genet.* **10**, 213–218 (1995).
67. Freudenreich, C. H., Stavenhagen, J. B. & Zakian, V. A. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell Biol.* **17**, 2090–2098 (1997).
68. Maurer, D. J., O'Callaghan, B. L. & Livingston, D. M. Orientation dependence of trinucleotide CAG repeat instability in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **16**, 6617–6622 (1996).
69. Schweitzer, J. K. & Livingston, D. M. The effect of DNA replication mutations on CAG tract stability in yeast. *Genetics* **152**, 953–963 (1999).
70. Lahiri, M., Gustafson, T. L., Majors, E. R. & Freudenreich, C. H. Expanded CAG repeats activate the DNA damage checkpoint pathway. *Mol. Cell* **15**, 287–293 (2004).
71. Spiro, C. *et al.* Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol. Cell* **4**, 1079–1085 (1999).
72. Glover, T. W., Berger, C., Coyle, J. & Echo, B. DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum. Genet.* **67**, 136–142 (1984).
73. Casper, A. M., Nghiem, P., Arlt, M. F. & Glover, T. W. ATR regulates fragile site stability. *Cell* **111**, 779–789 (2002).
This paper demonstrates the role of the replication checkpoint kinase ATR in the maintenance of fragile site stability.
74. Admire, A. *et al.* Cycles of chromosome instability are associated with a fragile site and are increased by defects in DNA replication and checkpoint controls in yeast. *Genes Dev.* **20**, 159–173 (2006).
75. Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R. & Mirkin, S. M. Replication and expansion of trinucleotide repeats in yeast. *Mol. Cell Biol.* **23**, 1349–1357 (2003).
76. Krasilnikova, M. M. & Mirkin, S. M. Replication stalling at Friedreich's ataxia (GAA)n repeats *in vivo*. *Mol. Cell Biol.* **24**, 2286–2295 (2004).
77. Schwartz, M. *et al.* Homologous recombination and nonhomologous end-joining repair pathways regulate fragile site stability. *Genes Dev.* **19**, 2715–2726 (2005).
78. Aguilera, A. The connection between transcription and genomic instability. *EMBO J.* **21**, 195–201 (2002).
79. Ikeda, H. & Matsumoto, T. Transcription promotes recA-independent recombination mediated by DNA-dependent RNA polymerase in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **76**, 4571–4575 (1979).
80. Voelkel-Meiman, K., Keil, R. L. & Roeder, G. S. Recombination-stimulating sequences in yeast ribosomal DNA correspond to sequences regulating transcription by RNA polymerase I. *Cell* **48**, 1071–1079 (1987).
81. Thomas, B. J. & Rothstein, R. Elevated recombination rates in transcriptionally active DNA. *Cell* **56**, 619–630 (1989).
82. Grimm, C., Schaer, P., Munz, P. & Kohli, J. The strong ADH1 promoter stimulates mitotic and meiotic recombination at the ADE6 gene of *Schizosaccharomyces pombe*. *Mol. Cell Biol.* **11**, 289–298 (1991).
83. Dul, J. L. & Drexler, H. Transcription stimulates recombination. II. Generalized transduction of *Escherichia coli* by phages T1 and T4. *Virology* **162**, 471–477 (1988).
84. Nickoloff, J. A. & Reynolds, R. J. Transcription stimulates homologous recombination in mammalian cells. *Mol. Cell Biol.* **10**, 4837–4845 (1990).
85. Brock, R. D. Differential mutation of the beta-galactosidase gene of *Escherichia coli*. *Mutat. Res.* **11**, 181–186 (1971).
86. Herman, R. K. & Dworkin, N. B. Effect of gene induction on the rate of mutagenesis by ICR-191 in *Escherichia coli*. *J. Bacteriol.* **106**, 543–550 (1971).
87. Beletskii, A. & Bhagwat, A. S. Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **93**, 13919–13924 (1996).
88. Datta, A. & Jinks-Robertson, S. Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* **268**, 1616–1619 (1995).
89. Green, P., Ewing, B., Miller, W., Thomas, P. J. & Green, E. D. Transcription-associated mutational asymmetry in mammalian evolution. *Nature Genet.* **33**, 514–517 (2003).
90. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
91. Frederico, L. A., Kunkel, T. A. & Shaw, B. R. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry* **29**, 2532–2537 (1990).
92. Kim, N., Abdulovic, A. L., Gealy, R., Lippert, M. J. & Jinks-Robertson, S. Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. *DNA Repair (Amst)* **6**, 1285–1296 (2007).
93. Garcia-Rubio, M., Huertas, P., Gonzalez-Barrera, S. & Aguilera, A. Recombinogenic effects of DNA-damaging agents are synergistically increased by transcription in *Saccharomyces cerevisiae*. New insights into transcription-associated recombination. *Genetics* **165**, 457–466 (2003).
94. Fix, D. F. & Glickman, B. W. Asymmetric cytosine deamination revealed by spontaneous mutational specificity in an Ung$^-$ strain of *Escherichia coli*. *Mol. Gen. Genet.* **209**, 78–82 (1987).
95. Skandalis, A., Ford, B. N. & Glickman, B. W. Strand bias in mutation involving 5-methylcytosine deamination in the human *HPRT* gene. *Mutat. Res.* **314**, 21–26 (1994).
96. Drolet, M. *et al.* Overexpression of RNase H partially complements the growth defect of an *Escherichia coli* delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc. Natl Acad. Sci. USA* **92**, 3526–3530 (1995).
97. Masse, E. & Drolet, M. *Escherichia coli* DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. *J. Biol. Chem.* **274**, 16659–16664 (1999).
98. Huertas, P. & Aguilera, A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* **12**, 711–721 (2003).
This paper shows that hyper-recombinant yeast THO mutants that are defective in mRNP biogenesis allow the co-transcriptional formation of R loops at particular DNA sequences.
99. Li, X. & Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365–378 (2005).
100. Aguilera, A. Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. *Curr. Opin. Cell Biol.* **17**, 242–250 (2005).
101. Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nature Rev. Genet.* **8**, 533–543 (2007).
102. Prado, F., Piruat, J. I. & Aguilera, A. Recombination between DNA repeats in yeast hpr1delta cells is linked to transcription elongation. *EMBO J.* **16**, 2826–2835 (1997).
103. Luna, R. *et al.* Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. *Mol. Cell* **18**, 711–722 (2005).
104. Gallardo, M., Luna, R., Erdjument-Bromage, H., Tempst, P. & Aguilera, A. Nab2p and the Thp1p–Sac3p complex functionally interact at the interface between transcription and mRNA metabolism. *J. Biol. Chem.* **278**, 24225–24232 (2003).
105. Fan, H. Y., Merker, R. J. & Klein, H. L. High-copy-number expression of *Sub2p*, a member of the RNA helicase superfamily, suppresses *hpr1*-mediated genomic instability. *Mol. Cell Biol.* **21**, 5459–5470 (2001).
106. Gomez-Gonzalez, B. & Aguilera, A. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc. Natl Acad. Sci. USA* **104**, 8409–8414 (2007).
107. Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E. & Lieber, M. R. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature Immunol.* **4**, 442–451 (2003).
This paper provides a molecular demonstration of the existence of R loops in the Ig S regions *in vivo*, which are important for CSR.
108. Wierdl, M., Greene, C. N., Datta, A., Jinks-Robertson, S. & Petes, T. D. Destabilization of simple repetitive DNA sequences by transcription in yeast. *Genetics* **143**, 713–721 (1996).
109. Krasilnikova, M. M., Samadashwily, G. M., Krasilnikov, A. S. & Mirkin, S. M. Transcription through a simple DNA repeat blocks replication elongation. *EMBO J.* **17**, 5095–5102 (1998).
110. Grabczyk, E., Mancuso, M. & Sammarco, M. C. A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP *in vitro*. *Nucleic Acids Res.* **35**, 5351–5359 (2007).
111. French, S. Consequences of replication fork movement through transcription units *in vivo*. *Science* **258**, 1362–1365 (1992).
112. Mirkin, E. V. & Mirkin, S. M. Mechanisms of transcription-replication collisions in bacteria. *Mol. Cell Biol.* **25**, 888–895 (2005).
113. Wellinger, R. E., Prado, F. & Aguilera, A. Replication fork progression is impaired by transcription in hyper-recombinant yeast cells lacking a functional THO complex. *Mol. Cell Biol.* **26**, 3327–3334 (2006).

114. Deshpande, A. M. & Newlon, C. S. DNA replication fork pause sites dependent on transcription. *Science* **272**, 1030–1033 (1996).
This paper uses two-dimensional-gel electrophoresis to identify RF pause sites in a transcriptionally active tRNA gene of *S. cerevisiae*.
115. Azvolinsky, A., Dunaway, S., Torres, J. Z., Bessler, J. B. & Zakian, V. A. The *S. cerevisiae* Rrm3p DNA helicase moves with the replication fork and affects replication of all yeast chromosomes. *Genes Dev.* **20**, 3104–3116 (2006).
116. McGlynn, P. & Lloyd, R. G. Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression. *Cell* **101**, 35–45 (2000).
117. Trautinger, B. W., Jaktaji, R. P., Rusakova, E. & Lloyd, R. G. RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Mol. Cell* **19**, 247–258 (2005).
118. Brewer, B. J. & Fangman, W. L. A replication fork barrier at the 3′ end of yeast ribosomal RNA genes. *Cell* **55**, 637–643 (1988).
119. Di Noia, J. M. & Neuberger, M. S. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22 (2007).
120. Daniels, G. A. & Lieber, M. R. RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Res.* **23**, 5006–5011 (1995).
121. Reaban, M. E., Lebowitz, J. & Griffin, J. A. Transcription induces the formation of a stable RNA. DNA hybrid in the immunoglobulin alpha switch region. *J. Biol. Chem.* **269**, 21850–21857 (1994).
122. Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F. & Maizels, N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* **18**, 1618–1629 (2004).
123. Bransteitter, R., Pham, P., Scharff, M. D. & Goodman, M. F. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl Acad. Sci. USA* **100**, 4102–4107 (2003).
124. Chaudhuri, J. *et al.* Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **422**, 726–730 (2003).
125. Xue, K., Rada, C. & Neuberger, M. S. The *in vivo* pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in *Msh2*$^{-/-}$ *Ung*$^{-/-}$ mice. *J. Exp. Med.* **203**, 2085–2094 (2006).
126. Nambu, Y. *et al.* Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science* **302**, 2137–2140 (2003).
127. Besmer, E., Market, E. & Papavasiliou, F. N. The transcription elongation complex directs activation-induced cytidine deaminase-mediated DNA deamination. *Mol. Cell Biol.* **26**, 4378–4385 (2006).
128. Longerich, S. & Storb, U. The contested role of uracil DNA glycosylase in immunoglobulin gene diversification. *Trends Genet.* **21**, 253–256 (2005).
129. Chaudhuri, J., Khuong, C. & Alt, F. W. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* **430**, 992–998 (2004).
130. Casellas, R. *et al.* Ku80 is required for immunoglobulin isotype switching. *EMBO J.* **17**, 2404–2411 (1998).
131. Manis, J. P. *et al.* Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. *J. Exp. Med.* **187**, 2081–2089 (1998).
132. Petersen, S. *et al.* AID is required to initiate Nbs1/γ-H2AX focus formation and mutations at sites of class switching. *Nature* **414**, 660–665 (2001).
133. Schrader, C. E., Guikema, J. E., Linehan, E. K., Selsing, E. & Stavnezer, J. Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. *J. Immunol.* **179**, 6064–6071 (2007).
134. Jankovic, M., Nussenzweig, A. & Nussenzweig, M. C. Antigen receptor diversification and chromosome translocations. *Nature Immunol.* **8**, 801–808 (2007).
135. Ramiro, A. R. *et al.* AID is required for c-Myc/IgH chromosome translocations *in vivo*. *Cell* **118**, 431–438 (2004).
This paper links CSR with chromosomal translocations involving Ig genes and the *c-Myc* proto-oncogene, and shows that AID is required for these translocations.
136. Duquette, M. L., Huber, M. D. & Maizels, N. G-rich proto-oncogenes are targeted for genomic instability in B-cell lymphomas. *Cancer Res.* **67**, 2586–2594 (2007).
137. Ramiro, A. *et al.* The role of activation-induced deaminase in antibody diversification and chromosome translocations. *Adv. Immunol.* **94**, 75–107 (2007).
138. Ramiro, A. R. *et al.* Role of genomic instability and p53 in AID-induced c-Myc-IgH translocations. *Nature* **440**, 105–109 (2006).
139. Yan, C. T. *et al.* IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449**, 478–482 (2007).
140. Ward, J. D., Barber, L. J., Petalcorin, M. I., Yanowitz, J. & Boulton, S. J. Replication blocking lesions present a unique substrate for homologous recombination. *EMBO J.* **26**, 3384–3396 (2007).
141. Tercero, J. A. & Diffley, J. F. Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature* **412**, 553–557 (2001).
This paper describes a primary role for the S-phase checkpoint in preventing irreversible RF collapse. It shows that the checkpoint proteins Mec1 and Rad53 are required for the completion of DNA replication in the presence of the DNA-alkylating agent methyl methanesulphonate.
142. Lopes, M. *et al.* The DNA replication checkpoint response stabilizes stalled replication forks. *Nature* **412**, 557–561 (2001).
143. Ivessa, A. S., Zhou, J. Q. & Zakian, V. A. The *Saccharomyces* Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell* **100**, 479–489 (2000).
144. Katou, Y. *et al.* S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature* **424**, 1078–1083 (2003).
145. Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A. & Labib, K. Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. *Genes Dev.* **19**, 1905–1919 (2005).
146. Franco, A. A., Lam, W. M., Burgers, P. M. & Kaufman, P. D. Histone deposition protein Asf1 maintains DNA replisome integrity and interacts with replication factor C. *Genes Dev.* **19**, 1365–1375 (2005).
147. Su, T. T. Cellular responses to DNA damage: one signal, multiple choices. *Annu. Rev. Genet.* **40**, 187–208 (2006).
148. Trenz, K., Smith, E., Smith, S. & Costanzo, V. ATM and ATR promote Mre11 dependent restart of collapsed replication forks and prevent accumulation of DNA breaks. *EMBO J.* **25**, 1764–1774 (2006).
149. Liberi, G. *et al.* Rad51-dependent DNA structures accumulate at damaged replication forks in sgs1 mutants defective in the yeast ortholog of BLM RecQ helicase. *Genes Dev.* **19**, 339–350 (2005).
150. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273**, 5858–5868 (1998).
151. Burma, S., Chen, B. P., Murphy, M., Kurimasa, A. & Chen, D. J. ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *J. Biol. Chem.* **276**, 42462–42467 (2001).
152. 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).
153. 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).

Acknowledgements

We thank H. Klein, K. Myung, A. Ramiro and F. Prado for comments on the manuscript and D. Haun for style supervision. We apologize to those whose publications could not be cited owing to space limitations. Research in A.A.'s laboratory is funded by grants from the Spanish Ministry of Science and Education (BFU2006-05260), Consolider Ingenio 2010 (CDS2007-015) and Junta de Andalucía (CVI102 and CVI624).

DATABASES

Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
c-Myc | *DNL4* | FRA10B | FRA16B | *MCM4* | *RAD9* | *RPA1* | *YKU70* | *YKU80*
OMIM: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Burkitt's lymphoma | fragile X mental retardation syndrome | Friedrich's ataxia | Huntington disease | myotonic dystrophy
UniProtKB: http://ca.expasy.org/sprot
AID | Asf1 | Cac1 | GreA | KU80 | LIG4 | MEC1 | MFD | Msh2 | Msh6 | Rad24 | Rad27 | Rfa1 | Rfc5 | Rrm3 | Sac3 | Sgs1 | Sic1 | Sub2 | Thp1 | UNG | Pif1 | XRCC4 | YRA1

FURTHER INFORMATION

Andrés Aguilera's homepage:
http://www.cabimer.es/en/dept/mb/genome-instability

SUPPLEMENTARY INFORMATION

See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF