

Endogenous DNA Damage as a Source of Genomic Instability in Cancer

Anthony Tubbs¹ and André Nussenzweig^{1,*}

¹Laboratory of Genome Integrity, NIH, Bethesda, MD 20892, USA

*Correspondence: andre_nussenzweig@nih.gov

<http://dx.doi.org/10.1016/j.cell.2017.01.002>

Genome instability, defined as higher than normal rates of mutation, is a double-edged sword. As a source of genetic diversity and natural selection, mutations are beneficial for evolution. On the other hand, genomic instability can have catastrophic consequences for age-related diseases such as cancer. Mutations arise either from inactivation of DNA repair pathways or in a repair-competent background due to genotoxic stress from cellular processes such as transcription and replication that overwhelm high-fidelity DNA repair. Here, we review recent studies that shed light on endogenous sources of mutation and epigenomic features that promote genomic instability during cancer evolution.

Introduction

DNA is the template for the basic processes of replication and transcription, making the maintenance of genetic stability critical for viability. Even before the discovery of the double helix in 1953, it was known that exogenous agents that we are exposed to in our everyday lives, such as X-rays, ultraviolet (UV) light, and various chemicals, can cause genetic changes that can promote cancer (Friedberg, 2008). It took an additional 10 years after its stable structure was elucidated to recognize that DNA is also subject to constant assault from endogenous sources during normal metabolism (Lindahl, 1993; Lindahl and Nyberg, 1972). Although both exogenous and endogenous lesions have the potential to modify the basic building blocks of genetic information, the relative contributions of intrinsic and extrinsic factors to organ-specific cancer incidence remains unclear (Tomasetti and Vogelstein, 2015; Wu et al., 2016).

It has been estimated that each human cell is subject to approximately 70,000 lesions per day (Figure 1) (Lindahl and Barnes, 2000). The majority of lesions (75%) are single-strand DNA (ssDNA) breaks, which can arise from oxidative damage during metabolism or base hydrolysis. ssDNA breaks can also be converted to DNA double-strand breaks (DSBs), which although much less frequent, are more dangerous. Given this extraordinary daily barrage of endogenous and exogenous DNA damage, it became evident that cells must have acquired enzymes during evolution that repaired DNA anomalies and thereby restored genome integrity (Friedberg, 2008; Lindahl and Barnes, 2000). In a personal perspective 21 years after the discovery of the double helix, Francis Crick wrote, “We totally missed the possible role of enzymes in repair although ... I later came to realize that DNA is so precious that probably many distinct mechanisms would exist. Nowadays, one could hardly discuss mutation without considering repair at the same time” (Crick, 1974).

Thanks to the pioneering work of Tomas Lindahl, Paul Modrich, and Aziz Sancar (who shared the 2015 Nobel Prize in Chemistry), among others, various DNA repair pathways were identified that

protect cells from different lesions to which they are subjected (Figure 2) (Kunkel, 2015). Tomas Lindahl discovered the pathway that repairs modified bases (base excision repair [BER]), and Paul Modrich discovered a distinct pathway that detects and removes bases that are mis-incorporated during DNA replication (mismatch repair [MMR]) (Figure 2D), whereas the mechanism for removal of bulky adducts in DNA (nucleotide excision repair [NER]) (Figure 2C), was proposed by Aziz Sancar. Each of these DNA repair pathways excise a damaged region and insert new bases to fill the gap (Figures 2A–2D).

The mechanisms for DSB repair are distinct and consist of two sub-pathways called homologous recombination (HR) and non-homologous end-joining (NHEJ). During HR, the DSB is repaired by exchanges of equivalent regions of DNA between homologous or sister chromosomes, whereas NHEJ religates the ends without the use of a template. NHEJ frequently leaves insertions or deletions at the breakpoint and therefore tends to be error prone whereas HR-mediated repair is of higher fidelity. Each of the DNA repair pathways needs to be coordinated with a series of signaling responses that arrest cell division or trigger cell death in case the lesions are irreparable (Haber, 2015).

Although DNA repair pathways have been extensively studied, the precise identity and sources of DNA damage that shape the mutational landscape of the cancer genome remain unclear. The Cancer Genome Project and the Cancer Genome Atlas have intersected with the field of DNA repair in the last few years as genetic mutations responsible for cancer have been identified. These large scale sequencing and bioinformatics approaches have revealed a remarkable diversity of somatic mutations, as well as specific signatures of DNA damage and errors in DNA repair in various cancers. These errors may occur if specific DNA repair and/or checkpoint pathways are inactivated, if DNA editing enzymes are deregulated, or if the damage load overwhelms DNA restorative capacity. Although exogenous agents such as carcinogens were once thought to be the major source of mutation, technological advances in the detection of DNA

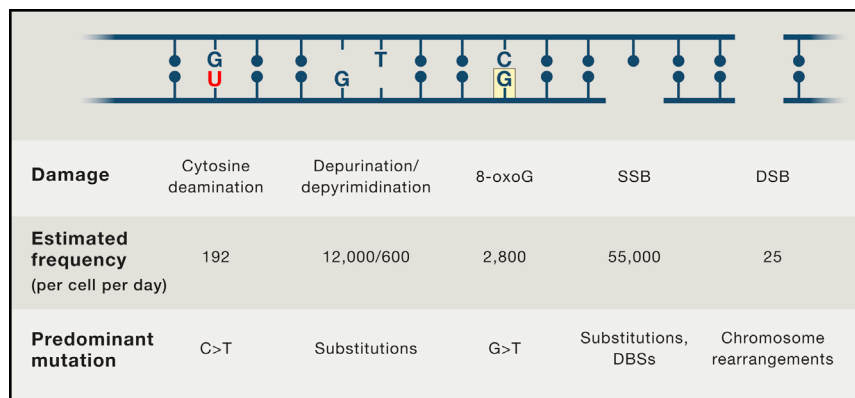


Figure 1. Estimated Frequencies of DNA Lesions and Mutations Associated with Dysfunctional DNA Repair

damage have revealed diverse and abundant types of endogenous DNA damage. Like noxious foreign chemicals, baseline DNA damage by endogenous processes may also overwhelm functional DNA repair machinery to generate mutations. Here, we review how endogenous sources of DNA damage and chromatin organization contribute to mutational processes that have been recorded in cancer genomes.

Inactivation of DNA Repair Pathways in Cancer

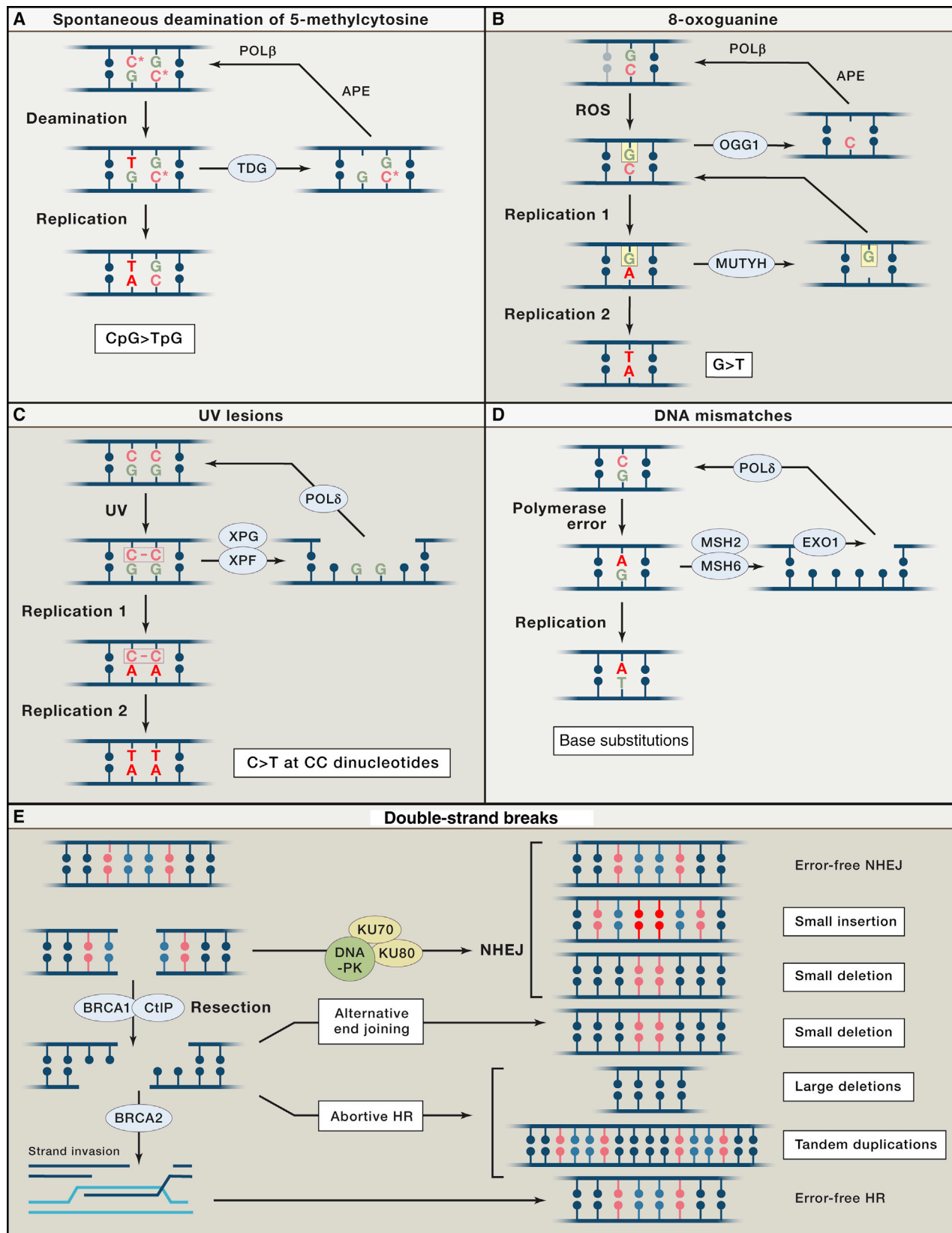
The classic mutator hypothesis (Loeb et al., 1974; Nowell, 1976) postulates that loss of DNA repair genes leads to genome instability, which in turn increases mutation rates at other genomic sites, leading to cellular transformation. Consistent with this hypothesis, an elevated mutation load associated with defective DNA repair underlies various hereditary cancers. For example, hereditary non-polyposis colon cancer is caused by defective MMR, and a large fraction of hereditary breast and ovarian cancers are accounted for by mutations in genes (*BRCA1* and *BRCA2*) that control DSB repair by HR (Figure 2E). Until recently, it has been difficult to detect germline or somatic DNA repair deficiencies in most types of cancer. However, next-generation sequencing has revealed more widespread deficits in DNA repair pathways across different cancers. For example, it has been found that men with advanced high-risk prostate cancer are more than five times as likely to harbor either heritable or somatic mutations in HR-mediated repair genes (most commonly *BRCA2* and *ATM*) and MMR genes (*MLH1* and *MLH2*) than patients with low-risk tumors are (Pritchard et al., 2016). Whole-genome sequences of breast cancers have also revealed deficiencies in both HR and MMR (Morganella et al., 2016; Nik-Zainal et al., 2016). Although these discoveries will undoubtedly provide new targets for therapy, it remains puzzling why malfunction of DNA repair genes is associated with organ-specific cancers despite their ubiquitous expression and function in all cell and tissue types. Variations in cancer risk could result from tissue-specific vulnerabilities to distinct intrinsic or extrinsic factors leading to distinct mutational processes.

Base Substitution and Rearrangement Signatures in Cancer

The classification of different mutation and rearrangement types into distinct signatures has uncovered pathways associated with

defective DNA maintenance (Alexandrov and Stratton, 2014). Although different mutational processes inevitably operate during the course of malignancy, there can be specific mutation types that dominate the repertoire for a particular cancer type. For example, a high proportion of thymine to guanine (T>G) base mutations found at the immunoglobulin (Ig) genes is the dominant feature of chronic lymphocytic leukemia, which reflects the activity of the error-prone DNA polymerase η during somatic hypermutation (Figure 3C) (Alexandrov et al., 2013). In melanoma, the majority of mutations are cytosine to thymine (C>T) at adjacent pyrimidine residues, which reflect the generation of thymine dimers after exposure to UV light. By classifying mutations according to the substitution and the sequence context immediately surrounding the mutated base and chromosomal rearrangements types (ie. deletions, duplication, inversions, and translocations), it is possible to deconvolve the spectrum of mutations into combinations of individual mutation types (Morganella et al., 2016; Nik-Zainal et al., 2016). In this way, tumors can be characterized by a combination of distinct signatures with different relative strengths that reflect the underlying mutational processes that occurred during tumorigenesis.

These analyses have revealed distinct signatures associated with deficiencies in HR, MMR, and NER in various cancers (Hellday et al., 2014). Interestingly, a mutational signature attributed to impaired BER activity has not been found so far, perhaps because loss of BER would compromise fitness. In some instances, the mutation(s) underlying the deregulated DNA repair pathway has not been determined, suggesting that new caretaker genes could be uncovered. It is also conceivable that in some cases, signatures might not reflect constitutive loss of DNA repair activity per se, but rather a strong intermittent process that temporarily overwhelmed the DNA repair capacity or provoked error-prone repair. Just as transient UV exposure will produce a signature similar to that expected in NER-deficient cancers, cellular processes such as oncogene activation, hormone signaling, and inflammation may occur sporadically in bursts, which could overwhelm an intact DNA repair pathway. In this context, recent single-cell analyses of triple negative breast cancer suggest that the majority of copy number aberrations are acquired in moments of crisis, followed by clonal expansions that produce the tumor mass (Gao et al., 2016). Examples include chromothripsis, in which a massive rearrangement was acquired in a single event, and chromoplexy, which involves coordinate and simultaneous rearrangements of multiple chromosomes (Zhang et al., 2013). Recent work also challenges the classic theory that pancreatic cancer evolves gradually through random acquisition of driving mutations, instead favoring a model where multiple driving mutations occur



(legend on next page)

simultaneously through genome rearrangements (Notta et al., 2016). These bursts of genomic instability may be facilitated by replication stress or transcriptional stress (described below) in an otherwise DNA-repair-proficient background.

Stochastic Errors in DNA Replication or Spontaneous Deamination as a Source of Mutation

Among the estimated 70,000 lesions per day that a normal cell is forced to cope with, each has the potential to be converted into a mutation. The most frequent single-nucleotide substitutions found in normal tissues and human tumors are transversion mutations that change a purine nucleotide to another purine (A>G) or a pyrimidine nucleotide to another pyrimidine (C>T) (Alexandrov and Stratton, 2014). Likely sources of these mutations are misincorporation by DNA polymerases or spontaneous deamination of cytosine to uracil.

Each time a cell divides, 6×10^9 nucleotides are replicated by DNA polymerases. DNA polymerases are also responsible for filling in single-strand gaps of DNA when mutated bases are excised during NER, MMR, and BER. Replication is predicted to generate mutations at a low but constant rate, and the variation in cancer risk across different tissues might be explained by the total number of cell divisions in normal stem cells derived from that tissue (Tomasetti and Vogelstein, 2015). Recently, mutation rates in human stem cells have been determined to accumulate at a steady rate of approximately 40 novel mutations per year in various tissue types (Blokzijl et al., 2016). Some of these mutations could be generated during translesion synthesis (TLS), which allows cells to efficiently complete DNA replication across sites of DNA damage (Goodman and Woodgate, 2013). DNA translesion polymerases are error prone and are thought to promote tolerance of DNA damage, which can contribute to cancer evolution.

The most recurrent set of mutations that accumulate in stem cells with high division rate represents the aging signature (Alexandrov and Stratton, 2014), which has also been found to be the most common base substitution signature in cancer. This signature, which correlates with patient age, is made up of predominantly C>T transitions at CpG dinucleotide motifs, reflecting the inherent mutability of methylated cytosine (Lindahl, 1993). Indeed, C>T substitutions at CpGs make up 25% of somatic mutations in TP53 codons, implying that mutations associated with this signature might drive tumorigenesis (Olivier et al., 2010).

The deaminated product of methylated cytosine is thymine, which leads to transition mutations if the T:G mismatch is not

recognized and repaired prior to DNA replication (Figure 2A). In addition to incorporating the mutation, DNA replication also increases the probability of spontaneous deamination through transient exposure of ssDNA and during post-replicative conversion of cytosine to 5-methylcytosine (Lindahl, 1993). Although 70% to 80% of CpG cytosines are methylated in mammalian cells, the CpG dinucleotide occurs with a frequency of less than one quarter of that which would be expected by random chance. It has been proposed that CpG deficiency is due to the increased mutation rate of 5-methylcytosine (Scarano et al., 1967), which likely stems from inefficient excision of T:G mismatches. 5-methylcytosine is demethylated by TET family DNA demethylases through a 5-hydroxymethylcytosine intermediate. Recent work shows that sites of 5-hydroxymethylcytosine are associated with C>G transversions in cancer genomes, and increased C>G correlates with the expression of TET enzymes (Supek et al., 2014). Thus, modified cytosine might also be inherently mutagenic.

Aging poses the greatest risk for cancer. The free radical theory of aging posits that reactive oxygen species produced by normal cellular metabolism are responsible for the damage to biomolecules resulting in organismal aging (Harman, 1956). One of the most common lesions arising from reactive oxygen is 8-oxoguanine (8-oxoG), which can result in mispairing with adenine resulting in G>T substitutions if 8-oxoG is not removed prior to DNA replication (Figure 2B). Even though 8-oxoG has long been believed to be a primary driver of mutagenesis in both nuclear and mitochondrial DNA, a mutational signature attributed to oxidative DNA lesions has not been found in primary cancers or in mitochondrial DNA during aging (Helleday et al., 2014; Kennedy et al., 2013). The lack of G>T substitutions associated with age might indicate that removal of 8-oxoG is extremely efficient (Banerjee et al., 2006) (Figure 2B). In contrast, T:G mismatches that arise during spontaneous deamination of 5-methylcytosine might either occur more frequently or might not be recognized as efficiently by the MMR machinery or the thymine DNA glycosylase (TDG), which recognizes T:G mismatches (Figure 2A).

DNA Editing as a Source of Mutation

ssDNA is a substrate for the cytosine deaminase family of enzymes, which includes APOBECs and activation-induced cytidine deaminase (AID) (see below) (Swanton et al., 2015). APOBECs normally function in protecting against retroviruses

Figure 2. DNA Repair Pathways and Mutations Associated with Dysfunctional Repair

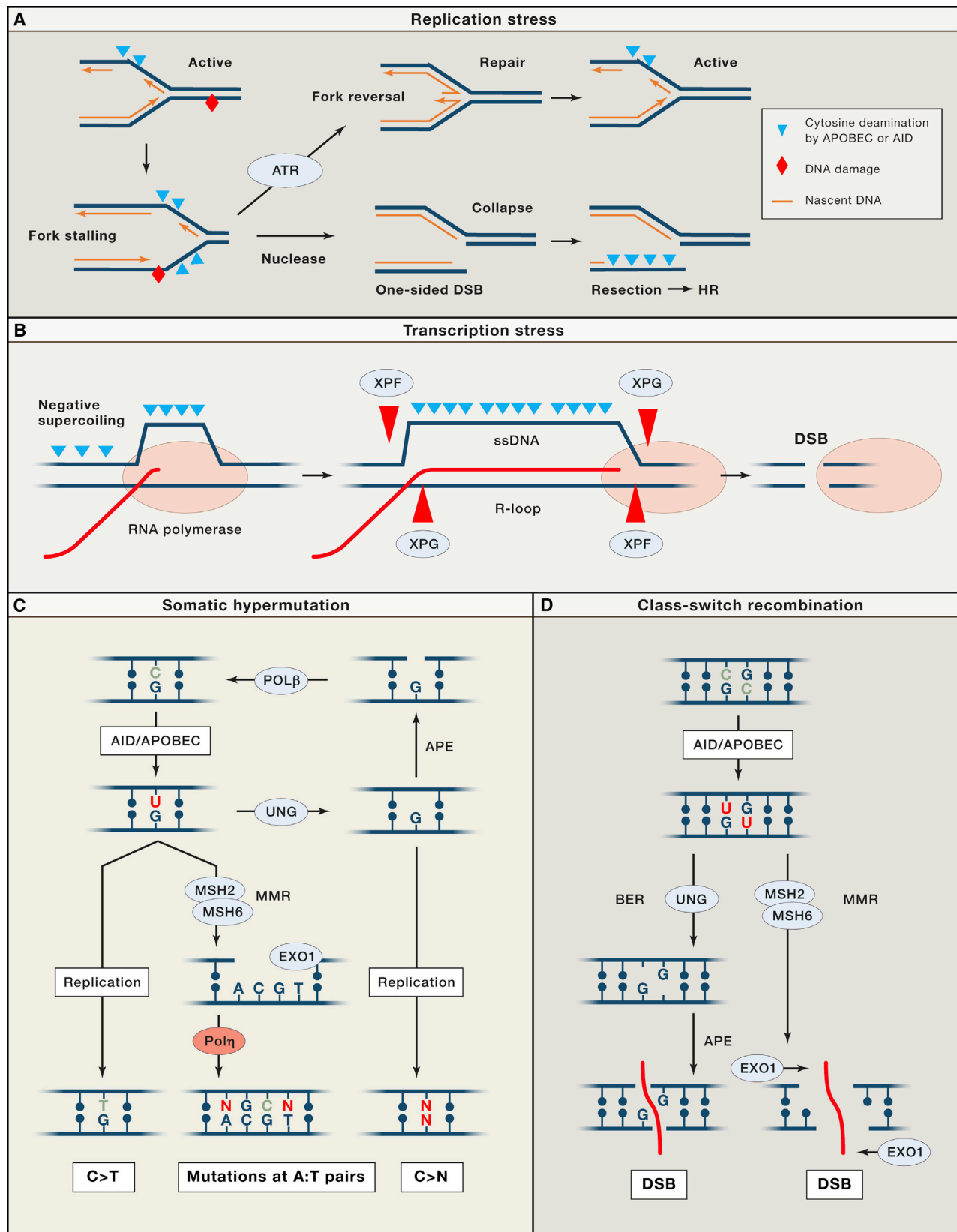
(A) Spontaneous deamination of 5-methylcytosine (C*) generates an aging signature, characterized by CpG>TpG transition mutations. 5-methylcytosine deamination converts C>T, which is recognized by thymine DNA glycosylase (TDG). Apurinic/apyrimidinic endodeoxyribonuclease (APE) excises the abasic nucleotide, which is replaced by polymerase beta (POLβ).

(B) Reactive oxygen species (ROS) oxidizes guanine to form 8-oxoG. 8-oxoG can pair with adenine and may generate G>T mutations if 8-oxoG:C pairing is not detected by OGG1 before replication or if the 8-oxoG:A pairing is not recognized by MTH1 after the first round of replication.

(C) Exposure to UV light generates cyclobutane pyrimidine dimers, which are recognized by nucleotide excision repair (NER) factors and excised by XPG and XPF flap endonucleases. Polymerase delta (POLδ) synthesizes DNA to restore C:G pairing. Failure to complete NER results in C>T transition mutations at CC dinucleotides.

(D) DNA polymerase errors induce mismatch repair (MMR) through MSH2/MSH6. Exonuclease I (EXO1) removes a short patch of DNA that is resynthesized by POLδ. Failure to complete MMR results in base substitutions.

(E) DNA DSBs are repaired using non-homologous end joining (NHEJ) or homologous recombination (HR). Normal NHEJ may result in small insertions or deletions, as does an inherently error-prone form of end-joining (alternative end joining), which relies on short homologous sequences (microhomologies) for base pairing. Normal HR is an error-free repair mechanism, but abortive HR may yield chromosomal rearrangements, such as tandem duplications or large deletions, which are common in HR-deficient cancers. BRCA1 and CtIP promote DNA end resection during HR, whereas BRCA2 promotes strand invasion to facilitate normal HR.



(legend on next page)

and other mobile elements (Swanton et al., 2015). APOBEC3B is upregulated in approximately half of all cancers and promotes chemotherapy resistance in estrogen-receptor-positive breast cancer (Alexandrov et al., 2013; Burns et al., 2013; Law et al., 2016). The APOBEC mutational burden ranks second only to the aging signature in cancer (Alexandrov et al., 2013; Burns et al., 2013). APOBEC-induced mutations occur either as strand-coordinated clusters (kataegis) or as scattered mutations (Alexandrov et al., 2013). APOBEC-associated mutational signatures occur at a higher rate on the lagging strand (Haradhvala et al., 2016; Morganella et al., 2016), which during DNA synthesis endures longer exposure as ssDNA relative to the leading strand (Figure 3A). In addition, collapsed replication forks are repaired by HR, which generates a long 3' ssDNA overhang during end resection that, in theory, can be a substrate for APOBEC (Figure 3A). Interestingly, localized APOBEC hypermutations show marked colocalization and appear to be coupled to genomic rearrangements (Alexandrov et al., 2013), perhaps because of high levels of ssDNA formed as a result of replication stress and transcription stress (see *Gene Activation Can Destabilize the Genome*) (Kanu et al., 2016) (Figures 3A and 3B).

The Ig genes are targets of high density AID-mediated cytosine deamination during somatic hypermutation and class-switch recombination in B cells (reviewed in (Di Noia and Neuberger, 2007)). AID has a preference for deaminating cytosine residues flanked by a 5' purine, making its signature distinguishable from APOBECs, which show a variety of sequence preferences (Swanton et al., 2015). In either case, the U:G mismatches, if replicated, will generate C>T transition mutations (Figure 3C). Alternatively, these mismatches can be processed into DSBs, leading to class-switch recombination (Figure 3D). In addition to the Ig heavy- and light-chain loci, AID initiates recurrent DSBs in non-Ig loci, predominantly within B cell super enhancers and regulatory clusters (Meng et al., 2014; Qian et al., 2014). Some of these off-target sites are proto-oncogenes that, when translocated to potent Ig enhancers, lead to their constitutive expression. Despite the risk of B cell genomic instability associated with AID off-target deamination activity and DSB formation (Figure 3D), the majority of the AID-generated uracils in non-Ig loci are corrected by high-fidelity DNA repair and hence are not detectable as mutations (Liu et al., 2008). Thus, AID-mediated cytosine deamination seems to be differentially repaired depending on genomic context. The mechanisms that allow high-fidelity BER and/or MMR machineries to perform error-prone repair at some loci whereas others are spared from muta-

tion is not known but might be linked to B cell transformation (Liu et al., 2008). Numerous alternative possibilities have been suggested for differential repair, including distinct chromatin structures that might recruit error-prone versus high-fidelity repair factors, variation in repair capacity during the cell cycle, and differences in the replication timing of the targeted genes (Liu and Schatz, 2009).

Influence of Chromatin Organization on Regional Mutation Rates

There is a great variation in the mutation load in cancer, ranging from frequencies of less than one mutation per Mb in some childhood cancers to greater than 250 mutations per Mb in ultramutated biallelic MMR-deficient brain tumors (Shlien et al., 2015). Most cancers have been found to harbor elevated rates of base substitutions in heterochromatic late replicating regions and reduced rates in early replicating regions (Schuster-Böckler and Lehner, 2012). High mutation densities correlate with repressive histone modifications (e.g., H3K9me3 and H4K20me3) and anti-correlate with marks of open chromatin such as H3K4me3, GC content, and genes with higher expression (Schuster-Böckler and Lehner, 2012) (Figure 4A). This association is upheld across diverse tissue and signature types and is also conserved in *S. cerevisiae*, in which nucleotide mis-incorporations are dominant during late phases of the cell cycle (Waters and Walker, 2006). This suggests that replication timing and the chromatin landscape can influence lesion formation and DNA repair fidelity.

It has been suggested that base substitutions in cancer are enriched in genomic regions that undergo late replication because of differential DNA repair rather than differences in mutation rates (Supek and Lehner, 2015; Zheng et al., 2014). One possibility is that early S phase templates have more time to recognize and repair mutations prior to mitosis. In addition, the DNA MMR machinery, which is coupled to DNA replication, appears to be more effective in euchromatic early replicating regions (Supek and Lehner, 2015; Zheng et al., 2014) (Figure 4B). NER also reduces the rate of mutation in genic regions (Zheng et al., 2014), and differences in repair rates between transcribed and non-transcribed DNA have been linked to high-fidelity-transcription-coupled NER (Haradhvala et al., 2016; Morganella et al., 2016; Nik-Zainal et al., 2012) (Figure 4B). Consistent with differences in repair across the genome, when either the NER or MMR pathway is lost, mutations are more evenly distributed (Supek and Lehner, 2015; Zheng et al., 2014).

Figure 3. Causes and Consequences of Enzyme-Induced Cytosine Deamination

(A) ssDNA exposed during replication fork progression is a substrate for cytosine deamination. Active replication forks expose ssDNA on the lagging strand. Stalled replication forks expose ssDNA on both the leading and lagging strand, which activates ATR to stop replication and stabilize the fork through controlled fork reversal. Endonuclease cleavage causes replication fork collapse into one-ended DNA DSBs, which must be resected before repair by HR.

(B) ssDNA exposed during transcription is a substrate for cytosine deamination. RNA polymerase (RNAP) exposes ssDNA while synthesizing nascent RNA (red) at the transcription bubble and behind the RNAP as a result of negative supercoiling (left). Transcription stress induces the formation of R-loops (middle), which exposes long stretches of ssDNA on the non-template strand. R-loops may be resolved in DSBs through cleavage by TC-NER components XPF and XPG (right).

(C) Mechanism of enzyme-induced somatic hypermutation (SHM). Cytosine deamination generates a U:G mismatch, which is recognized by uracil glycosylase UNG to generate an abasic site opposite the G. Apurinic/apyrimidinic endodeoxyribonuclease (APE) excises the abasic nucleotide which is repaired by POL β . If the abasic site is replicated over prior to repair, this can give rise to transition or transversion mutations (N). If the U:G mismatch is replicated prior to repair, C>T substitution occurs. Mutations at neighboring A:T base pairs can occur if the U:G pair is recognized by the MMR pathway and repaired by error prone polymerase η (Pol η).

(D) Mechanisms of enzyme-induced rearrangements and class-switch recombination (CSR). If cytosine deamination occurs nearby on opposite strands, repair through BER or MMR will form DNA DSBs, leading to chromosomal rearrangements or CSR.

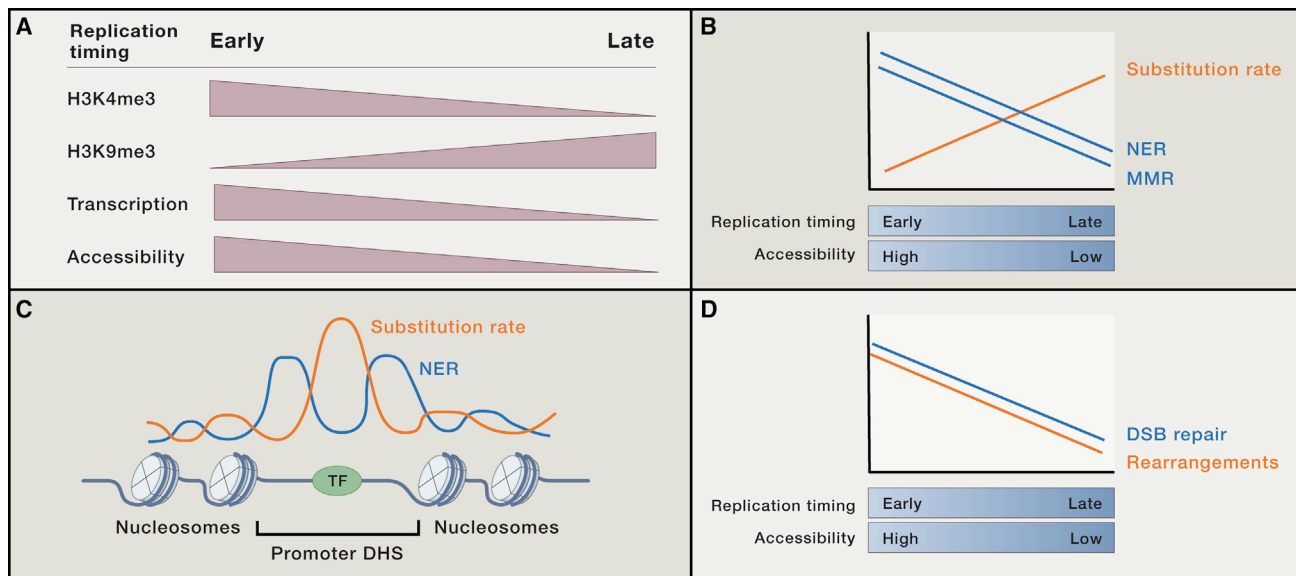


Figure 4. Replication Timing and Protein-DNA Interactions Influence Regional Mutation Rates

(A) Early replicating regions of the genome are associated with high levels of H3K4me3, transcription, increased chromatin accessibility, and low levels of H3K9me3.

(B) In cancer genomes, accessible early replicating regions have fewer base-pair substitutions. MMR is more likely to recognize mismatches in early replication regions. NER is also more efficient in transcribed, accessible chromatin.

(C) DNA accessibility determines efficiency of NER (adapted from Sabarinathan et al., 2016). Repair of UV-induced lesions is efficient within DNase I hypersensitive sites (DHSs), but is impaired by nucleosome occupancy and transcription factor binding. Conversely, UV-induced mutations are more frequent where DNA-binding proteins interfere with NER machinery.

(D) Whereas DSBs are more efficiently repaired in accessible chromatin, chromosome rearrangements in cancer genomes accumulate at early replicating regions.

Even within accessible chromatin, there is variation in mutation density (Figure 4C). For example, in several cancer types, including melanoma, ovarian, lung, astrocytoma, esophageal, and prostate cancers, there is increased mutation at the peaks of DNase I hypersensitive sites relative to accessible flanking regions, even though both regions are nucleosome free (Perera et al., 2016; Sabarinathan et al., 2016) (Figure 4C). In melanoma, such regulatory regions have mutation rates four times higher than flanking sequences. Evidence suggests that peaks of accessible chromatin within regulatory DNA (i.e., promoters) are bound by proteins such as the transcription initiation machineries that exclude NER and other repair proteins (Perera et al., 2016; Sabarinathan et al., 2016) (Figure 4C). Genome-wide analysis of excision repair after UV damage in primary skin fibroblasts has demonstrated a reduction in repair within transcription factor binding sites at DNase I hypersensitive sites, whereas NER-deficient cells do not show any increased promoter mutation density (Perera et al., 2016; Sabarinathan et al., 2016). Interestingly, recent sequencing of individual neurons from the human prefrontal cortex revealed that each neuron has a distinct genome with more than 1,000 mutations per cell (Lodato et al., 2015). Mutation hotspots in neurons occurred predominantly at transcriptionally-active loci and at DNase I hypersensitive sites (Lodato et al., 2015). This suggests the possibility that even in neurons, mutation hotspots can result from the exclusion of repair proteins by DNA-bound factors. Impaired DNA damage repair is also a common feature of neurodegenerative diseases (McKinnon, 2013). In neuronal progenitors, however, recurrent

DNA DSB clusters occur in gene bodies of late-replicating genes, possibly contributing to functional diversity during neuronal development (Wei et al., 2016).

Influence of Chromatin Organization on Large-Scale Chromosomal Rearrangements

Chromosomal rearrangements (deletions, tandem duplications, inversions, and translocations) are initiated by DNA DSBs. Overall, rearrangements are 10- to 1,000-fold less frequent than somatic mutations but also show regional variation across a cancer genome. However, in stark contrast to substitution mutations, which are predominant in heterochromatin, rearrangements are enriched in early replicating regions (Morganella et al., 2016; Sima and Gilbert, 2014), including those that may arise through a single catastrophic event (Baca et al., 2013; Zhang et al., 2013).

Unlike the case for base substitutions, it appears that increased DNA damage rather than deficient DNA repair might underlie the predominance of rearrangements in early replicating regions (Figure 4D). While different chromatin environments may selectively utilize NHEJ or HR pathways (Chiolo et al., 2011; Tsouroula et al., 2016), it is generally believed that the heterochromatin superstructure confers a barrier to DSB repair. Consistent with this, euchromatic DSBs are more rapidly repaired than heterochromatic DSBs (Goodarzi et al., 2008) (Figure 4D). Thus, early replicating regions of the genome might be expected to harbor fewer rearrangements, unless there were an overabundance of DSBs in these regions relative to late

replicating regions. One hallmark feature of early replicating euchromatin is the association with gene-rich, transcriptionally active, and DNase I accessible regions (Figure 4A). Nearly three quarters of expressed genes are estimated to be early replicating (Sima and Gilbert, 2014). As discussed below, there is evidence that high levels of transcription and the increased probability of conflicts with the replication machinery pose a substantial threat to genome integrity (Aguilera, 2002).

Oncogenic Stress Promotes DNA Damage and Genomic Rearrangements in Early Replicating Regions

Oncogene activation is a major driving force in cancer development (Halazonetis et al., 2008). Oncogenes stimulate new replication origin firing, which depletes nucleotide pools, decreases replication fork speed, and results in stalled replication forks, which are prone to collapse into DSBs (Halazonetis et al., 2008). Replication stress leads to exposure of tracts of ssDNA that are substrates for endonuclease cleavage, which generates DSBs (Cortez, 2015), or for APOBEC3-deaminase-mediated mutagenesis (Kanu et al., 2016). By deregulating the timing of replication initiation and progression, oncogenes might also disrupt the spatio-temporal segregation that normally prevents conflicts between transcription and DNA replication (Halazonetis et al., 2008). This might be exacerbated by oncogenes such as *c-myc*, which amplifies gene expression at all active genes (Nie et al., 2012), thereby posing a further impediment to replication. Recent studies in bacteria indicate that replication-transcription collisions can cause replication fork stalling and DNA breaks leading to two types of mutation signatures: duplications and deletions within the transcriptional unit and promoter-localized base substitutions (Sankar et al., 2016).

Oncogene-induced DNA damage is not random, but preferentially occurs at fragile sites (Sarni and Kerem, 2016). In contrast to late replicating common fragile sites (CFSs), early replicating fragile sites (ERFSs) appear near mammalian origins of replication within highly expressed gene clusters enriched for repetitive elements and CpG dinucleotides (Barlow et al., 2013). These regions may be particularly susceptible to replication-transcription collisions, creating the potential for genome rearrangements. Indeed, ERFSs identified in primary B cells account for a large fraction of recurrent amplifications and/or deletions in human diffuse large B cell lymphomas (Barlow et al., 2013).

In addition to oncogenes, recent cancer genome sequencing studies have identified the histone methyltransferases KMT2C and KMT2D, among the most broadly mutated genes in human cancer (Lawrence et al., 2014). KMT2C and KMT2D belong to the family of mammalian mixed-lineage leukemia (MLL) genes that encode histone methyltransferases, responsible for methylation at H3K4 at a subset of active gene promoters and enhancers (Lee et al., 2013). Despite their mutation in many cancers, the roles of KMT2C and KMT2D in genome stability have only recently been uncovered. KMT2C- and KMT2D-dependent H3K4 methylation was found to be induced at replication forks upon replication stress (Ray Chaudhuri et al., 2016), and KMT2D physically interacts with the helicase RECQL5 to prevent collisions between transcription and replication machineries (Kantidakis et al., 2016; Saponaro et al., 2014). Like RECQL5-deficient cells, KMT2D-deficient cells exhibited

transcription stress associated with RNAPII pausing, stalling, and backtracking, leading to chromosomal aberrations primarily within ERFSs (Kantidakis et al., 2016). Thus, cancer-driving oncogenes and mutations in KMT2C and KMT2D could potentiate DNA breaks and chromosomal rearrangements preferentially in early replicating regions.

Gene Activation Can Destabilize the Genome

Like oncogenes, hormone stimulation of gene expression may induce genome instability (Haffner et al., 2010; Ju et al., 2006; Lin et al., 2009; Williamson and Lees-Miller, 2011). In estrogen-receptor-positive breast cancer cells, estrogen stimulation boosts expression of a broad cohort of genes, promotes cell cycle entry, and increases DNA damage (Ju et al., 2006; Williamson and Lees-Miller, 2011). Robust transcription at estrogen-responsive genes results in co-transcriptional structures known as R-loops, or RNA:DNA hybrids associated with unannealed ssDNA (Stork et al., 2016) (Figure 3B). Although RNA processing factors normally resolve RNA:DNA hybrids, it has been suggested that hormones or other stimuli that promote high levels of transcription might overload the system, resulting in the accumulation of R-loops (Stork et al., 2016). The displaced ssDNA non-transcribed strand can then act as a substrate for cytosine deamination by APOBEC, which through the activation of BER pathways can lead to DSB formation (Periyasamy et al., 2015) (Figure 3B). Alternatively, unprocessed R-loops can activate the transcription-coupled NER (TC-NER) endonucleolytic machinery, which can excise the RNA:DNA hybrid, generating a DSB upon encounter with a replication fork (Cortez, 2015; Stork et al., 2016) (Figure 3B). Interestingly, a correlation was observed between genomic rearrangements in breast tumors and estrogen-responsive genes, suggesting a tissue-specific vulnerability to an endogenous but potentially genotoxic agent (Stork et al., 2016). In summary, ssDNA associated with R-loops can be mutagenetic and recombinogenic.

Independent of collisions with replication forks, transcription itself might contribute to genome instability. Over 20 years ago, it was demonstrated that an increased transcription level stimulates spontaneous mutagenesis in yeast (Datta and Jinks-Robertson, 1995). Recent genome-wide sequencing approaches in mammalian cells have revealed that transcriptionally-active regions also harbor more translocations and are more fragile than less-active regions (Barlow et al., 2013; Chiarle et al., 2011; Klein et al., 2011). One potential source of transcriptional stress is the superhelical tension produced as RNA polymerase traverses DNA. For example, negative supercoiling (underwinding) can lead to stretches of ssDNA behind the advancing polymerase, with potential for enzyme-induced cytosine deamination and mutation (Figure 5A). Torsional stress is normally relieved by topoisomerases, enzymes that catalyze transient cleavage of the DNA backbone, followed by one rotation of the DNA strand before resealing (Pommier et al., 2016). The broken DNA intermediates are thought to be sequestered from the DNA damage surveillance machinery. However, at some unknown frequency, topoisomerases fail to complete the reaction, which can result in trapped cleavage complexes and persistent DNA damage on one or both DNA strands (Figure 5B). Notably, several DNA alterations, including oxidative

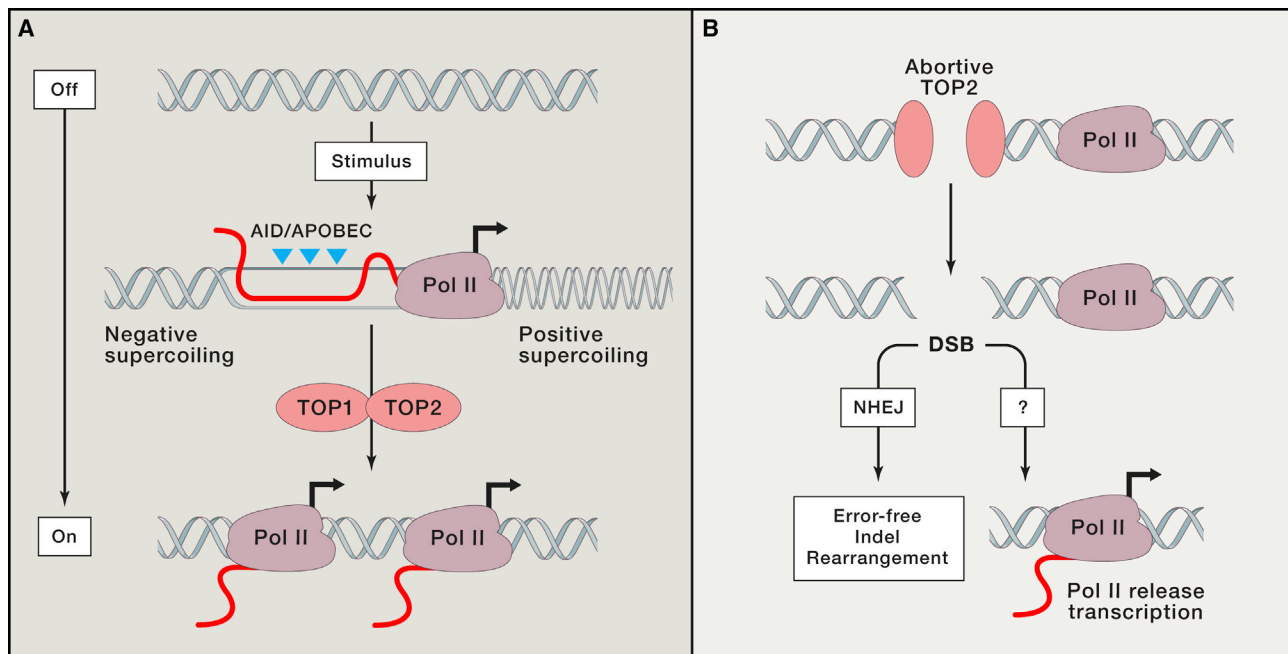


Figure 5. Canonical and Non-canonical Topoisomerase Activity and Gene Activation

(A) Upon gene activation, canonical topoisomerase activity promotes gene expression. The enzymes TOP1 and TOP2 cleave one or both strands to alleviate positive supercoiling ahead of and negative supercoiling behind RNA polymerase (Pol II). ssDNA behind Pol II can be a substrate for cytosine deamination. (B) Non-canonical, abortive TOP2 reactions result in protein-linked DNA breaks that are resolved the NHEJ repair pathway. Abortive TOP2 activity also promotes transcription by an unknown mechanism. Persistent DNA damage may also result in chromosomal alterations.

lesions, base modifications, mismatches, and ssDNA breaks enhance the frequency of persistent topoisomerase cleavage complexes (Pommier et al., 2016). These can eventually give rise to DSBs and short deletions (Lippert et al., 2011; Takahashi et al., 2011), implicating trapped cleavage complexes as a source of mutation.

Recently, non-canonical roles for topoisomerases have also been observed, in which prolonged topoisomerase-mediated DNA damage is necessary to promote transcription (Figure 5B). In cells activated by diverse stimuli, including heat shock, serum induction, insulin, androgen, estrogen, and various neuronal stimuli, the transcriptional program appears to require persistent topoisomerase-mediated DNA damage that is recognized by the DNA damage transducers ATM, DNA-PKcs, and PARP1 (Bunch et al., 2015; Haffner et al., 2010; Ju et al., 2006; Lin et al., 2009; Madabhushi et al., 2015; Perillo et al., 2008; Wong et al., 2009). For example, upon activation of postmitotic neurons with NMDA, H2AX is rapidly phosphorylated in transcribed regions of early-response genes (Madabhushi et al., 2015). It is postulated that DNA damage signaling might be necessary for the release of paused RNA polymerase at promoters or for conformational changes that facilitate promoter-enhancer interactions (Calderwood, 2016).

Although persistent DNA damage induced by non-canonical topoisomerase activity promotes transcription, it can also lead to genome instability. For example, androgen stimulation in prostate cells leads to TOP2 β -dependent androgen receptor target gene expression, as well as genomic recombination events at activated loci, leading to prostate cancer rearrange-

ments (Haffner et al., 2010; Lin et al., 2009). The probability of aberrant rearrangements is likely to be exacerbated by deficiencies in HR, which are found in as many as 25% of advanced prostate cancers (Pritchard et al., 2016). It has been speculated that the inflammatory milieu might stabilize TOP2-DNA cleavage complexes, perhaps due to increases in reactive oxygen species surrounding the prostate lesion (De Marzo et al., 2007; Pommier et al., 2016). Inflammation also induces the expression of hundreds of pathogen-associated genes, which require topoisomerase I for nucleosome remodeling and transcriptional elongation (Rialdi et al., 2016). If such inflammatory transcriptional responses similarly result in persistent topoisomerase lesions, they might contribute to mutation and/or chromosomal rearrangements.

Conclusions and Future Perspectives

The mutational landscape of cancer is generated through a combination of environmental and endogenous stresses that cause base substitutions, insertions, deletions, and chromosomal rearrangements. A major lesson from recent work is that even in repair-sufficient cells, endogenous and oncogenic stress can occasionally overwhelm normal genome maintenance pathways (Figure 6). One major outcome of these stresses is exposure of ssDNA, which is subject to nuclease activity, spontaneous hydrolysis, and enzyme-induced deamination that often results in base substitutions or processing into DSBs. Hormonal signaling and inflammatory responses may act similar to oncogenes by driving cells to undergo high levels of transcriptional activity or replication stress that leads to genome instability.

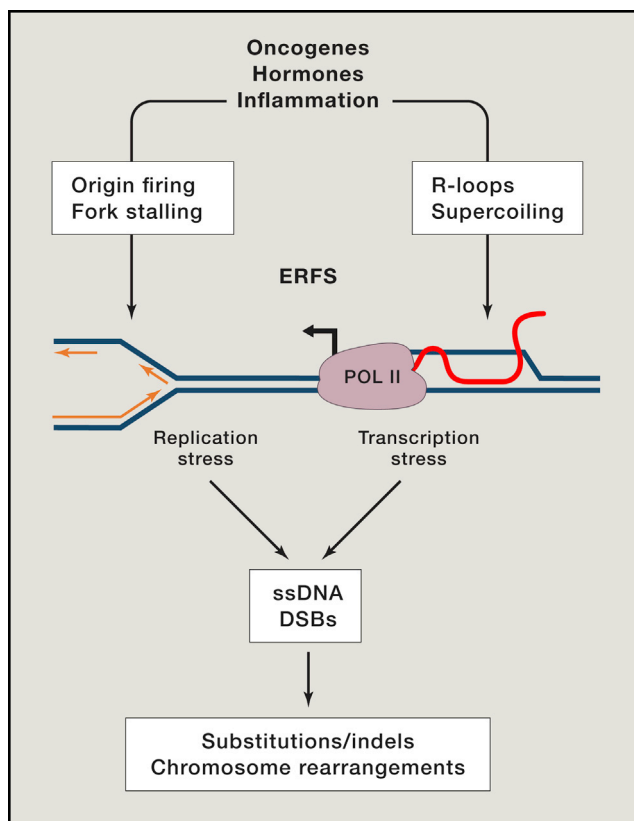


Figure 6. Replication and Transcription Stress Contribute to the Mutational Landscape of Cancer

An emerging model for how mitogenic signaling promotes genome instability at early replicating fragile sites (ERFSs). Oncogenes, hormones, and inflammation can promote unscheduled activation of replication and transcription programs. This can result in replication stress and transcription stress through mechanisms that are largely unknown. The generation of ssDNA and DSBs at ERFSs can lead to genomic instability.

Despite the fact that each cell in the human body is exposed to thousands of DNA lesions per day, most are efficiently repaired. Unfortunately, it is not possible with current technologies to accurately quantify the steady-state levels of several types of DNA lesions. For example, the most reliable method to detect ssDNA breaks (i.e., the alkali comet assay) is sensitive to only a few thousand lesions per cell (or one ssDNA break every million bp). Similarly, it is difficult to estimate the frequency and location of replication-associated DSBs and protein-blocked lesions arising from abortive topoisomerase activity. Although the phosphorylation of the histone variant H2AX, or γ -H2AX, has been extensively used as a surrogate marker of DSBs, it also marks single-stranded regions exposed during replication stress and spreads over hundreds of kilobases. However, recent DNA break capture and next-generation-sequencing-based methods have begun to quantitate and precisely locate genomic co-ordinates of low frequency DNA lesions, translocations, and other infrequent complex rearrangements (Canela et al., 2016; Frock et al., 2015; Lensing et al., 2016; Zhang et al., 2015). These emerging technological and computational advances should enable a better understanding

of the various sources, risk factors, and mechanisms that contribute to genome instability.

Our comprehensive understanding of genome maintenance pathways is a direct result of decades of work investigating DNA repair mechanisms. This is the basis for current clinical cancer treatments predicated on chemotherapy and ionizing radiation that damage DNA and in turn overwhelm the cellular DNA repair capacity in highly proliferative tumor cells. As effective as they have been, treatments that produce DNA damage on a global scale frequently result in side effects, secondary tumors, and drug resistance. Precision therapies based on synthetic lethality exploit cancer-specific mutations by targeting deficient DNA repair and thereby reduce collateral damage to normal cells (Lord et al., 2015). Tumors with defective DNA repair might also respond well to immune checkpoint inhibitors by generating mutations that produce neoantigens that enhance T cell reactivity (Schumacher and Schreiber, 2015). Additionally, by manipulating highly intricate DNA repair pathways, it has become possible to generate site-specific nucleotide substitutions of DNA through CRISPR/Cas9 base editing (Komor et al., 2016; Nishida et al., 2016; Paquet et al., 2016), which might provide a complementary approach to correcting known germline mutations that predispose individuals to cancer. In conclusion, a deeper understanding of how DNA lesions are generated, processed, and repaired will continue to provide insights and new opportunities for cancer prevention and treatment.

ACKNOWLEDGMENTS

We are especially grateful to Joshua Waterfall, Ferenc Livak, Avinash Bhandoola, and Sam John for comments on the manuscript and to Jiri Lukas, Keith Caldecott, and Yossi Shiloh for discussions. This work was supported by the Intramural Research Program of the NIH, the National Cancer Institute, and the Center for Cancer Research. A.N. was also supported by the US Department of Defense (BCRP DOD Idea Expansion Award BC133858 and BCRP Breakthrough Award BC151331), the Ellison Foundation Award for Aging Research, and Alex's Lemonade Stand Foundation Reach Award. A.T. has been supported by a fellowship from the American Cancer Society (PF-16-037-01-DMC).

REFERENCES

- Aguilera, A. (2002). The connection between transcription and genomic instability. *EMBO J.* 21, 195–201.
- Alexandrov, L.B., and Stratton, M.R. (2014). Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* 24, 52–60.
- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.
- Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., et al. (2013). Punctuated evolution of prostate cancer genomes. *Cell* 153, 666–677.
- Banerjee, A., Santos, W.L., and Verdine, G.L. (2006). Structure of a DNA glycosylase searching for lesions. *Science* 311, 1153–1157.
- Barlow, J.H., Faryabi, R.B., Callén, E., Wong, N., Malhowski, A., Chen, H.T., Gutierrez-Cruz, G., Sun, H.W., McKinnon, P., Wright, G., et al. (2013). Identification of early replicating fragile sites that contribute to genome instability. *Cell* 152, 620–632.

- Blokzijl, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kuijk, E., Prins, P., et al. (2016). Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 538, 260–264.
- Bunch, H., Lawney, B.P., Lin, Y.F., Asaithamby, A., Murshid, A., Wang, Y.E., Chen, B.P., and Calderwood, S.K. (2015). Transcriptional elongation requires DNA break-induced signalling. *Nat. Commun.* 6, 10191.
- Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., et al. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494, 366–370.
- Calderwood, S.K. (2016). A critical role for topoisomerase IIb and DNA double strand breaks in transcription. *Transcription* 7, 75–83.
- Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B.P., and Nussenzweig, A. (2016). DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Mol. Cell* 63, 898–911.
- Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., et al. (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* 147, 107–119.
- Chiolo, I., Minoda, A., Colmenares, S.U., Polyzos, A., Costes, S.V., and Karpen, G.H. (2011). Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* 144, 732–744.
- Cortez, D. (2015). Preventing replication fork collapse to maintain genome integrity. *DNA Repair (Amst.)* 32, 149–157.
- Crick, F. (1974). The double helix: a personal view. *Nature* 248, 766–769.
- Datta, A., and Jinks-Robertson, S. (1995). Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* 268, 1616–1619.
- De Marzo, A.M., Platz, E.A., Sutcliffe, S., Xu, J., Grönberg, H., Drake, C.G., Nakai, Y., Isaacs, W.B., and Nelson, W.G. (2007). Inflammation in prostate carcinogenesis. *Nat. Rev. Cancer* 7, 256–269.
- Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* 76, 1–22.
- Friedberg, E.C. (2008). A brief history of the DNA repair field. *Cell Res.* 18, 3–7.
- Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* 33, 179–186.
- Gao, R., Davis, A., McDonald, T.O., Sei, E., Shi, X., Wang, Y., Tsai, P.C., Casasent, A., Waters, J., Zhang, H., et al. (2016). Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat. Genet.* 48, 1119–1130.
- Goodarzi, A.A., Noon, A.T., Deckbar, D., Ziv, Y., Shiloh, Y., Löbrich, M., and Jeggo, P.A. (2008). ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Mol. Cell* 31, 167–177.
- Goodman, M.F., and Woodgate, R. (2013). Translesion DNA polymerases. *Cold Spring Harb. Perspect. Biol.* 5, a010363.
- Haber, J.E. (2015). Deciphering the DNA Damage Response. *Cell* 162, 1183–1185.
- Haffner, M.C., Aryee, M.J., Toubajji, A., Esopi, D.M., Albadine, R., Gurel, B., Isaacs, W.B., Bova, G.S., Liu, W., Xu, J., et al. (2010). Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. *Nat. Genet.* 42, 668–675.
- Halazonetis, T.D., Gorgoulis, V.G., and Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. *Science* 319, 1352–1355.
- Haradhvala, N.J., Polak, P., Stojanov, P., Covington, K.R., Shinbrot, E., Hess, J.M., Rheinbay, E., Kim, J., Maruvka, Y.E., Braunstein, L.Z., et al. (2016). Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell* 164, 538–549.
- Harman, D. (1956). Aging: a theory based on free radical and radiation chemistry. *J. Gerontol.* 11, 298–300.
- Helleday, T., Eshtad, S., and Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* 15, 585–598.
- Ju, B.G., Lunyak, V.V., Perissi, V., Garcia-Bassets, I., Rose, D.W., Glass, C.K., and Rosenfeld, M.G. (2006). A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science* 312, 1798–1802.
- Kantidakis, T., Saponaro, M., Mitter, R., Horswell, S., Kranz, A., Boeing, S., Aygün, O., Kelly, G.P., Matthews, N., Stewart, A., et al. (2016). Mutation of cancer driver MLL2 results in transcription stress and genome instability. *Genes Dev.* 30, 408–420.
- Kanu, N., Cerone, M.A., Goh, G., Zalmas, L.P., Bartkova, J., Dietzen, M., McGranahan, N., Rogers, R., Law, E.K., Gromova, I., et al. (2016). DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. *Genome Biol.* 17, 185.
- Kennedy, S.R., Salk, J.J., Schmitt, M.W., and Loeb, L.A. (2013). Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. *PLoS Genet.* 9, e1003794.
- Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* 147, 95–106.
- Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424.
- Kunkel, T.A. (2015). Celebrating DNA's Repair Crew. *Cell* 163, 1301–1303.
- Law, E.K., Sieuwerts, A.M., LaPara, K., Leonard, B., Starrett, G.J., Molan, A.M., Temiz, N.A., Vogel, R.I., Meijer-van Gelder, M.E., Sweep, F.C., et al. (2016). The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci. Adv.* 2, e1601737.
- Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501.
- Lee, J.E., Wang, C., Xu, S., Cho, Y.W., Wang, L., Feng, X., Baldrige, A., Sartorelli, V., Zhuang, L., Peng, W., and Ge, K. (2013). H3K4 mono- and dimethyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife* 2, e01503.
- Lensing, S.V., Marsico, G., Hänsel-Hertsch, R., Lam, E.Y., Tannahill, D., and Balasubramanian, S. (2016). DSBcapture: in situ capture and sequencing of DNA breaks. *Nat. Methods* 13, 855–857.
- Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K., and Rosenfeld, M.G. (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* 139, 1069–1083.
- Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* 362, 709–715.
- Lindahl, T., and Barnes, D.E. (2000). Repair of endogenous DNA damage. *Cold Spring Harb. Symp. Quant. Biol.* 65, 127–133.
- Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. *Biochemistry* 11, 3610–3618.
- Lippert, M.J., Kim, N., Cho, J.E., Larson, R.P., Schoenly, N.E., O'Shea, S.H., and Jinks-Robertson, S. (2011). Role for topoisomerase 1 in transcription-associated mutagenesis in yeast. *Proc. Natl. Acad. Sci. USA* 108, 698–703.
- Liu, M., and Schatz, D.G. (2009). Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol.* 30, 173–181.
- Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 457, 841–845.
- Lodato, M.A., Woodworth, M.B., Lee, S., Evrony, G.D., Mehta, B.K., Karger, A., Lee, S., Chittenden, T.W., D'Gama, A.M., Cai, X., et al. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* 350, 94–98.
- Loeb, L.A., Springgate, C.F., and Battula, N. (1974). Errors in DNA replication as a basis of malignant changes. *Cancer Res.* 34, 2311–2321.

- Lord, C.J., Tutt, A.N., and Ashworth, A. (2015). Synthetic lethality and cancer therapy: lessons learned from the development of PARP inhibitors. *Annu. Rev. Med.* 66, 455–470.
- Madabhushi, R., Gao, F., Pfening, A.R., Pan, L., Yamakawa, S., Seo, J., Rueda, R., Phan, T.X., Yamakawa, H., Pao, P.C., et al. (2015). Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* 161, 1592–1605.
- McKinnon, P.J. (2013). Maintaining genome stability in the nervous system. *Nat. Neurosci.* 16, 1523–1529.
- Meng, F.L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberg, D., et al. (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159, 1538–1548.
- Morganella, S., Alexandrov, L.B., Glodzik, D., Zou, X., Davies, H., Staaf, J., Sieuwerts, A.M., Brinkman, A.B., Martin, S., Ramakrishna, M., et al. (2016). The topography of mutational processes in breast cancer genomes. *Nat. Commun.* 7, 11383.
- Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R., et al. (2012). c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* 151, 68–79.
- Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell* 149, 979–993.
- Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54.
- Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353.
- Notta, F., Chan-Seng-Yue, M., Lemire, M., Li, Y., Wilson, G.W., Connor, A.A., Denroche, R.E., Liang, S.B., Brown, A.M., Kim, J.C., et al. (2016). A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* 538, 378–382.
- Nowell, P.C. (1976). The clonal evolution of tumor cell populations. *Science* 194, 23–28.
- Olivier, M., Hollstein, M., and Hainaut, P. (2010). TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb. Perspect. Biol.* 2, a001008.
- Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S., and Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* 533, 125–129.
- Perera, D., Poulos, R.C., Shah, A., Beck, D., Pimanda, J.E., and Wong, J.W. (2016). Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. *Nature* 532, 259–263.
- Perillo, B., Ombra, M.N., Bertoni, A., Cuozzo, C., Sacchetti, S., Sasso, A., Chiariotti, L., Malorni, A., Abbondanza, C., and Avvedimento, E.V. (2008). DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. *Science* 319, 202–206.
- Periyasamy, M., Patel, H., Lai, C.F., Nguyen, V.T., Nevedomskaya, E., Harrod, A., Russell, R., Remenyi, J., Ochocka, A.M., Thomas, R.S., et al. (2015). APOBEC3B-Mediated Cytidine Deamination Is Required for Estrogen Receptor Action in Breast Cancer. *Cell Rep.* 13, 108–121.
- Pommier, Y., Sun, Y., Huang, S.N., and Nitiss, J.L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* 17, 703–721.
- Pritchard, C.C., Mateo, J., Walsh, M.F., De Sarkar, N., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R., et al. (2016). Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N. Engl. J. Med.* 375, 443–453.
- Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K.R., Resch, W., Liang, G., Tang, Z., Mathé, E., Benner, C., et al. (2014). B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* 159, 1524–1537.
- Ray Chaudhuri, A., Callen, E., Ding, X., Gogola, E., Duarte, A.A., Lee, J.E., Wong, N., Lafarga, V., Calvo, J.A., Panzarino, N.J., et al. (2016). Replication fork stability confers chemoresistance in BRCA-deficient cells. *Nature* 535, 382–387.
- Rialdi, A., Campisi, L., Zhao, N., Lagda, A.C., Pietzsch, C., Ho, J.S., Martinez-Gil, L., Fenouil, R., Chen, X., Edwards, M., et al. (2016). Topoisomerase 1 inhibition suppresses inflammatory genes and protects from death by inflammation. *Science* 352, aad7993.
- Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A., and López-Bigas, N. (2016). Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* 532, 264–267.
- Sankar, T.S., Wastuwidyaningtyas, B.D., Dong, Y., Lewis, S.A., and Wang, J.D. (2016). The nature of mutations induced by replication–transcription collisions. *Nature* 535, 178–181.
- Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G.P., Heron, M., Williams, H., Söding, J., Stewart, A., and Svejstrup, J.Q. (2014). RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* 157, 1037–1049.
- Sarni, D., and Kerem, B. (2016). The complex nature of fragile site plasticity and its importance in cancer. *Curr. Opin. Cell Biol.* 40, 131–136.
- Scarano, E., Iaccarino, M., Grippo, P., and Parisi, E. (1967). The heterogeneity of thymine methyl group origin in DNA pyrimidine isostichs of developing sea urchin embryos. *Proc. Natl. Acad. Sci. USA* 57, 1394–1400.
- Schumacher, T.N., and Schreiber, R.D. (2015). Neoantigens in cancer immunotherapy. *Science* 348, 69–74.
- Schuster-Böckler, B., and Lehner, B. (2012). Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* 488, 504–507.
- Shlien, A., Campbell, B.B., de Borja, R., Alexandrov, L.B., Merico, D., Wedge, D., Van Loo, P., Tarpey, P.S., Coupland, P., Behjati, S., et al.; Biallelic Mismatch Repair Deficiency Consortium (2015). Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermuted cancers. *Nat. Genet.* 47, 257–262.
- Sima, J., and Gilbert, D.M. (2014). Complex correlations: replication timing and mutational landscapes during cancer and genome evolution. *Curr. Opin. Genet. Dev.* 25, 93–100.
- Stork, C.T., Bocek, M., Crossley, M.P., Sollier, J., Sanz, L.A., Chédin, F., Swigut, T., and Cimprich, K.A. (2016). Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *eLife* 5.
- Supek, F., and Lehner, B. (2015). Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* 521, 81–84.
- Supek, F., Lehner, B., Hajkova, P., and Warnecke, T. (2014). Hydroxymethylated cytosines are associated with elevated C to G transversion rates. *PLoS Genet.* 10, e1004585.
- Swanton, C., McGranahan, N., Starrett, G.J., and Harris, R.S. (2015). APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov.* 5, 704–712.
- Takahashi, T., Burguiere-Slezak, G., Van der Kemp, P.A., and Boiteux, S. (2011). Topoisomerase 1 provokes the formation of short deletions in repeated sequences upon high transcription in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 108, 692–697.
- Tomasetti, C., and Vogelstein, B. (2015). Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 347, 78–81.
- Tsouroula, K., Furst, A., Rogier, M., Heyer, V., Maglott-Roth, A., Ferrand, A., Reina-San-Martin, B., and Soutoglou, E. (2016). Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Mol. Cell* 63, 293–305.

- Waters, L.S., and Walker, G.C. (2006). The critical mutagenic translesion DNA polymerase Rev1 is highly expressed during G(2)/M phase rather than S phase. *Proc. Natl. Acad. Sci. USA* 103, 8971–8976.
- Wei, P.C., Chang, A.N., Kao, J., Du, Z., Meyers, R.M., Alt, F.W., and Schwer, B. (2016). Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell* 164, 644–655.
- Williamson, L.M., and Lees-Miller, S.P. (2011). Estrogen receptor α -mediated transcription induces cell cycle-dependent DNA double-strand breaks. *Carcinogenesis* 32, 279–285.
- Wong, R.H., Chang, I., Hudak, C.S., Hyun, S., Kwan, H.Y., and Sul, H.S. (2009). A role of DNA-PK for the metabolic gene regulation in response to insulin. *Cell* 136, 1056–1072.
- Wu, S., Powers, S., Zhu, W., and Hannun, Y.A. (2016). Substantial contribution of extrinsic risk factors to cancer development. *Nature* 529, 43–47.
- Zhang, C.Z., Leibowitz, M.L., and Pellman, D. (2013). Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes Dev.* 27, 2513–2530.
- Zhang, C.Z., Spektor, A., Cornils, H., Francis, J.M., Jackson, E.K., Liu, S., Meyerson, M., and Pellman, D. (2015). Chromothripsis from DNA damage in micronuclei. *Nature* 522, 179–184.
- Zheng, C.L., Wang, N.J., Chung, J., Moslehi, H., Sanborn, J.Z., Hur, J.S., Collisson, E.A., Vemula, S.S., Naujokas, A., Chiotti, K.E., et al. (2014). Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. *Cell Rep.* 9, 1228–1234.