

RESEARCH

Open Access



# Mutational signature distribution varies with DNA replication timing and strand asymmetry

Marketa Tomkova<sup>1</sup>, Jakub Tomek<sup>2</sup>, Skirmantas Kriaucionis<sup>1</sup> and Benjamin Schuster-Böckler<sup>1\*</sup> 

## Abstract

**Background:** DNA replication plays an important role in mutagenesis, yet little is known about how it interacts with other mutagenic processes. Here, we use somatic mutation signatures—each representing a mutagenic process—derived from 3056 patients spanning 19 cancer types to quantify the strand asymmetry of mutational signatures around replication origins and between early and late replicating regions.

**Results:** We observe that most of the detected mutational signatures are significantly correlated with the timing or direction of DNA replication. The properties of these associations are distinct for different signatures and shed new light on several mutagenic processes. For example, our results suggest that oxidative damage to the nucleotide pool substantially contributes to the mutational landscape of esophageal adenocarcinoma.

**Conclusions:** Together, our results indicate an interaction between DNA replication, the associated damage repair, and most mutagenic processes.

**Keywords:** Mutagenesis, DNA replication, DNA repair

## Background

Understanding the mechanisms of mutagenesis in cancer is important for the prevention and treatment of the disease [1, 2]. Mounting evidence suggests replication itself contributes to cancer risk [3]. Copying of DNA is intrinsically asymmetrical, with leading and lagging strands being processed by distinct sets of enzymes [4], and different genomic regions replicating at defined times during S phase [5]. Previous analyses have focused either on the genome-wide distribution of mutation rate or on the strand specificity of individual base changes. These studies revealed that the average mutation frequency is increased in late-replicating regions [6, 7], and that the asymmetric synthesis of DNA during replication leads to strand-specific frequencies of base changes [8–11]. However, the extent to which DNA replication influences distinct mutational mechanisms, with their manifold possible causes, remains incompletely understood.

Mutational signatures have been established as a powerful approach to quantify the presence of distinct mutational mechanisms in cancer [12]. A mutational signature is a unique combination of the frequencies of all base-pair mutation types (C:G > A:T, T:A > G:C, etc.) and their flanking nucleotides. Since it is usually not known which base in a pair was the source of a mutation, the convention is to annotate mutations from the pyrimidine (C > A, T > A, etc.), leading to 96 possible combinations of mutation types and neighboring bases. Non-negative matrix factorization is used to extract mutational signatures from somatic mutations in cancer samples [12]. This approach has the important advantage of being able to distinguish between processes that have the same major mutation type (such as C > T transitions) but differ in their sequence context. We built upon this feature of mutational signatures and developed a computational framework to identify the replication-strand-specific impact of distinct mutational processes. Using this system, we quantified the replication strand and timing bias of mutational signatures across 19 cancer types. We show that replication affects the distribution of nearly all mutational signatures across the genome, including those that represent chemical

\* Correspondence: [benjamin.schuster-boeckler@ludwig.ox.ac.uk](mailto:benjamin.schuster-boeckler@ludwig.ox.ac.uk)

<sup>1</sup>Ludwig Cancer Research Oxford, University of Oxford, Old Road Campus Research Building, Oxford OX3 7DQ, UK

Full list of author information is available at the end of the article



mutagens. The unique strand asymmetry and replication timing profile of different signatures reveal novel aspects of the underlying mechanism. For example, we discovered a strong lagging strand bias of T > G mutations in esophageal adenocarcinoma, suggesting an involvement of oxidative damage to the nucleotide pool in the etiology of the disease. Together, our results highlight the critical role of DNA replication and the associated repair in the accumulation of somatic mutations.

## Results and discussion

### Replication bias of mutational signatures

DNA replication in eukaryotic cells is initiated around replication origins (ORI), from where it proceeds in both directions, synthesizing the leading strand continuously and the lagging strand discontinuously (Fig. 1a). We used two independent data sets to describe replication direction relative to the reference sequence, one derived from high-resolution replication timing data [11] and the other from direct detection of ORIs by short nascent strand sequencing (SNS-seq) [13], corrected for technical artifacts [14] (see “Methods”). The former provides information for more genomic loci, while the latter is of higher resolution. As a third measure of DNA replication, we compared regions replicating early during S phase to regions replicating late [11]. We calculated *strand-specific* signatures [15] that add strand information to each mutation type, based on the direction of DNA replication [11] (Fig. 1b). We clustered the strand-specific signatures and further condensed them into *directional signatures* consisting of 96 mutation types, each assigned either “leading” or “lagging” direction depending on the frequency in the strand-specific signature (Fig. 1c; see “Methods”). These directional signatures can be used to separately compute the presence of the signature on the leading and lagging strands in individual samples, which is analogous to what is called the *exposure* to the signature in a sample [16], representing the genome-scale normalized contribution of mutations to the signature (Fig. 1d). Depending on whether the strand bias matches the consensus of the directional signature, the exposure can be *matching* or *inverse*. The latter can occur if the strand bias of a signature in a subset of samples does not match the bias observed in the samples that most strongly contributed to the definition of the signature. We applied this novel algorithm to somatic mutations detected in whole-genome sequencing of 3056 tumor samples from 19 cancer types (Additional file 1: Table S1). We excluded protein-coding genes from the analysis in order to prevent potential confounding of the results by transcription strand asymmetry [11, 12] or selection. Samples with microsatellite instability (MSI) and POLE mutations were treated as separate groups, since they are associated with specific mutational processes. In total, we detected 25 mutational signatures that each corresponded to one of the COSMIC

signatures (<http://cancer.sanger.ac.uk/cosmic/signatures>) and four novel signatures, which were primarily found in samples that had not been previously used for signature extraction (myeloid blood, skin, MSI, and ovarian cancers; Additional file 2: Figures S1–S5).

In total, 21 out of 29 signatures exhibited significant replication strand asymmetry, 23 were significantly correlated with replication timing, and 27 were significant in at least one of these two metrics (signtest  $p < 0.05$ , with Benjamini-Hochberg correction; Fig. 2, Additional file 2: Figures S7–S8, S13–S24, Additional file 3: Table S2, Additional file 4: Table S3). Such widespread replication bias across the mutational landscape is surprising, considering that previous reports documented strand bias for only a few mutational processes, such as activity of the APOBEC class of enzymes that selectively edit exposed single-stranded cytosines on the lagging strand [11, 15, 17–19]. Including protein coding genes did not qualitatively change the results (Additional file 2: Figure S9a–c), nor did the exclusion of non-coding in addition to protein-coding genes (Additional file 2: Figure S9d–f). Similarly, using SNS-seq data to determine replication strand direction leads to highly similar findings (Additional file 2: Figure S9g–i). Furthermore, we validated that the strand asymmetry and correlation with replication timing is lost when using random genomic loci as replication origins or replication timing domains (Additional file 2: Figures S10 and S11).

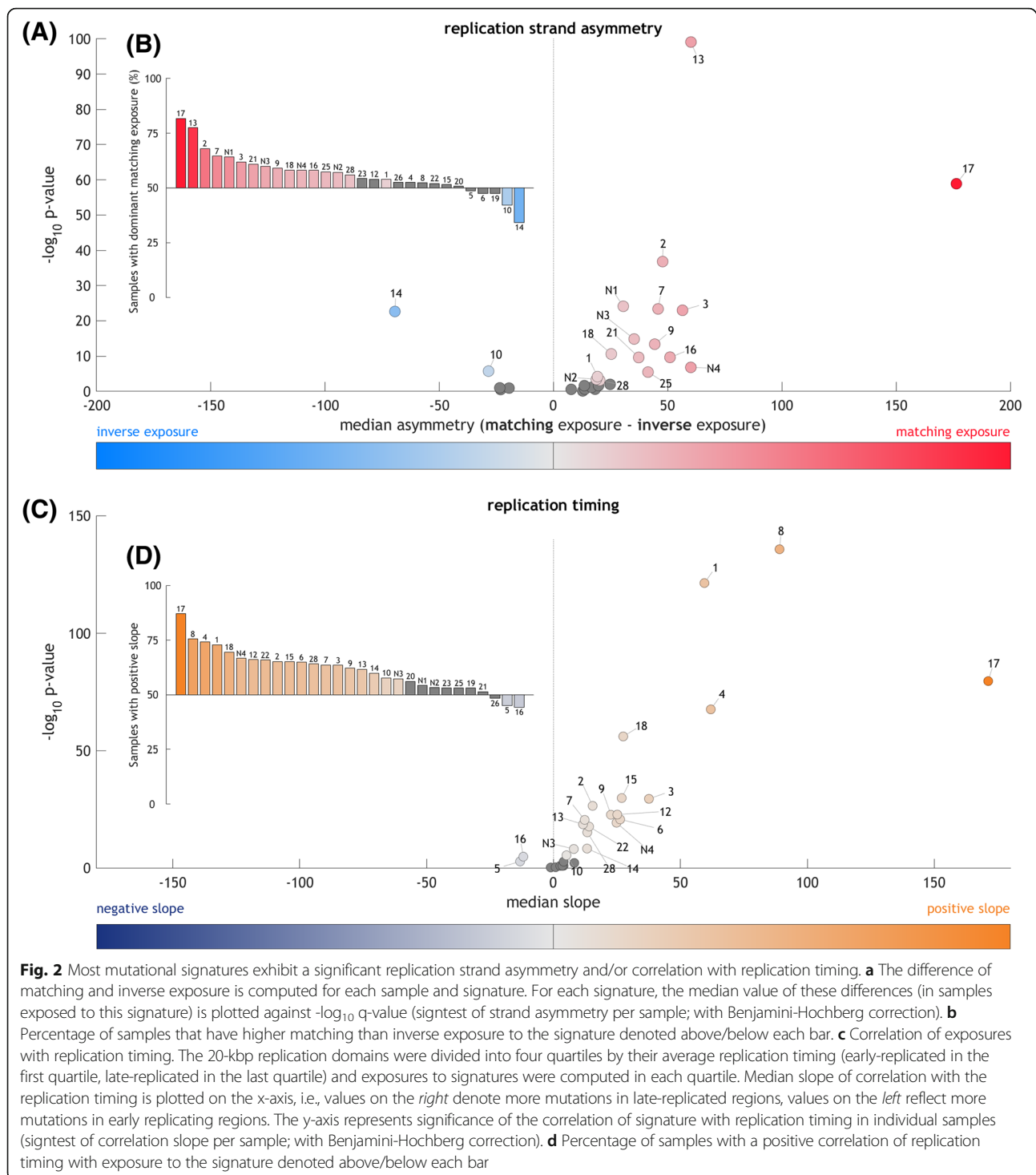
Our observations confirm that both APOBEC signatures (2 and 13) exhibit clear strand asymmetry, with signature 13 being the most significantly asymmetric signature (q-value  $2e^{-98}$ ). In breast cancer samples, we observed differences in these signatures with respect to replication timing: signature 2, but not signature 13, shows a significant enrichment in late replicating regions (Fig. 3), which is consistent with previous reports [15]. These results validate that our approach is able to correctly identify strand and timing asymmetries of mutagenic processes. Consequently, we next tried to interpret the replication biases we observed in other mutational signatures.

Interestingly, the lack of correlation with replication timing in signature 13 seems to be specific to breast cancers, as other cancer types (such as lung squamous, esophageal adenocarcinoma, and pancreas) show a significant correlation (Additional file 2: Figure S12). Nevertheless, the replication strand asymmetry of both signatures 2 and 13 is significantly present in all these tissues.

### Processes directly involving DNA replication or repair

Amongst the better understood mutational mechanisms, several involve replicative processes and DNA repair, such as mismatch-repair deficiency (MMR) [20] or mutations in the proofreading domain of Pol  $\epsilon$  (“POLE-MUT samples”) [8, 21]. We first analyzed the signatures representing these mechanisms, since they can be directly attributed to a

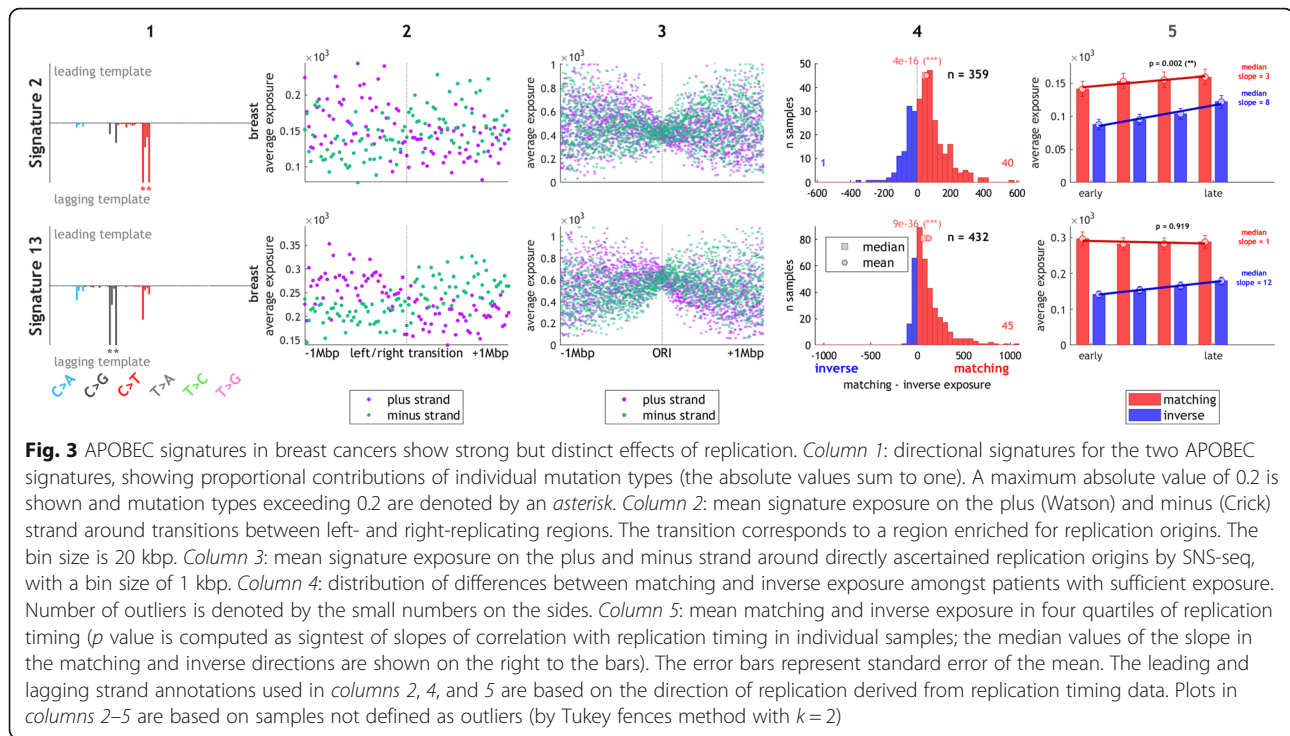




known molecular process. All five signatures previously associated with MMR and the novel MSI-linked signature N4 exhibit a clear trend of replication strand asymmetry (significant in signature (sig.) 6 and N4 in MSI), generally with enrichment of C > T mutations on the leading strand template and C > A and T > C mutations on the lagging strand template (Fig. 4, Additional file 2: Figure S13), in line

with the previously suggested role of MMR to balance mutational asymmetries generated by DNA polymerases during replication [9, 11].

It has previously been proposed that the correlation of overall mutation rate with replication timing (as shown in Fig. 2b) is a direct result of the activity of MMR [22]. In contrast, when splitting the signal by mutational signatures,



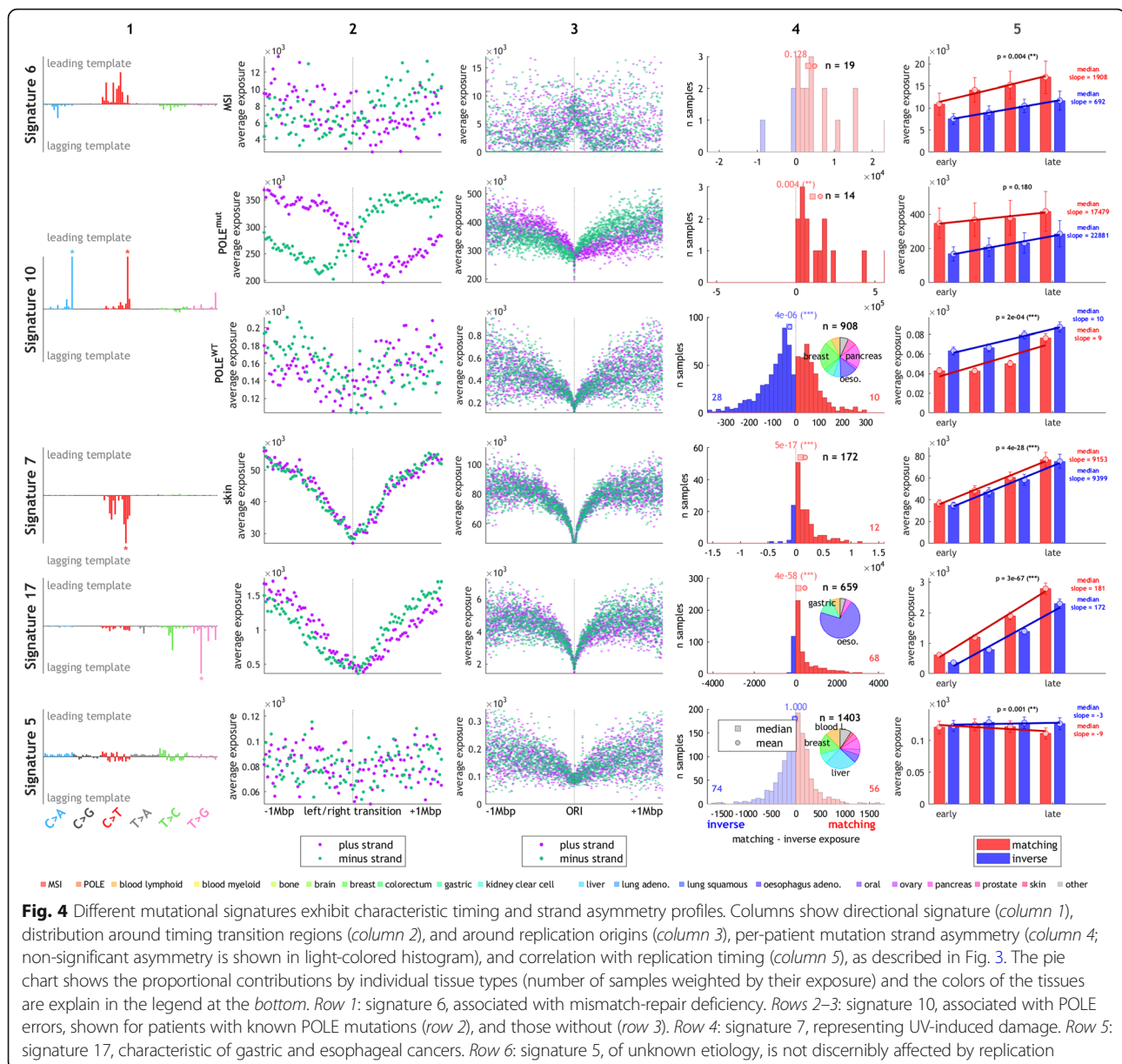
we observed a more complex relationship. Some MMR signatures in MMR-deficient patients do not significantly correlate with replication timing (sig. 15, 21, 26) or do so only in one direction of replication (such as a negative correlation in the leading direction in sig. 20), whereas others show a clear increase in the late replicated regions (sig. 6 and N4; Additional file 2: Figure S13). We next explored the effect of MMR on correlation with replication timing (irrespective of the replication strand) in all signatures with exposure  $> 10$  in at least four MSI samples. Four signatures exhibited a significant correlation with replication timing in MSI samples (a positive correlation in sig. 6, 18, and N4; negative correlation in sig. 20; Additional file 2: Figure S25). Interestingly, in all these four signatures, the slope of the correlation was significantly steeper in MSI than MSS samples (Additional file 2: Figure S26). Furthermore, some of the other signatures significantly correlated with replication timing in MSS (e.g., sig. 17 and 8) also showed a weak but consistent correlation in MSI. In other signatures (e.g., sig. 1) the correlation is lost. Altogether, these results indicate that MMR is only one of several factors influencing mutagenesis in a timing-dependent manner.

Unexpectedly, two MMR signatures (sig. 6 and N4) showed increased exposures around ORIs (Fig. 4, Additional file 2: Figures S13, S14, S27). Based on experiments in yeast, it has been suggested that MMR is involved in balancing the differences in fidelity of the leading and lagging polymerases [9], in particular repairing errors made by Pol  $\alpha$  [9], which primes the leading strand at ORIs and each lagging strand Okazaki

fragment [23] and lacks intrinsic proofreading capabilities [24]. It has been recently shown that error-prone Pol  $\alpha$ -synthesized DNA is retained in vivo, causing an increase of mutations on the lagging strand [10]. Since regions around ORIs have a higher density of Pol  $\alpha$ -synthesized DNA (as discussed, e.g., in [25]), it is possible that increased exposure to signatures 6 and N4 around ORIs is caused by incomplete repair of Pol  $\alpha$ -induced errors. The most common Pol  $\alpha$ -induced mismatches normally repaired by MMR are G-dT and C-dT, leading to C  $>$  T mutations on the leading strand and C  $>$  A mutations on the lagging strand [26], matching our observations in the MMR-linked signatures. Notably, we also detected weaker but still significant exposure to MMR signatures in samples with seemingly intact mismatch repair (Additional file 2: Figure S14). Replication strand asymmetry in these samples was substantially smaller, but the higher exposure to signatures 6 and N4 around ORIs remained (Additional file 2: Figure S27). These findings are compatible with a model in which one of the functions of mismatch repair is to balance the effect of mis-incorporation of nucleotides by Pol  $\alpha$ . Signatures 6, N4, and possibly 26 appear to reflect this mechanism, while the other MMR signatures might be a result of unrelated functions of MMR, such as its involvement in balancing errors made by other polymerases, e.g., Pol  $\delta$ .

POLE-MUT samples were previously reported to be “ultra-hypermutated” with excessive C  $>$  A and C  $>$  T mutations on the leading strand [8, 11, 21]. Mutational





signature 10 has been associated with mutations in the proofreading domain of Pol  $\epsilon$ , the main leading strand polymerase [23, 27]. We noticed that mutational signature 14 is also strongly associated with POLE mutations in the data sets by Shlien et al. [21], Alexandrov et al. [12], and in The Cancer Genome Atlas (TCGA). Since then, Andrianova et al. confirmed this observation, showing that signature 14 is enriched in POLE-MUT samples with mismatch repair deficiency [28]. As expected, we observed very strong strand asymmetry for these two signatures in all POLE-MUT samples, with an increase of C > A, C > T, and T > G mutations on the leading strand (Fig. 4, Additional file 2: Figure S15). As with MMR signatures, we also found weak but significant evidence of signature

10 and 14 in samples without Pol  $\epsilon$  defects (POLE-WT). Strikingly, however, in these samples the strand asymmetry was in the inverse orientation compared to the POLE-MUT samples, i.e., more C > A, C > T, and T > G mutations on the lagging strand (Fig. 4, Additional file 2: Figure S16). Conversely, we detected the presence of two signatures of unknown etiology, signatures 18 and 28, in POLE-MUT samples, but in the inverse orientation compared to POLE-WT samples. We performed two additional analyses in order to validate that this is not an artifact of spurious/wrong associations in the signature exposures decomposition. First, we removed exposures not robust to perturbations (see “Methods”) and confirmed that all four signatures (10, 14, 18, and 28)

remained significantly strand asymmetric in both POLE-MUT and POLE-WT samples (Additional file 2: Figures S17–S18). Second, we directly compared the frequencies of the most prominent mutation types for each of the four signatures (sig. 10, 14, 18, and 28) in POLE-MUT and POLE-WT samples on the leading and lagging strands. The inverse strand preference observed in the signatures was also detected for individual mutation types. For example, the frequency of mutations in TCT > A, TCG > T, and TTT > G, the three major components of signature 10, is higher on the lagging strand than on the leading strand in POLE-WT samples, whereas it is higher on the leading strand in POLE-MUT (Additional file 2: Figures S28–S31). We therefore hypothesize that POLE-linked signatures are originally caused by a process that affects both strands and, under normal circumstances, is slightly enriched on the lagging strand. This could be caused by certain types of DNA lesions which under normal circumstances are less accurately replicated when on the template of the lagging strand (e.g., due to a lower fidelity of Pol  $\delta$  or Pol  $\alpha$  compared to wild-type Pol  $\epsilon$  when replicating these lesions). In POLE-MUT samples the lack of replication-associated proofreading would then lead to a strong relative increase in these mutations on the leading strand, explaining the flipped orientation of signatures.

Apart from replication strand asymmetry, we also observed a significant correlation with replication timing in signatures 10, 14, 18, and 28 in POLE-WT samples (Additional file 2: Figure S16). The correlation was significant only in signature 28 in POLE-MUT samples, but at least 75% of samples showed a positive slope of the correlation also in signatures 10, 14, and 18, and the slope was significantly increased in POLE-MUT compared to MSS POLE-WT samples in signatures 10, 18, and 28 (Additional file 2: Figures S15, S16, S32, S33). Future studies with larger sample size are needed to evaluate the effect of POLE mutations and MMR on the replication timing characteristics on this group of mutational signatures.

### Signatures linked to environmental mutagens

We next focused on signatures that have not previously been reported to be connected to replication, or for which the causal mechanism is unknown. Our data show a link between DNA replication and exogenous mutagens such as UV light (signature 7), tobacco smoke (signature 4), or aristolochic acid (AA; signature 22) [29]. In these signatures, we observed marked correlation with replication timing (Fig. 4, Additional file 2: Figures S19 and S20). Higher mutation frequency late in replication has been observed in mouse embryonic fibroblast (MEFs) treated with AA or Benzo[a]pyrene (B[a]P; a mutagen in tobacco smoke) [30]. This increased mutagenicity might be attributed to different DNA damage tolerance pathways being

active during early and late replication. Regions replicated early in S-phase are thought to prefer high-fidelity template switching, whereas regions replicated late are more likely to require translesion synthesis (TLS), which has a higher error rate [31–37]. This is consistent with the observation in yeast that a disruption of TLS leads to decreased mutation frequency in late-replicating regions and therefore a more even distribution of mutation frequency between early and late-replicating regions [32]. In particular, TLS has been observed to increase in activity and mutagenicity later in the cell cycle when replicating DNA damaged by B[a]P [38]. Alternatively, differences in chromatin accessibility could be responsible for the decreased mutagenicity in early-replicated regions. Open chromatin is, on average, replicated earlier and is also more accessible to repair enzymes, which could contribute to the decreased mutation frequency in early-replicating regions [39].

We also observed weak but significant replication strand asymmetry in the mutagen-induced signatures in the tissues associated with the respective mutagen (Additional file 2: Figure S19). Signature 4 has a significant strand asymmetry in lung cancers, similar to signature 7 in skin cancers. Signature 22 in kidney cancers has a small sample size but shows the same trend. This matches a previously observed lower efficiency of bypass of DNA damage on the lagging strand [40] and strong mutational strand asymmetry in cells lacking Pol  $\eta$ , the main TLS polymerase responsible for the replication of UV-induced photolesions [41]. Altogether, our data highlight the importance of replication in converting DNA damage into actual mutations and suggest that bypass of DNA damage occurring on the lagging template results in detectably lower fidelity on this strand.

Signature 17 had the largest median strand asymmetry (sixth largest  $\log_2$  ratio of the two strands) and also was one of the signatures with the strongest correlations with replication timing (Figs. 2 and 4, Additional file 2: Figure S7). The mutational process causing this signature is unclear. We noted that the timing asymmetry and exposure distribution around ORIs to signature 17 closely resembled that of signatures 4 and 7, suggesting a possible link to DNA damage. Signature 17 is most prominent in gastric cancers and esophageal adenocarcinoma (EAC), where it appears early during disease development [42], and it is also present in Barrett's esophagus (BE), a precursor to EAC [43]. Due to the importance of gastro-esophageal and duodeno-gastric reflux in the development of BE and EAC [44–46] and the resulting oxidative stress [47–50], it has been speculated that oxidative damage could cause the mutation patterns characteristic for signature 17 [51, 52]. Increased oxidative damage to guanine has been reported in the epithelial cells of dysplastic BE as well as after incubation of BE tissue with a cocktail mimicking bile reflux [50]. Oxidative stress affects bases in not only the DNA but also the nucleotide pool,

such as the oxidation of dGTP to 8-oxo-dGTP. This oxidized dGTP derivative has been shown to induce T > G transversions [53–55] through incorporation by TLS polymerases into DNA opposite A on the template strand [56]. In contrast, oxidation of guanine in the DNA produces 8-oxo-G, which has been shown to result in C > A mutations when paired with adenine during replication [57]. These C > A mutations are normally prevented by DNA glycosylases in the base excision repair pathway, such as MUTYH and OGG1, which repair 8-oxo-G:A pairs to G:C. However, if an 8-oxo-G:A mismatch resulted from incorporation of 8-oxo-dGTP in the de novo synthesized strand, the “repair” to G:C would actually lead to a T > G mutation [57]. Consequently, depletion of MUTYH led to an increase of C > A mutations [57, 58] but a decrease of T > G mutations induced by 8-oxo-dGTP [59]. Importantly, the mismatch of 8-oxo-G and A has been shown in yeast to be more efficiently repaired into G:C when 8-oxo-G is on the lagging strand template [60, 61], resulting in an enrichment of T > G mutations on the lagging strand template if the 8-oxoG:A mismatch originated from incorporation of 8-oxo-dGTP opposite A. Our data show strong lagging-strand bias of T > G mutations and overall higher exposure to signature 17 on the lagging strand, supporting the hypothesis that signature 17 is a by-product of oxidative damage.

#### DNA methylation-linked mutagenesis

A small but significant strand asymmetry was detected in signature 1 (Additional file 2: Figure S22). This observation is difficult to explain with spontaneous deamination of 5-methylcytosine, the assumed etiology of signature 1. However, the observation would be in line with a previously hypothesized model in which Pol  $\epsilon$  has decreased fidelity of replicating 5-methylcytosine, causing an enrichment of C > T mutations in methylated cytosines on the leading strand, especially in samples with deficiency in MMR or Pol  $\epsilon$  proofreading [62]. Our analysis shows that signature 1 is slightly enriched on the leading strand, even in samples proficient for post-replicative proofreading and repair. Moreover, we observed that signature 1 is significantly correlated with replication timing in MSS, but not in MSI samples (Additional file 2: Figure S26), in line with the possibility that MMR repairs C > T errors in a CpG context introduced by Pol  $\epsilon$ , as MMR is thought to be active primarily in the early-replicated regions [22].

#### Conclusions

Our findings demonstrate how the relationship between mutational signatures and DNA replication can help to illuminate the mechanisms underlying several currently unexplained mutational processes, as exemplified by signature 17 in esophageal cancer. Crucially, our computational analysis produces testable hypotheses which we anticipate

to be experimentally validated in the future; for instance, that bypass of external-mutagen-induced DNA damage (such as UV light, tobacco smoking, and aristolochic acid) is more error prone during synthesis of the lagging strand, or that oxidative damage to the dNTP pool contributes to the etiology of signature 17. Our results also add a new perspective to the recent debate regarding the correlation of tissue-specific cell division rates with cancer risk [3]. It has been argued that this correlation is primarily attributable to “bad luck” in the form of random errors that are introduced during replication by DNA polymerases. Critics of this theory have pointed out that the range of mutational signatures observed in cancer samples makes a purely replication-driven etiology of cancer mutations unlikely [63, 64]. Our analysis at least partially reconciles the two arguments, showing that most mutational signatures are themselves affected by DNA replication, including signatures linked to environmental mutagens. The presence of mutational signatures on the one hand and a strong relationship between replication and the risk of cancer on the other therefore need not be mutually exclusive. In conclusion, our results provide evidence that DNA replication interacts with most processes that introduce mutations in the genome, suggesting that differences amongst DNA polymerases and post-replicative repair enzymes might play a larger part in the accumulation of mutations than previously appreciated.

#### Methods

##### Somatic mutations

Cancer somatic mutations in 3056 whole-genome sequencing samples (Additional file 1: Table S1, Additional file 5: Table S4) were obtained from the data portal of TCGA, the data portal of the International Cancer Genome Consortium (ICGC), and previously published data in peer review journals [12, 21, 51, 65, 66]. For TCGA samples, aligned reads of paired tumor and normal samples were downloaded from the Genomic Data Commons Data Portal website under TCGA access request #10140 and somatic variants were called using Strelka (version 1.0.14) [67] with default parameters. The status of POLE-MUT and MSI samples were obtained from the supplementary data of [11, 21, 66].

##### Direction of replication

Left- and right-replicating domains were taken from [11] where replication timing profiles were generated in six lymphoblastoid cell lines [68], valleys and peaks (defined as regions with a slope with a magnitude lower than 250 rtu per Mb) were removed, after which left- and right-replicating domains were defined as timing transition regions with a negative and positive slope, respectively [11]. In the left-replicated regions, the reference strand is used as a template for the leading strand, while the opposite strand is



used as a template for the lagging strand, and vice versa for the right-replicated regions. Each domain (called territory in the original source code and data) is 20 kbp wide and annotated with the direction of replication and with replication timing.

### Excluded regions

The following regions were excluded: regions with low unique mappability of sequencing reads (positions with mean mappability in 100-bp sliding windows below 0.99 from UCSC mappability track: alignability of 50mers, accession number wgEncodeEH000320), gencode protein coding genes, and blacklisted regions defined by Anshul Kundaje [69] (Anshul\_Hg19UltraHighSignalArtifactRegions.bed, Duke\_Hg19SignalRepeatArtifactRegions.bed, and wgEncodeHg19ConsensusSignalArtifactRegions.bed from <http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg19-human/>).

### Mutation frequency analysis

All variants were classified by the pyrimidine of the mutated Watson-Crick base pair (C or T), strand of this base pair (C or T), and the immediate 5' and 3' sequence context into 96 possible mutation types as described by Alexandrov et al. [12]. The frequency of trinucleotides on each strand was computed for each replication domain. Then the mutation frequency of each mutation type in each replication domain on the leading (plus = Watson strand in left-replicating domains; minus = Crick strand in right-replicating domains) and lagging strand (vice versa) was computed for each sample.

### Extraction of mutational signatures

Matlab code [12] was used for extraction of strand-specific mutational signatures. The input data were the mutation counts on the leading and lagging strands (summed from all replicating domains together, but without the excluded regions) in each sample. The 192-elements-long mutational signatures (example in Fig. 1b) were extracted in each cancer type separately (for  $K$  number of signatures between 2 and 7). The best  $K$  with minimal error and maximal stability (minimizing  $\text{error}_K / \max(\text{error}) + (1 - \text{stability}_K)$  and with stability of at least 0.8) was selected for each cancer type. The stability metric (computed as average silhouette width of clusters of signatures computed by Non-negative Matrix Factorisation) represents reproducibility of the model, while the error (computed as the average Frobenius reconstruction error) evaluates the accuracy with which the deciphered mutational signatures and their respective exposures describe the original matrix of mutations. Signatures present in only a small number of samples with very low exposures were excluded ((95th percentile of exposures of this signature)/(Mean total exposure per samples) < 0.2). The remaining signatures were then normalized by the

frequency of trinucleotides in the leading and lagging strands and subsequently multiplied by the frequency of trinucleotides in the genome. This made them comparable with the 30 previously identified whole-genome-based COSMIC signatures (<http://cancer.sanger.ac.uk/cosmic/signatures>). Signatures extracted in each cancer type and COSMIC signatures were all pooled together (with equal values in the leading and lagging parts in the COSMIC signatures) and were clustered using unsupervised hierarchical clustering (with cosine distance and complete linkage). A threshold of 27 signatures was selected to identify clusters of similar signatures. Mis-clustering was avoided by manual examination (and whenever necessary re-assignment) of all signatures in all clusters (Additional file 2: Figure S6). The resulting 29 signatures (representing the detected clusters) contained 25 previously observed (COSMIC) and four new signatures. For the subsequent analysis, the signatures were converted back to 96 values: the 25 COSMIC signatures were used in their original form (i.e., having the same values as on the COSMIC website) and for the four newly identified signatures we used an average of the leading and lagging parts of the 192-elements-long signatures.

### Annotation of signatures with leading and lagging direction

Each signature was annotated with the dominant strand direction (leading vs lagging) in each of the 96 mutation types (Fig. 1c). This was based on the dominant strand direction within the signature's cluster. Types with unclear direction and small values were assigned according to the predominant direction of other trinucleotides of the same mutation group, such as C > T.

### Calculating strand-specific exposures in individual samples

Exposures to leading and lagging parts of the signatures on the leading and lagging strands in individual samples were quantified using non-negative least squares regression using the Matlab function  $e = \text{lsqnonneg}(S, m)$ , where:

$$S = \begin{pmatrix} S_{LD} & S_{LG} \\ S_{LG} & S_{LD} \end{pmatrix}, m = \begin{pmatrix} m_{LD} \\ m_{LG} \end{pmatrix}, e = \begin{pmatrix} e_{\text{matching}} \\ e_{\text{inverse}} \end{pmatrix}.$$

The matrix  $S_{LD}$  has 96 rows and 29 columns and represents the leading parts of the signatures, i.e., the elements of the lagging parts contain zeros in this matrix. Similarly,  $S_{LG}$  has the same size but contains zeros in the leading parts. The vector  $m_{LD}$  of length 96 contains mutations on the leading strand (again normalized by trinucleotides in leading strand/whole genome), and similarly  $m_{LG}$  contains mutations from the lagging strand. Finally,  $\text{lsqnonneg}$  finds a non-negative vector of exposures  $e$  such that it minimizes a function  $|m - S \cdot e|$ . A similar approach has been used in [70] for finding exposures to a given set of

signatures. Our extension includes the strand-specificity of the signatures. The interpretation of the model is that the *matching exposure*  $e_{matching}$  represents exposure of the leading part of the signature on the leading strand and exposure of the lagging part of the signature on the lagging strand, whereas  $e_{inverse}$  represents the two remaining options. It is important to note that the direction of the mutation is relative to the nucleotide in the base pair chosen as the reference, i.e., mutations of a pyrimidine on the leading strand correspond to mutations of a purine on the lagging strand. In order to minimize the number of spurious signature exposures, the least exposed signature was incrementally removed (in both leading and lagging parts) as long as the resulting error did not exceed the original error by 0.5%. The resulting reported values in each sample and signature were the difference (or fold change) of  $e_{matching}$  and  $e_{inverse}$ . In each signature, the sigttest was used to compare matching and inverse exposures across samples with sufficient minimal exposure (at least 10) to the signature. Benjamini-Hochberg correction was applied to correct for multiple testing.

### Replication origins

The left/right transitions of the replication domains represent regions with, on average, higher density of replication origins. In order to get better resolution of the replication origins, and to validate the results using an independent estimates of left- and right-replicating domains, genome-wide maps of human replication origins from SNS-seq by [13] were used. Eight fastq files (HeLa, iPS, hESC, IMR; each with two replicates) were downloaded and mapped to hg19 using bowtie2 (version 2.1.0). To control for the inefficient digestion of  $\lambda$ -exo step of SNS-seq, reads from non-replicating genomic DNA (LexoG0) were used as a control [14]. Peaks were called using “macs callpeak” with parameters `--gsize = hs --bw = 200 --qvalue = 0.05 --mfold 5 50` and LexoG0 mapped reads as a control. Only peaks covered in at least seven of the eight samples were used. We generated 1000 1-kbp bins to the left and right of each origin, as long as they did not reach half the distance to the next origin. We then used these replication direction annotations in the 1-kbp bins to calculate strand-specific exposures in individual samples as above and ascertained that both approaches lead to qualitatively very similar mutational strand asymmetries in individual signatures (Additional file 2: Figure S9).

### Quantification of exposures with respect to replication timing, left/right transitions, and replication origins

Replication domains were divided into four quartiles by their average replication timing. The entire exposure quantification was computed separately in each quartile, or bin around left/right transition or bin around replication origin. In replication timing plots, a linear regression

model (function fitlm in MatLab) was fitted to the mean exposure in each quartile (separately for matching and inverse exposures) and sigttest on slopes of the fits in individual samples was used to evaluate significance of correlation with replication timing across the cohort. Benjamini-Hochberg correction was applied to correct for multiple testing.

### Evaluation of robustness to noise in signature exposures

For each sample, 1000 perturbations of the input mutation frequency vector were performed, adding noise generated from normal distribution with mean 0 and standard deviation as 5% of the original mutation frequency. Exposures to signatures were quantified for each perturbation. Signatures with extremely variable exposures ( $(\text{qtl}75 - x)/x \geq 0.3$  or  $(x - \text{qtl}25)/x \geq 0.3$ ) were removed for this sample and then the exposures were re-computed using only the signatures that passed the filtering.

### Additional files

**Additional file 1: Table S1.** Overview of the used whole-genome sequencing samples. (PDF 211 kb)

**Additional file 2: Figures S1–S33.** Supplementary figures. (PDF 6116 kb)

**Additional file 3: Table S2.** Overview of the strand asymmetry and correlation with replication timing in mutational signatures. (PDF 205 kb)

**Additional file 4: Table S3.** Values of data points. (XLSX 12 kb)

**Additional file 5: Table S4.** List of all samples. (XLSX 1426 kb)

**Additional file 6: Table S5.** Values of the strand-specific mutational signatures. (XLSX 51 kb)

**Additional file 7:** Review history. (DOCX 64 kb)

### Acknowledgements

We thank Dr. Mary Muers for comments on the manuscript.

### Funding

S.K. and B.S.-B. are funded by Ludwig Cancer Research. S.K. received funding from BBSRC grant BB/M001873/1. M.T. and J.T. are funded by EPSRC (EP/F500394/1).

### Availability of data and materials

All data used in this study were downloaded from public repositories. A complete list of whole-genome sequencing data sets and their accession information can be found in Additional file 1: Table S1 and the list of all whole-genome sequencing samples and signature exposures are in Additional file 5: Table S4. The replication direction [11] was taken from <http://software.broadinstitute.org/cancer/cga/AsymTools>, table per\_base\_territories\_20kb.mat. The replication origins measured by SNS-seq [13] were based on the supplementary files of National Center for Biotechnology Information Gene Expression Omnibus accession GSE37757. The strand-specific signatures are in Additional file 6: Table S5. The code is available at <https://bitbucket.org/bsbludwig/replicationasymmetry> [71] and Figshare (<https://figshare.com/s/21174d60d594ddb9b83d>, DOI <https://doi.org/10.6084/m9.figshare.6941456>) [72] under GPL3 license.

### Authors' contributions

BS-B and MT designed the study; MT performed the analysis with contributions from JT; BS-B and MT wrote the manuscript with contributions from SK and JT. All authors read and approved the final manuscript.

### Review history

The review history is available as Additional file 7.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Ludwig Cancer Research Oxford, University of Oxford, Old Road Campus Research Building, Oxford OX3 7DQ, UK. <sup>2</sup>Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford OX1 3PT, UK.

Received: 6 February 2018 Accepted: 16 August 2018

Published online: 10 September 2018

**References**

- Secrier M, Li X, de Silva N, Eldridge MD, Contino G, Bornschein J, et al. Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat Genet.* 2016;2016:1131–41. Available from: <http://www.nature.com/doi/10.1038/ng.3659>.
- Stenzinger A, Pfarr N, Endris V, Penzel R, Jansen L, Wolf T, et al. Mutations in POLE and survival of colorectal cancer patients—link to disease stage and treatment. *Cancer Med.* 2014;3:1527–38.
- Tomasetti C, Vogelstein B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science.* 2015;347:78–81. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25554788>.
- Lujan SA, Williams JS, Kunkel TA. DNA polymerases divide the labor of genome replication. *Trends Cell Biol.* 2016;26:640–54. Available from: <https://doi.org/10.1016/j.tcb.2016.04.012>.
- Fragkos M, Ganier O, Coulombe P, Méchali M. DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol.* 2015;16:360–74. Available from: <https://doi.org/10.1038/nrm4002%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/25999062>.
- Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. Human mutation rate associated with DNA replication timing. *Nat Genet.* 2009;41:393–5.
- Lawrence MS, Stojanov P, Polak P, Kryukov G V, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499:214–218. Available from: <https://doi.org/10.1038/nature12213>.
- Shinbrot E, Henninger EE, Weinhold N, Covington KR, Göksenin AY, Schultz N, et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res.* 2014;1740–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25228659>.
- Lujan SA, Williams JS, Pursell ZF, Abdulovic-Cui AA, Clark AB, Nick McElhinny SA, et al. Mismatch repair balances leading and lagging strand DNA replication Fidelity. *PLoS Genet.* 2012;8:e1003016.
- Reijns MAM, Kemp H, Ding J, de Procé SM, Jackson AP, Taylor MS. Lagging-strand replication shapes the mutational landscape of the genome. *Nature.* 2015;518:502–6. Available from: <http://www.nature.com/nature/journal/v518/n7540/full/nature14183.html>.
- Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, et al. Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. *Cell.* 2016;164:538–49. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0092867415017146>.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature.* 2013;500:415–21. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3776390&tool=pmcentrez&rendertype=abstract>.
- Besnard E, Babled A, Lapasset L, Milhavel O, Parrinello H, Dantec C, et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol.* 2012;19:837–44.
- Foulk MS, Urban JM, Casella C, Gerbi SA. Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Res.* 2015;25:725–35.
- Morganella S, Alexandrov LB, Glodzik D, Zou X, Davies H, Staaf J, et al. The topography of mutational processes in breast cancer genomes. *Nat Commun.* 2016;7:11383. Available from: <http://www.nature.com/doi/10.1038/ncomms11383>.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* 2013;3:246–59. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3588146&tool=pmcentrez&rendertype=abstract>.
- Hoopes JJ, Cortez LM, Mertz TM, Malc EP, Mieczkowski PA, Roberts SA. APOBEC3A and APOBEC3B preferentially deaminate the lagging strand template during DNA replication. *Cell Rep.* 2016;14:1273–82. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S2211124716000425>.
- Green AM, Landry S, Budagyan K, Avgousti DC, Shalhout S, Bhagwat AS, et al. APOBEC3A damages the cellular genome during DNA replication. *Cell Cycle.* 2016;15:998–1008. Available from: <https://doi.org/10.1080/15384101.2016.1152426>.
- Seplyarskiy VB, Soldatov RA, Popadin KY, Antonarakis SE, Bazykin GA, Nikolaev SI. APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. *Genome Res.* 2016;26:174–82.
- Zhao H, Thienpont B, Yesilyurt BT, Moisse M, Reumers J, Coenegrachts L, et al. Mismatch repair deficiency endows tumors with a unique mutation signature and sensitivity to DNA double-strand breaks. *elife.* 2014;3:e02725.
- Shlien A, Campbell BB, de Borja R, Alexandrov LB, Merico D, Wedge D, et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat Genet.* 2015;47:257–62. Available from: <http://www.nature.com/doi/10.1038/ng.3202>.
- Supek F, Lehner B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature.* 2015;521:81–4.
- Stillman B. DNA polymerases at the replication fork in eukaryotes. *Mol Cell.* 2008;30:259–60.
- McCulloch SD, Kunkel TA. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res.* 2008;18:148–61.
- Waisenteiger IS-R, Liston VG, Menezes MR, Kim H-M, Lobachev KS, Stepchenkova EI, et al. Modulation of mutagenesis in eukaryotes by DNA replication fork dynamics and quality of nucleotide pools. *Environ Mol Mutagen.* 2012;53:699–724. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23055184>.
- Nick McElhinny SA, Kissling GE, Kunkel TA. Differential correction of lagging-strand replication errors made by DNA polymerases  $\alpha$  and  $\delta$ . *Proc Natl Acad Sci U S A.* 2010;107:21070–5.
- Georgescu RE, Schauer GD, Yao NY, Langston LD, Yurieva O, Zhang D, et al. Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. *elife.* 2015;2015:1–20.
- Andrianova MA, Bazykin GA, Nikolaev SI, Seplyarskiy VB. Human mismatch repair system balances mutation rates between strands by removing more mismatches from the lagging strand. *Genome Res.* 2017;27:1336–43.
- Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet.* 2014;15:585–98. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24981601>.
- Nik-Zainal S, Kucab JE, Morganella S, Glodzik D, Alexandrov LB, Arlt VM, et al. The genome as a record of environmental exposure. *Mutagenesis.* 2015;30:763–70.
- Waters LS, Walker GC. The critical mutagenic translesion DNA polymerase Rev1 is highly expressed during G(2)/M phase rather than S phase. *Proc Natl Acad Sci U S A.* 2006;103:8971–6.
- Lang GI, Murray AW. Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome Biol Evol.* 2011;3:799–811.
- Karras GI, Fumasoni M, Sienski G, Vanoli F, Branzei D. Article noncanonical role of the 9-1-1 clamp in the error-free DNA damage tolerance pathway. *Mol Cell.* 2013;49:536–46. Available from: <https://doi.org/10.1016/j.molcel.2012.11.016>.
- Gonzalez-Huici V, Szakal B, Urulangodi M, Psakhye I, Castellucci F, Menolfi D, et al. DNA bending facilitates the error-free DNA damage tolerance pathway and upholds genome integrity. *EMBO J.* 2014;33:327–40.
- Bi X. Mechanism of DNA damage tolerance. *World J Biol Chem.* 2015;6:48. Available from: <http://www.wjgnet.com/1949-8454/full/v6/i3/48.htm>.
- Branzei D, Szakal B. DNA damage tolerance by recombination: Molecular pathways and DNA structures. *DNA Repair (Amst).* 2016;44:68–75. Available from: <https://doi.org/10.1016/j.dnarep.2016.05.008>.

37. D'Souza S, Yamanaka K, Walker GC. Non mutagenic and mutagenic DNA damage tolerance. *Cell Cycle*. 2016;15:314–5. Available from: <https://doi.org/10.1080/15384101.2015.1132909>.
38. Diamant N, Hendl A, Vered I, Carell T, Reißner T, De Wind N, et al. DNA damage bypass operates in the S and G2 phases of the cell cycle and exhibits differential mutagenicity. *Nucleic Acids Res*. 2012;40:170–80.
39. Adar S, Hu J, Lieb JD, Sancar A. Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. *Proc Natl Acad Sci U S A*. 2016;201603388. Available from: <http://www.pnas.org/lookup/doi/10.1073/pnas.1603388113>.
40. Cordeiro-Stone M, Nikolaishvili-Feinberg N. Asymmetry of DNA replication and translesion synthesis of UV-induced thymine dimers. *Mutat Res*. 2002;510:91–106.
41. McGregor WG, Wei D, Maher VM, McCormick JJ. Abnormal, error-prone bypass of photoproducts by xeroderma pigmentosum variant cell extracts results in extreme strand bias for the kinds of mutations induced by UV light. *Mol Cell Biol*. 1999;19:147–54. Available from: <http://mcb.asm.org/content/19/1/147.abstract>.
42. Murugaesu N, Wilson GA, Birkbak NJ, Watkins TBK, McGranahan N, Kumar S, et al. Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov*. 2015;5:821–32.
43. Ross-Innes CS, Becq J, Warren A, Cheetham RK, Northen H, O'Donovan M, et al. Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat Genet*. 2015;47:1–11. Available from: <http://www.nature.com/doi/doi/10.1038/ng.3357>.
44. Souza RF. The role of acid and bile reflux in oesophagitis and Barrett's metaplasia. *Biochem Soc Trans*. 2010;38:348–52. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3072824&tool=pmcentrez&rendertype=abstract>.
45. Erichsen R, Robertson D, Farkas DK, Pedersen L, Pohl H, Baron JA, et al. Erosive reflux disease increases risk for esophageal adenocarcinoma, compared with nonerosive reflux. *Clin Gastroenterol Hepatol*. 2012;10:475–480.e1. Available from: <https://doi.org/10.1016/j.cgh.2011.12.038>.
46. Fein M, Maroske J, Fuchs KH. Importance of duodenogastric reflux in gastro-oesophageal reflux disease. *Br J Surg*. 2006;93:1475–82.
47. Kauppi J, Räsänen J, Sihvo E, Nieminen U, Arkkila P, Ahotupa M, et al. Increased oxidative stress in the proximal stomach of patients with Barrett's esophagus and adenocarcinoma of the esophagus and Esophagogastric junction. *Transl Oncol*. 2016;9:336–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27567957>.
48. Rasanen JV, Sihvo EIT, Ahotupa MO, Färkkilä MA, Salo JA. The expression of 8-hydroxydeoxyguanosine in oesophageal tissues and tumours. *Eur J Surg Oncol*. 2007;33:1164–8.
49. Jimenez P, Piazuelo E, Sanchez MT, Ortego J, Soteras F, Lanás A. Free radicals and antioxidant systems in reflux esophagitis and Barrett's esophagus. *World J Gastroenterol*. 2005;11:2697–703. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15884106>.
50. Dvorak K, Payne CM, Chavarria M, Ramsey L, Dvorakova B, Bernstein H, et al. Bile acids in combination with low pH induce oxidative stress and oxidative DNA damage: relevance to the pathogenesis of Barrett's oesophagus. *Gut*. 2007;56:763–71.
51. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet*. 2013;45:478–86. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3678719&tool=pmcentrez&rendertype=abstract>.
52. Nones K, Waddell N, Wayne N, Patch A-M, Bailey P, Newell F, et al. Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nat Commun*. 2015;5:1–9. Available from: <https://doi.org/10.1038/ncomms6224>.
53. Inoue M, Kamiya H, Fujikawa K, Ootsuyama Y, Murata-Kamiya N, Osaki T, et al. Induction of chromosomal gene mutations in *Escherichia coli* by direct incorporation of oxidatively damaged nucleotides: new evaluation method for mutagenesis by damaged dna precursors in vivo. *J Biol Chem*. 1998;273:11069–74.
54. Satou K, Kawai K, Kasai H, Harashima H, Kamiya H. Mutagenic effects of 8-hydroxy-dGTP in live mammalian cells. *Free Radic Biol Med*. 2007;42:1552–60. Available from: <https://doi.org/10.1016/j.freeradbiomed.2007.02.024>.
55. Satou K, Hori M, Kawai K, Kasai H, Harashima H, Kamiya H. Involvement of specialized DNA polymerases in mutagenesis by 8-hydroxy-dGTP in human cells. *DNA Repair (Amst)*. 2009;8:637–42.
56. Kamiya H. Mutations induced by oxidized DNA precursors and their prevention by nucleotide pool sanitization enzymes. *Genes Environ*. 2007;29:133–40.
57. Suzuki T, Kamiya H. Mutations induced by 8-hydroxyguanine (8-oxo-7,8-dihydroguanine), a representative oxidized base, in mammalian cells. *Genes Environ*. 2017;39:2. Available from: <https://doi.org/10.1186/s41021-016-0051-y>.
58. Rashid M, Fischer A, Wilson CH, Tiffen J, Rust AG, Stevens P, et al. Adenoma development in familial adenomatous polyposis and MUTYH-associated polyposis: somatic landscape and driver genes. *J Pathol*. 2016;238:98–108.
59. Suzuki T, Harashima H, Kamiya H. Effects of base excision repair proteins on mutagenesis by 8-oxo-7,8-dihydroguanine (8-hydroxyguanine) paired with cytosine and adenine. *DNA Repair (Amst)*. 2010;9:542–50. Available from: <https://doi.org/10.1016/j.dnarep.2010.02.004>.
60. Pavlov YI, Mian IM, Kunkel TA. Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. *Curr Biol*. 2003;13:744–8.
61. Mudrak SV, Welz-Voegele C, Jinks-Robertson S. The polymerase eta translesion synthesis DNA polymerase acts independently of the mismatch repair system to limit mutagenesis caused by 7,8-dihydro-8-oxoguanine in yeast. *Mol Cell Biol*. 2009;29:5316–26. Available from: <http://www.scopus.com/inward/record.url?eid=2-s2.0-70349329456&partnerID=tZ0tx3y1>.
62. Tomkova M, McClellan M, Kriacounis S, Schuster-Böckler B. DNA replication and associated repair pathways are involved in the mutagenesis of methylated cytosine. *DNA Repair (Amst)*. 2017. Available from: <https://www.sciencedirect.com/science/article/pii/S1568786417303464>.
63. Gao Z, Wyman MJ, Sella G, Przeworski M. Interpreting the dependence of mutation rates on age and time. *PLoS Biol*. 2016;14:e1002355. Available from: <http://dx.plos.org/10.1371/journal.pbio.1002355>.
64. Crossan GP, Garaycochea JI, Patel KJ. Do mutational dynamics in stem cells explain the origin of common cancers? *Cell Stem Cell*. 2015;16:111–2. Available from: <https://doi.org/10.1016/j.stem.2015.01.009>.
65. Bass AJ, Lawrence MS, Brace LE, Ramos AH, Drier Y, Cibulskis K, et al. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. *Nat Genet*. 2011;43:964–8. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3802528&tool=pmcentrez&rendertype=abstract>.
66. Wang K, Yuen ST, Xu J, Lee SP, Yan HHN, Shi ST, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet*. 2014;46:573–82. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24816253>.
67. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28:1811–7.
68. Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, et al. Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am J Hum Genet*. 2012;91:1033–40. Available from: <https://doi.org/10.1016/j.ajhg.2012.10.018>.
69. Encode Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74. Available from: <http://www.nature.com/nature/journal/v489/n7414/full/nature11247.html>.
70. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol*. 2016;17:31. Available from: <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0893-4>.
71. Tomkova M, Tomek J, Kriacounis S, Schuster-Böckler B. Mutational signature distribution varies with DNA replication timing and strand asymmetry. Source code. Bitbucket. <https://bitbucket.org/bsblabudwig/replicationasymmetry> (2018).
72. Tomkova M, Tomek J, Kriacounis S, Schuster-Böckler B. Mutational signature distribution varies with DNA replication timing and strand asymmetry Source code figshare. 2018; <https://doi.org/10.6084/m9.figshare.6941456>.