H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation

Received for publication, March 9, 2018, and in revised form, March 26, 2018. Published, Papers in Press, April 2, 2018, DOI 10.1074/jbc.RA118.002839

Yaping Huang‡, Liya Gu‡, and Guo-Min Li‡§1

From the ‡Department of Basic Medical Sciences, Tsinghua University School of Medicine, 100084 Beijing, China and the §Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75390

Edited by Xiao-Fan Wang

Histone H3 trimethylation at lysine 36 (H3K36me3) is an important histone mark involved in both transcription elongation and DNA mismatch repair (MMR). It is known that H3K36me3 recruits the mismatch-recognition protein MutSα to replicating chromatin via its physical interaction with MutSα’s PWPP domain, but the exact role of H3K36me3 in transcription is undefined. Using ChIP combined with whole-genome DNA sequencing analysis, we demonstrate here that H3K36me3, together with MutSα, is involved in protecting against mutation, preferentially in actively transcribed genomic regions. We found that H3K36me3 and MutSα are much more co-enriched in exons and actively transcribed regions than in introns and nontranscribed regions. The H3K36me3–MutSα co-enrichment correlated with a much lower mutation frequency in exons and actively transcribed regions than in introns and nontranscribed regions. Correspondingly, depleting H3K36me3 or disrupting the H3K36me3–MutSα interaction elevated the spontaneous mutation frequency in actively transcribed genes, but it had little influence on the mutation frequency in nontranscribed or transcriptionally inactive regions. Similarly, H2O2-induced mutations, which mainly cause base oxidations, preferentially occurred in actively transcribed genes in MMR-deficient cells. The data presented here suggest that H3K36me3-mediated MMR preferentially safeguards actively transcribed genes not only during replication by efficiently correcting mispairs in early replicating chromatin but also during transcription by directly or indirectly removing DNA lesions associated with a persistently open chromatin structure.

DNA mismatch repair (MMR)2 is a highly-conserved DNA repair pathway whose primary role is to correct mispairs arising during DNA replication (1–4). Genetic or epigenetic alterations in genes encoding MMR components result in a mutator phenotype and increase susceptibility to tumorigenesis (1, 2). The minimal essential components required for human MMR have been identified (5, 6). The MSH2–MSH6 heterodimer, called MutSα, is one of the most important components in human MMR, and its primary role is to recognize mismatches. MutSα binding to a mismatch triggers the downstream steps of MMR, including recruitment of other MMR factors, mismatch excision and repair DNA synthesis (3, 4).

Recent studies demonstrate that chromatin structure and epigenetic signals regulate many biological pathways and processes, including DNA repair and carcinogenesis. Our earlier studies show that trimethylated histone H3 lysine 36 (H3K36me3) recruits MutSα to chromatin by interacting specifically with the PWPP domain of MutSα (7). Deleting the MSH6 PWPP domain or inactivating SETD2, the major enzyme that trimethylates H3K36 (8), strongly suppresses MutSα recruitment to chromatin and causes a mutator phenotype (7). Similarly, H3K36M and H3K36L mutations also increase susceptibility to cancer (9, 10).

In addition to its role in MMR during DNA replication (7), H3K36me3 is highly associated with transcription processes, including transcription elongation and splicing (11). The histone mark is enriched in actively transcribed open chromatin (11) and is more abundant in exonic than in intronic regions. However, the exact role that H3K36me3 plays in transcription is unknown. Conversely, MMR has been reported to participate in transcription-coupled nucleotide excision repair (12), but the idea was highly controversial (13, 14). Even if MMR is involved in transcription, it is unclear how it works.

Previous studies have shown that, in addition to recognizing mispairs, MutSα specifically interacts with DNA lesions that are usually processed by base excision repair and nucleotide excision repair pathways, which include 8-oxo-7, 8-dihydro-2’-deoxyguanosine (15), and many bulky DNA adducts (16). These studies have established MutSα as a general DNA damage sensor, although the biology of this function is not quite understood. It is likely that MutSα’s broad lesion recognition function constitutes its role in transcription.

This work was supported in part by National Natural Science Foundation of China Grants 31370766, 31570814, and 81630077; National Natural Science Foundation of China–Israel Science Foundation Joint Research Program Grant 3146143005; National Key R&D Program of China Grant 2016YFC1302300; National Institutes of Health Grants GM112702 and CA19203; and funds from Tsinghua-Peking Joint Center for Life Sciences. The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article contains Figs. S1–S5 and Tables S1 and S2.

1 Holds the Reece A. Overcash, Jr., Chair for Center for Research on Colon Cancer. To whom correspondence should be addressed. Tel.: 214-648-9955; Fax: 214-648-9995; E-mail: guo-min.li@utswwestern.edu.

2 The abbreviations used are: MMR, mismatch repair; SNV, single nucleotide variant; Seq, sequencing; TSG, tumor suppressor gene; WGS, whole-genome sequencing; NER, nucleotide excision repair; BER, base excision repair; indels, insertions/deletions; RPKM, reads per kilobase pair per million; PCNA, proliferating cell nuclear antigen; HMM, hidden Markov model; BWA, Burrows-Wheeler Aligner.

This is an Open Access article under the CC BY license.
Mismatch repair preferentially protects transcribed genes

Based on our earlier demonstration that H3K36me3 interacts specifically with the MSH6 subunit of MutSα (7), it is possible that highly enriched H3K36me3 recruits MutSα to actively transcribed regions to mitigate the risk of DNA damage during transcription and/or actively promote lesion repair in the transcribed regions. To test this hypothesis, we used ChIP combined with sequencing (ChIP-Seq) and whole-genome sequencing (WGS) to characterize the distribution of H3K36me3 and MSH6 on chromatin and to create genome-wide profiles of relative mutation frequency in HeLa cells with or without disruption of the H3K36me3–MutSα interaction. The results presented here demonstrate the co-enrichment of H3K36me3 and MutSα in actively transcribed regions, which are consistent with the idea that H3K36me3 and MutSα play roles in protecting the integrity of DNA in open chromatin during transcription, as disrupting the H3K36me3–MutSα interaction preferentially promotes mutations in actively transcribed regions. This notion is further supported by the fact that H2O2-induced DNA lesions, which are substrates of base excision repair, preferentially cause mutations in actively transcribed genes in MMR-deficient cells. These results suggest that MMR is involved in removing oxidative DNA damage during transcription, which is distinct from its traditional mismatch correction function during DNA replication. This study also provides evidence for a novel function of H3K36me3 in removing transcription-associated lesions via MMR.

Results
Genome-wide distribution and co-enrichment of H3K36me3 and MSH6

We performed whole-genome ChIP-Seq analyses in HeLa cells (referred to as P0, see “Experimental procedures”) arrested at G1/S or S phase using antibodies against H3K36me3 or MSH6 that had been used for ChIP assays previously (see “Experimental procedures”). We used the RSEG pipeline (17) to map the genome-wide distributions of H3K36me3 and MSH6 with high confidence (see “Experimental procedures”). As expected, H3K36me3 was enriched in gene bodies and more enriched in exonic regions than in intronic regions (Fig. 1A). Binding domains were overlaid on transcribed genes in the reference genome Hg19 (http://genome.ucsc.edu/).3 H3K36me3-enriched domains aligned with 15,884 genes, and MSH6-enriched domains aligned with 5100 genes (Fig. 1B). The limited MSH6 enrichment is likely due to competition between MSH6 and other H3K36me3 reader proteins (18) and/or inefficient pulldown of chromatin fragments by the MSH6 antibody used because of a relatively weak interaction between MSH6 and H3K36me3. Nevertheless, ~97% (4924/5100) of MSH6-enriched genes are enriched for H3K36me3 (Fig. 1B), which indicates MSH6 recruitment to chromatin via H3K36me3.

To further define the relationship between H3K36me3 and MSH6 and their distribution and abundance in the genome, we classified the whole genome into groups based on H3K36me3 ChIP intensity and then plotted these data against the corresponding MSH6 ChIP intensity. H3K36me3 and MSH6 ChIP intensities correlate with each other in a linear fashion throughout the entire genome (Fig. 1C). Representative loci demonstrating this correlation are shown in Fig. 1D. To determine the specific distributions of H3K36me3 and MSH6 and the differences between these distributions in replicating and nonreplicating cells, we performed ChIP-Seq analyses in HeLa cells arrested in S phase or G1/S boundary. We found that 60–80% of H3K36me3-enriched regions were co-bound in both phases (Fig. 1E). Interestingly, the ChIP intensities for H3K36me3 and MSH6 also correlated with each other in both phases (Fig. 1F).

Finally, we analyzed the relationship between transcription activity and H3K36me3/MSH6 distribution and demonstrated that regions bound by both H3K36me3 and MSH6 show higher transcription activities than those bound by H3K36me3 alone (Fig. 1G). Taken together, these results suggest that MSH6 is preferentially recruited to actively transcribed regions in both replicating and nonreplicating cells. Given MMR’s essential role in correcting replication errors, it is understandable that MSH6 is recruited to chromatin in replicating cells. Why MSH6 is also associated with chromatin in nonreplicating cells, particularly in actively transcribed regions, remains to be investigated.

H3K36me3–MSH6 co-enrichment is inversely correlated to local mutation frequency

We used whole-genome sequencing data from HeLa P0 cells to construct a genome-wide map of mutations, i.e. single nucleotide variants (SNVs) and insertions/deletions (indels). These mutations were aligned with the genome-wide distributions of H3K36me3 and MSH6. To determine the impact of H3K36me3/MSH6 enrichment and distribution on local mutations, we classified the whole genome into functional segments, i.e. promoter, exon, intron, and intergenic regions, and mapped H3K36me3/MSH6 enrichments and mutation frequencies to the individual segments. Consistent with the data shown in Fig. 1A, H3K36me3 was more enriched in exonic regions than in intergenic regions, as was MSH6 (Fig. 2A). However, relative mutation frequency was inversely related to the enrichment of both H3K36me3 and MSH6, as exons were less susceptible to mutation, and introns and intergenic regions were more susceptible to mutation (Fig. 2, A and B). In addition, exons with higher H3K36me3 intensities exhibited lower mutation frequencies than those with lower H3K36me3 intensities (Fig. S1A). These results reveal that mutation frequency is inversely correlated to H3K36me3/MSH6 enrichment.

We also found that the H3K36me3–MSH6 interaction is critical for mutation avoidance, as mutations were found in only 18.8% of the H3K36me3–MSH6 co-bound genes but 44.6% of the genes bound by H3K36me3 alone. Conversely, the mean mutation frequency in H3K36me3-only genes is higher than in the co-bound ones (Fig. S1B). These data indicate that H3K36me3 executes its genome stability function via MMR.

It is known that H3K36me3 enrichment tends to increase from the 5’ end to the 3’ end in gene bodies (19). To determine

---

3 Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.
this distribution tendency’s influence on mutation frequency, we divided all gene bodies in the genome into 10 segments from 5’ to 3’ and analyzed their relative H3K36me3- and MSH6-ChIP intensities and mutation frequencies. As shown in Fig. 2C, H3K36me3-ChIP signals indeed increase from the 5’ end to the 3’ end in the gene bodies, as do the MSH6-ChIP signals. Mapping mutations in the corresponding sequences revealed that mutation frequencies in 5’ sequences are generally higher than those in 3’ sequences (Fig. 2C), suggesting that even within the gene body H3K36me3-dependent MMR determines mutation frequency.

Replication timing substantially influences gene mutability (20). To determine the involvement of H3K36me3 in this process, we examined the relationship between replication timing, enrichment for H3K36me3, and mutation frequency. The results show that regions relatively enriched for H3K36me3 tend to replicate earlier and have a lower relative mutation frequency than regions with fewer H3K36me3 signals (Fig. 2D). These observations imply that H3K36me3-dependent MMR is critical in replication timing-associated mutability. It is worth mentioning that the replication timing-associated mutation frequency difference observed in HeLa P0 cells is not as big as those reported in certain cancer genomes (20–22). We believe that this may be related to DNA repair capacities. HeLa cells are proficient in all known DNA repair pathways, including MMR (23), but many cancer genome studies reported in the literature used cancer cases that usually carry a defective DNA repair pathway.
Gene bodies highly enriched for H3K36me3 display high mutation frequencies

The data shown above strongly suggest that H3K36me3/MSH6 enrichment is inversely associated with local mutation frequency. However, we found that this rule does not apply to a small portion (0.4%) of the genome, where both H3K36me3/MSH6 intensity and mutation frequency are high (highlighted in the gray box in Fig. 3, A and B). We divided the whole genome into 18 groups according to their relative H3K36me3/MSH6 ChIP-intensity (i.e., 1–6, 7–12, and 13–18 classified as low, intermediate, and high intensity groups, respectively) and quantified the mutation frequency in each of these groups from the whole-genome sequencing data of HeLa P0 cells. As shown in Fig. 3, A and B, when H3K36me3/MSH6 intensity is less than 9, the number of both SNVs and indels per kb gradually decreases as the abundance of H3K36me3/MSH6 increases; however, when H3K36me3/MSH6 intensity reaches a relatively high level (higher than level 13), we observed increased mutation frequencies in these regions. To determine whether this unusual phenomenon is related to transcription, we analyzed mutation frequencies of gene bodies and intergenic sequences in the three groups. The results show that, unlike in the gene bodies, mutation frequencies in these intergenic sequences were inversely correlated to the abundance of H3K36me3 (Fig. 3C), consistent with the overall inverse correlation between H3K36me3 abundance and mutation frequency (see above). Because the major difference between gene bodies and intergenic sequences is their transcriptional activity, it is possible that the excess mutations identified in H3K36me3-high regions are associated with transcription.

Previous studies have demonstrated mutational signatures of base substitutions associated with transcription-coupled damage in humans (24) and bacteria (25). We therefore analyzed the mutational signature of the base substitutions by incorporating information regarding the 5’ and 3’ neighboring sites of each mutated base in the WGS data from HeLa P0 cells, as described (24). The results reveal predominant C→T mutations occurring at NpCpG trinucleotides (Fig. S2A), a characteristic of mutational signature1A/B (24, 26). This mutational signature occurred more frequently: 1) in H3K36me3-high regions than H3K36me3-low regions (Fig. 3D and Fig. S2A) and 2) in H3K36me3-high gene bodies than H3K36me3-high intergenic regions (Fig. S2, B, upper panel). However, the latter does not apply to H3K36me3-low gene bodies and intergenic regions (Fig. S2, B, bottom panel). These results imply that the excess mutations in H3K36me3-high gene bodies are likely related to transcription-associated DNA damage and repair.

In addition to mutational signatures, strand bias of C→T, C→A, and T→C mutations have also been linked to transcription-coupled damage and repair (24, 26). To determine whether the excess mutations in H3K36me3-high gene bodies are strand-biased, we designated genomic regions as tx (+) when they code genes on the reference strand and tx (−) when they code on the complementary strand, as described by Har-
adhvala et al. (26), and we calculated the strand asymmetry scores (see “Experimental procedures”) of each of the three mutation types in H3K36me3-high (intensity groups 13–18), -intermediate (intensity groups 7–12), and -low (intensity groups 1–6) regions. Fig. 3E shows the strand asymmetry analysis of C→T, C→A, and T→C mutations. We detected little strand bias in these mutations over the entire genome (Fig. 3E, upper panel) but a relatively strong strand bias in H3K36me3-low regions (middle and bottom panels). Interestingly, despite excess mutations in H3K36me3-high regions (Fig. 3A–D), the predominant C→T and T→C mutations did not show the strong strand bias as seen in H3K36me3-low regions (Fig. 3E), which indicates that DNA repair systems (27–29), including H3K36me3-dependent MMR, repair such transcription-coupled damages in transcribed and nontranscribed DNA strands in a balanced manner. Thus, at least some of the excess mutations in H3K36me3-high regions likely accumulated because of insufficient DNA lesion repair on the transcribed strand during active transcription.

**Actively transcribed genes are highly susceptible to mutation despite being enriched for H3K36me3**

To determine transcription’s association with the excess mutations observed in H3K36me3-high regions, we performed RNA-Seq analysis in HeLa P0 cells, as described under “Experimental procedures.” Nontranscribed genomic regions were defined as bins with zero reads mapped, whereas transcribed genomic regions were identified as bins with one or more reads mapped. This analysis excluded rRNA transcripts, but genomic regions encompassed by spliced reads (e.g. introns) were binned with transcribed genomic regions. The overall fold-enrichment for H3K36me3/MSH6 was significantly higher (Fig. 4A, p < 0.01 and p < 0.05 for H3K36me3 and MSH6, respectively), and relative mutation frequency was significantly lower (Fig. 4B, p < 0.01) in all transcribed regions, including noncoding RNA genes, than in nontranscribed regions. Mutational analysis of the actively transcribed 5S RNA gene clusters and their flanking nontranscribed intergenic spacer regions revealed that only 4 of the 17 gene regions harbored a mutation, but 11 of the 16 nontranscribed intergenic spacer regions contained 1–5 mutations. Although we did not link excess mutations in H3K36me3-high gene bodies to all transcribed genes, these data further suggest that H3K36me3-dependent MMR preferentially protects transcribed genes from mutations.

Given that highly transcribed genes are more susceptible to mutation than less active and silent genes (26, 30, 31), we next analyzed transcription activity and its relationship with H3K36me3 abundance and mutation frequency. We classified all regions (including those for noncoding RNAs) in the whole
genome into six groups based on their transcription levels, with group 1 and group 6 being the lowest and highest level in transcription activity, respectively. We found that mutation frequency is inversely correlated to H3K36me3 abundance in genomic regions in transcription groups 1–3 (Fig. 4C), which consist of the vast majority of the regions analyzed. For group 4 regions, mutation frequency does not correlate with the abundance of H3K36me3. However, for regions in groups 5 and 6, which represent a very small portion of all the regions analyzed, mutation frequencies increase gradually as H3K36me3 intensity increases (Fig. 4C, see the regions highlighted in the gray box). Thus, an overall biphasic correlation between mutation frequency and H3K36me3 intensity was observed (Fig. 4C, green solid line), which is consistent with the data shown in Fig. 3A. We indeed found that the vast majority (~78%) of these highly transcribed genes are located in the regions with excess mutations shown in Fig. 3A (regions between CHP intensity 12–18). These unexpected results suggest that highly transcribed genes, although heavily protected by H3K36me3-dependent MMR, are highly susceptible to mutations (see below for details). These mutations are probably derived from DNA damage in open chromatin during transcription (32, 33).

**Disrupting the H3K36me3–MSH6 interaction preferentially increases mutation frequency in highly transcribed regions**

SETD2 is the major methyltransferase that trimethylates H3K36 (8), and the MSH6 PWWP domain interacts with H3K36me3 (7). To determine the actual involvement of H3K36me3-dependent MMR in transcription-associated damage removal, we examined the relationships between H3K36me3/MSH6 abundance, transcription activity, and de novo mutations from various HeLa cells (i.e. SETD2-knockout or MSH6 PWWP-domain disrupted) with 100 passages (P100, see “Experimental procedures”). As expected, we detected very low levels of H3K36me3 and MSH6 on chromatin in SETD2-knockout (SETD2-KO) HeLa cells (Figs. S3, A and B, and S4, A and B). Consistent with previous studies (7), the de novo mutation frequency was elevated in SETD2-KO cells (Fig. S4C). Interestingly, the de novo mutations in SETD2-KO cells but not...
Mismatch repair preferentially protects transcribed genes

in control HeLa cells occurred preferentially in exonic regions originally enriched for H3K36me3/MSH6 (Fig. 5A), and the relative mutation frequency in SETD2-KO cells correlates with the original fold-enrichment for H3K36me3 in control HeLa cells (Fig. 5B). In addition, relative mutation frequencies were higher in transcribed genes than in nontranscribed genes in SETD2-KO HeLa cells, and the opposite was observed in control HeLa cells (Fig. 5C). We similarly analyzed the relationship between transcription level and de novo mutation frequency and found that mutation frequency correlates positively with transcription activity in SETD2-KO cells but not in control HeLa cells (Fig. 5D).

To rule out the possibility that H3K36me3 prevents mutations in actively transcribed regions by recruiting other repair proteins, we also performed similar analyses in HeLa cells carrying a complete knockout of MSH6 or an MSH6 defective in the PWWP domain (Fig. S3, C–E), which only disrupted the MMR pathway or disrupted the interaction between H3K36me3 and MSH6 without impairing other DNA repair pathways. We observed similar results to those obtained in the SETD2-KO HeLa cells (Fig. 5A–D).

To test whether the de novo mutations observed in H3K36me3–MSH6-deficient HeLa cells are associated with transcription-associated DNA lesions, we analyzed the mutational signatures in all three MMR-deficient cells and their control HeLa cells, as described above (see Fig. S2 and Fig. 3). The overall mutation spectrum is quite similar in these cells (Fig. S4D) and also similar to that observed in HeLa P0 cells (Fig. S2). However, when we analyzed the detailed mutational signatures on different gene groups with different transcription levels, we found an increase in the C→T mutational signature associated with an increase in transcriptional activities in all three MMR-deficient cells (Fig. S4E), consistent with the results shown in Fig. 3D. More strikingly, analysis of mutational strand bias revealed that, compared with wildtype (WT) HeLa P100 cells, strand-biased mutations in cells defective in the SETD2, MSH6, and MSH6 PWWP domains were all significantly compromised for C→T mutations (Fig. 5E, middle and bottom panels). Similarly, the degree of mutational strand bias for C→A and T→C mutations was also reduced (Fig. 5E). Given that all cell lines used here are proficient in both base excision repair (BER) and nucleotide excision repair (NER), but differ in MMR, we conclude that the observed mutations and the loss of mutational strand bias in actively transcribed genes in H3K36me3–MSH6-defective cells are due to the loss of MMR function.

To selectively examined enrichment for H3K36me3 and relative mutation frequency in 123 tumor suppressor genes anno-
Mismatch repair preferentially protects transcribed genes

The results in control HeLa cells revealed the following: 1) H3K36me3 was more enriched in these genes than in control genes (Fig. 6A); 2) relative mutation frequency was lower in these tumor suppressor genes transcribed at a higher level than control genes (Fig. 6B), consistent with a previous study (35). In H3K36me3-depleted cells (i.e. SETD2-KO), nonsilent mutations occurred preferentially in these tumor suppressor genes. However, not all tumor suppressor genes carried a new mutation in SETD2-KO cells. For example, we detected de novo mutations in BRCA2, DAXX, NOTCH1, TET1, TGFBR2, ARID2, NF2, and FAT1 but not in many other tumor suppressor genes analyzed in this study (Table S1). Interestingly, 5 of the 11 (~46%) nonsilent point mutations shown in Table S1 are C→T mutations.

We reasoned that the discrepancy in mutation frequencies among these tumor suppressor genes can be explained by their transcription activity, which has been shown to influence the stability of the local sequences (Figs. 4 and 5). We therefore compared the mutagenic activities of tumor suppressor genes that were enriched for H3K36me3/MSH6 prior to SETD2 knockout, but possessed different transcription activities. In general, relatively highly transcribed tumor suppressor genes were more prone to mutation in MMR-deficient cells (Fig. 6D). Fig. 6E shows six representative genes in this analysis. Even though the SETD2 knockout depleted the H3K36me3 signal in all ARID2, NF2, FAT1, CEBPA, CDH1, and DNMT3A genes, we detected mutations in the highly transcribed genes ARID2, NF2, and FAT1 but not in the much less transcribed genes CEBPA, CDH1, and DNMT3A. These results indicate that actively transcribed tumor suppressor genes are preferentially safeguarded by H3K36me3-dependent MMR in normal circumstances to avoid mutations during transcription, but they are susceptible to mutation when the MMR system is down. Our data provide a further molecular explanation for why actively transcribed genes are targeted for mutation (36–38).

**Actively transcribed cancer genes are more susceptible to oxidation-induced mutations in H3K36me3- or MSH6-depleted cells**

To further confirm that the observed mutations in actively transcribed genes are due to loss of MMR activity associated with transcription, we treated MMR-proficient and -deficient HeLa cells with a low concentration of H2O2, which mainly induces oxidative DNA damage, such as 8-oxo-guanine and strand breaks. These oxidative DNA lesions are independent of DNA replication and are not classical substrates for MMR. The H2O2 treatment did not block loading of H3K36me3 and MSH6 to chromatin in the transcription-active G1 phase (Fig. S5, A and B). We analyzed H2O2-induced mutations by DNA sequencing of three actively transcribed cancer genes CALR, MYC, and TP53 and two transcriptionally inactive genes CD79A and GATA1. We observed that the relative mutation frequencies were significantly higher in the actively transcribed genes than in the transcriptionally inactive ones in SETD2-KO, MSH6-KO, or MSH6-PWWP-defective cells (Table 1 and Table S2). However, we did not detect this discrepancy in mutation frequencies between actively transcribed and transcrip-
Mismatch repair preferentially protects transcribed genes

Table 1

Mutation frequency induced by H$_2$O$_2$ in actively transcribed or inactively transcribed cancer genes

Cells were treated with 50 μM H$_2$O$_2$ for 6 days, and five cancer genes were selected for mutation analysis. The expression level was indicated by reads/kb/million reads (RPKM) value. Relative mutation frequency was calculated as mutations/sequence length/generations. Fold change was calculated by dividing median mutation frequency of actively transcribed genes by that of inactively transcribed ones.

| Gene | Actively transcribed | Inactively transcribed | Fold change (Actively/Inactively) |
|------|----------------------|------------------------|----------------------------------|
| CALR | 286.54               | 62.71                  | 19.60                            |
| MYC  | 1.1                  | 2.4                    | 1.1                              |
| TP53 | 10.9                 | 4.8                    | 3.3                              |
| CD79A| 0.05                 | 0.2                    | 0.02                             |
| GATA1| 5.4                  | 4.8                    | 2.2                              |

It is worth noting that in H3K36me3- and MSH6-defective cells, H$_2$O$_2$-induced mutation frequencies in these genes are essentially proportional to their transcriptional activities, e.g. CALR > MYC > TP53 (Table 1), consistent with the data shown in Fig. 5D. Because all cell lines used here differ only in terms of their MMR activity, these results confirm that H3K36me3-mediated MMR preferentially protects actively transcribed genes from mutation.

Discussion

MMR is well known for correcting mismatches generated during DNA replication to ensure high DNA replication fidelity. We have recently shown that H3K36me3 is an important MMR factor, as it recruits MutSα to replicating chromatin in human cells (7). Here, we provide the evidence showing that H3K36me3-dependent MMR preferentially protects actively transcribed genes from mutation. We believe that MMR does this through both replication and transcription. Therefore, this study has identified a new genome maintenance function for H3K36me3 and MMR in repairing transcription-associated DNA lesions.

Under normal circumstances, both H3K36me3 and MSH6 are more abundant not only in gene bodies than in other genomic regions but also in actively transcribed genes than in less actively transcribed and silent genes (Figs. 2 and 4). Correspondingly, local mutation frequency is inversely correlated with H3K36me3/MSH6 intensity (Figs. 2 and 4). However, in H3K36me3-depleted cells or cells in which the H3K36me3–MSH6 interaction had been disrupted, mutations occurred preferentially in actively transcribed regions (Fig. 5). These observations clearly suggest that H3K36me3-mediated MMR preferentially protects actively transcribed regions from mutation.

It is known that active protein-coding genes resided in early replicating chromatin (45–47), where H3K36me3/MSH6 are highly enriched (Fig. 2D). Thus, the MMR system can efficiently correct errors generated in actively transcribed genes during early replication timing, resulting in error-free replication. In addition, the following facts support a role for MMR in transcription-associated DNA lesion removal. We observed that a number of very actively transcribed genes, which are highly enriched for H3K36me3/MSH6 and supposed to be in early replicating chromatin, had a mutation frequency higher than that of genes with less transcription activity and less abundance in H3K36me3/MSH6 (Fig. 4, C and D). This phenomenon cannot be explained by replication-associated mutations, as the very actively transcribed genes would have been better protected by H3K36me3-mediated MMR. However, during transcription very actively-transcribed genes are persistently exposed in open chromatin structure and thus suffer more DNA damage-induced mutations than less actively transcribed genes. This assumption is supported by the fact that actively transcribed genes displayed a higher H$_2$O$_2$-induced mutation frequency than less actively transcribed ones in MMR-deficient cells (Table 1). Given that all cells used are competent in other DNA repair pathways and that H$_2$O$_2$-induced damage is not replication-specific, the simplest explanation is that defects in H3K36me3-mediated MMR are responsible for the observed mutations in actively transcribed genes during transcription. These results agree with recent studies demonstrating that human cells differentially repair DNA lesions or mispairs in euchromatin and heterochromatin (36) and that mutations preferentially occur in active genes in tumors defective in MMR (37, 48).

Transcription-coupled NER has been well documented to selectively remove some DNA lesions from transcribed strands (39, 40). However, NER’s DNA lesion specificity may limit its role in transcription-coupled repair. In contrast, MutSα has a very broad DNA substrate capacity and is capable of recognizing many DNA lesions that are normally processed by BER and NER pathways, such as G:U mispairs, 8-oxo-guanine, O6-methylguanine, and UV dimers (15, 16, 29, 41–44). In addition, unlike other DNA repair genes, whose expressions are induced in response to DNA damage, the expressions of MMR genes are
Mismatch repair preferentially protects transcribed genes

maintained at a very stable level in all cell cycle phases (7), making them available at any time when needed. Thus, both the availability and capability render MMR an ideal genome maintenance system.

It is known that during DNA replication, MMR corrects biosynthetic errors in a strand-specific manner and that strand specificity is directed either by a pre-existing strand break (e.g., the ends of Okazaki fragments) or by endonucleolytic cleavage by MutLα. How the MMR system removes the strand-specific damage during transcription is unknown. However, previous studies have provided some hints to the answer. MMR has been shown to occur in nondividing cells exposed to high levels of DNA damage, a reaction called noncanonical MMR (49–51). Recent studies by Modrich and co-workers (52–54) have revealed that covalently closed circular DNA containing a lesion or a helix perturbation that MutSα recognizes stimulates PCNA loading and subsequently activates the MutLα endonuclease activity. Although the endonucleolytic cleavage by MutLα lacks strand bias in vitro, it has been proposed that in vivo interactions between PCNA and MutSα (55) and/or unidentified DNA signals confer strand specificity on the reaction. We therefore propose that, after H3K36me3 recruits MutSα to actively transcribed chromatin containing a DNA lesion, MutSα directly promotes repair of the DNA lesion, thereby mitigating the risk of locally higher mutation frequency in actively transcribed genes and ensuring both the stability of the transcribed genes and transcription accuracy. It is also possible that MutSα recruits other DNA repair proteins to DNA lesions in actively transcribed regions, similar to the process known as transcription-coupled nucleotide excision repair (56).

We also show that mutation frequency is higher in introns than in exons in MMR-proficient HeLa cells (Fig. 2A), but the order of mutation frequencies in these regions is reversed when H3K36me3 or the H3K36me3–MSH6 interaction was depleted or disrupted (Fig. 5A), suggesting a critical role for MMR in maintaining exon stability. These results are in agreement with a recent bioinformatics study that analyzed the whole-genome expression and mutation data of several hundred MMR-proficient and MMR-deficient tumors deposited in the TCGA database. The data suggest that mismatches in exonic DNA are repaired more efficiently than their intronic counterparts in a manner that depends on H3K36me3-dependent MMR (48). However, because many introns are transcribed together with their corresponding exons, how does the MMR system selectively repair DNA lesions in exons? Given that H3K36me3 is highly enriched in exons, Schwartz et al. (57) have proposed that H3K36me3 levels define exon–intron boundaries during mRNA synthesis and H3K36me3’s preferential recruitment of MutSα to exon-containing nucleosomes allows MMR to specifically target exons. Based on their bioinformatics analysis, Frigola et al. (48) suggest that a cross-talk between the RNA splicing machinery and MMR could account for the differential mutation frequencies in exons and introns. However, further studies are required to elucidate the detailed mechanism by which H3K36me3 and MutSα promote stability in actively transcribed genomic regions.

Experimental procedures

Cell lines and cell culture

SETD2 and MSH6 knockout HeLa (SETD2-KO and MSH6-KO) and PWWP-mutant HeLa (PWWP-MT1/2) cells were created using CRISPR–Cas9 genome-editing technology with small guide RNAs targeting to the third and the second exon of SETD2 and MSH6, respectively, as described previously (58). The PWWP-mutant cells are by-products of an MSH6 knock-out that deleted a fragment (PWWP-MT1, 112–120–amino acid deletion; PWWP-MT2, 122–128–amino acid deletion) in the MSH6 PWWP domain without disrupting the whole protein. Cells were grown in Dulbecco’s modified Eagle’s medium (Invitrogen) supplemented with 10% fetal bovine serum.

Cell synchronization and cell cycle analysis

We synchronized cells as described (7) and arrested cells at G1/S by culturing for 18 h in complete medium with 2 mM thymidine, 10 h in thymidine-free medium, and then 15 h in thymidine-containing medium. We harvested early S phase cells by releasing G1/S cells into thymidine-free medium for 1.5 h and confirmed cell cycle status by flow cytometry.

Western blotting and chromatin fractionation

We performed Western blotting analyses using antibodies against SETD2 (A3194, Abclonal), MSH6 (sc-10798, Santa Cruz Biotechnology), H3 (ab12079, Abcam), H3K36me3 (ab9050, Abcam) or β-actin (BE0033, EasyBio), according to the manufacturer’s recommendations.

SETD2-KO, PWWP-MT, and control HeLa cells were fractionated as described (59) with some modifications. Briefly, cells were lysed for 10 min on ice in 1 ml of ice-cold 0.5% Triton X-100/mCSK buffer containing multiple protease inhibitors and 200 μM Na3VO4. Cell lysates were subjected to centrifugation (2000 × g, 3 min) to obtain nuclear fractions. Nuclei were digested with 250 units/ml DNase I in 0.1% Triton X-100/mCSK containing 2 mM MgCl2 and 1 mM MnCl2 at room temperature for 20 min. The samples were centrifuged (2000 × g, 3 min), and the supernatant fraction, which contains solubilized chromatin, was saved.

Microscopy and immunofluorescence analysis

We performed immunofluorescence analysis as described (7) with antibodies against H3K36me3 (ab9050, Abcam) and MSH6 (610968, BD Biosciences). Fluorescence images were obtained using Axio Observer Z1 microscope (Carl Zeiss MicroImaging), and images were processed and analyzed using ImageJ software.

ChIP and ChIP-Seq analysis

We performed ChIP using extracts from 2 × 10^7 cells and 4 μg of ChIP-grade H3K36me3 (ab9050, Abcam) or MSH6 (sc-10798, Santa Cruz Biotechnology) antibody, according to the native ChIP (N-ChIP) protocol, as described previously (60). For MSH6 ChIP, we first cross-linked cells with 1% formaldehyde at room temperature for 10 min before micrococcal nuclease digestion. We kept 5% chromatin fragments digested by micrococcal nuclease as input. Both the input and ChIP
products were deproteinized using phenol/chloroform extraction, followed by sequencing library construction and sequencing in single-end 50-bp (SE50) mode using the Illumina Next-Generation-Sequencing platform.

We subjected the sequencing reads to the standard quality control pipeline and mapped clean reads to a human reference genome (UCSC hg19) by the short oligonucleotide alignment package (SOAP2.1) with default parameters (61). Reads mapping to more than one position in the genome were filtered out. Multiple reads mapping to the same position were counted once to avoid potential PCR bias. We used the RSEG pipeline (17) to identify H3K36me3- and MSH6-enriched regions and ChIP-Seq intensities. RSEG models the read counts with a negative binomial distribution and subsequently uses a two-state hidden Markov model (HMM) to segment the genome into foreground domains and background domains. To identify H3K36me3- and MSH6-enriched domains, we regarded sequencing reads from DNA fragments without enrichment by any antibody (input) as background compared with reads from ChIP-ed samples under rseg-diff mode-2 with default settings. We computed the bin size based on total read counts, and the effective genome size with the posterior probability of each bin obtained by HMM decoding is larger than 0.95 (false discovery rate <5%). The mean of read counts within a region is above the 90th percentile of foreground emission distribution. Finally, we merged adjacent enriched bins into enriched domains. We used domain score (> = 10), which measures both the quality and size of the domain, as a filter parameter to keep only highly confident domains.

**Whole-genome sequencing analysis**

We performed whole-genome sequencing using the Illumina HiSeq 2500 (San Diego) platform in PE150 (paired-end 150 bp) mode with an average 30× depth for each base. We processed raw image files using the Illumina Pipeline (version 1.3.4) for base-calling with default parameters. The clean reads were aligned to the HeLa genome reference (62) by Burrows-Wheeler Aligner (BWA) alignment pipeline with default parameters (63). We then processed the BAM files derived from BWA alignment by the Genome Analysis Toolkit (GATK version 1.6) (64) to perform re-alignment around known indel sites. We subjected all aligned reads to GATK count covariates based on known SNVs (dbSNP135) and recalibrated base quality using GATK table recalibration. Both uniquely-mapped reads and multiple-mapped reads were used to identify variations. Multiple aligned reads with a mismatch = 1 were randomly assigned to one site for each mismatch. We used the GATK unified genotyper to identify genome-wide SNVs and indels with default parameters and filtered variations by quality control.

To elucidate the accumulated de novo mutations in SETD2-KO, MSH6-KO, PWWP-MT, and WT HeLa cells, we used an experimental evolution strategy, as described previously (65). All cells were cultured for 100 generations (P100) and then subjected to WGS analysis, as described above. We regarded all the cells (SETD2-KO P100, MSH6-KO P100, and PWWP-MT P100 and two replicates of HeLa P100 and HeLa P0) as a uniform population to discard the common mutations identified in all the samples. Then, we compared the P100 data to the HeLa P0 data to identify de novo mutations that occurred only in P100 cells but not in P0 cells. We used mutations in HeLa P0 cells to analyze the H3K36me3/MSH6–mutation relationship under normal circumstances (Figs. 2–4). We used de novo mutations to address mutations that preferentially targeted actively transcribed genes in H3K36me3–MMR-deficient cells (Figs. 5 and 6, D and E). The sequencing data are accessible in NCBI with SRA accession no. SRP128160.

**Mutational signature and strand bias analyses**

We performed mutational signature analysis as described (24). We counted 96 substitution classifications defined by six substitution classes (C→A/G→T, C→G/G→C, C→T/G→A, T→A/A→T, T→C/A→G, and T→G/A→C) and analyzed the sequence context immediately 5’ and 3’ to the mutated nucleotide in the whole genome or in certain genomic regions. Then, we calculated the percentage of each classification and normalized to its corresponding trinucleotide frequency in the whole genome or in certain genomic regions.

We analyzed mutational strand bias as described (26). We annotated genomic regions as tx(+) if they encoded genes on the reference strand and as tx(−) if they encoded genes on the complementary strand. For each of the six mutation types, we calculated the mutation frequencies of the pair of complementary substitutions (e.g. C→T and G→A) in both the tx(+) and tx(−) groups. Then, we calculated the strand asymmetry score as the log2 ratio of the mutation frequencies for the complementary substitutions (e.g. log2(C→T/G→A)) in the entire genome or in the tx(+) and tx(−) groups. The degree of strand bias increases as the absolute value of the strand asymmetry score increases.

**Processing Repli-Seq data and HeLa transcriptome data**

We obtained Repli-Seq data from ENCODE for HeLa-S3 cells with accession no. ENCSR647UES (66). Signal intensities were averaged over 1-kb intervals throughout the genome. We downloaded HeLa-S3 RPKM data for Hg19 annotated genes with UCSC accession no. wgEncodeEH000130 (66). These genes were stratified into groups by RPKM value, ranking from low to high.

**Identifying transcribed and nontranscribed genomic regions in HeLa**

HeLa RNA-Seq reads were processed using a standard filter pipeline and aligned to Hg19 using Bowtie2 (67) with default parameters. The mapped reads were then mapped to H3K36me3 bins. Nontranscribed genomic regions were identified as a single bin corresponding to zero reads mapped; transcribed genomic regions were identified as bins with one or more reads mapped. Genomic regions encompassed by spliced reads (e.g. introns) were binned with transcript genomic regions. As defined by these parameters, 952,358,472-kb regions were identified as transcribed, and 1,966,089,658-kb regions were identified as nontranscribed.

**Oxidation-induced mutation analysis**

Cells were plated onto two 60-mm dishes for each of the four cell lines (SETD2-KO, MSH6-KO, PWWP-MT, and control...
Mismatch repair preferentially protects transcribed genes

HeLa cells) with 50% confluence. Cells in one dish were treated with 50 μM H₂O₂ for 6 days, and cells in the other dish were grown in H₂O₂-free medium. Genomic DNA was isolated and PCR-amplified using primers specific for three actively expressed genes (CD79A, MYC, and TP53) and two inactively expressed genes (CDT79A and GATA1). The length of PCR products is ~700 bp (Table S2). We subjected PCR products to TA cloning and randomly selected 20 positive colonies from each PCR for DNA-sequencing analysis. We aligned the sequences to the reference sequence to identify H₂O₂-induced mutations. Mutation percentage was calculated by dividing the mutation-containing colonies by the total colonies (n = 20), as shown in Table S2. Relative mutation frequency was calculated as shown in Table 1: relative mutation frequency = number of mutations/total sequencing length (20 × PCR length)/generations.

Statistical analysis

All statistical analyses were performed with Welch’s two sample t test using R 3.1.2 package. Data were considered statistically significant if p values were less than 0.05 or 0.01, as indicated.

Author contributions—Y. H. and G.-M. L. data curation; Y. H. formal analysis; Y. H. investigation; Y. H. writing-original draft; L. G. and G.-M. L. project administration; L. G. and G.-M. L. writing-review and editing; G.-M. L. conceptualization; G.-M. L. supervision; G.-M. L. project administration; L. G. and G.-M. L. writing-review and editing; G.-M. L. conceptualization; G.-M. L. supervision; G.-M. L. funding acquisition.

Acknowledgments—We thank Dong Wang and Guangshuo Ou for making their bioinformatics resource and microscopy available for the sequencing data analysis and immunofluorescence experiments, respectively; the NCBI dbGap (the Database of Genotypes and Phenotypes) committee for allowing us to use the HeLa genome database for this study; and Jonathan Feinberg for helpful comments.

References

1. Modrich, P., and Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu. Rev. Biochem. 65, 101–133 CrossRef Medline
2. Kolodner, R. (1996) Biochemistry and genetics of eukaryotic mismatch repair. Genes Dev. 10, 1433–1442 CrossRef Medline
3. Kunkel, T. A., and Erie, D. A. (2015) Eukaryotic mismatch repair in relation to DNA replication. Annu. Rev. Genet. 49, 291–313 CrossRef Medline
4. Li, G. M. (2008) Mechanisms and functions of DNA mismatch repair. Cell Res. 18, 85–98 CrossRef Medline
5. Constantin, N., Dzantiev, L., Kadyrov, F. A., and Modrich, P. (2005) Human mismatch repair: reconstitution of a nick-directed bidirectional reaction. J. Biol. Chem. 280, 39752–39761 CrossRef Medline
6. Zhang, Y., Yuan, F., Prensell, S. R., Tian, K., Gao, Y., Tomkinson, A. E., Gu, L., and Li, G. M. (2005) Reconstitution of 5'-directed human mismatch repair in a purified system. Cell 122, 693–705 CrossRef Medline
7. Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G. M. (2013) The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. Cell 153, 590–600 CrossRef Medline
8. Edmunds, J. W., Mahadevan, L. C., and Clayton, A. L. (2008) Dynamic histone H3 methylation during gene induction: HYB/PSet2 mediates all H3K36 trimethylation. EMBO J. 27, 406–420 CrossRef Medline
9. Fang, D., Gan, H., Lee, J. H., Han, J., Wang, Z., Riester, S. M., Jin, L., Chen, J., Zhou, H., Wang, J., Zhang, H., Yang, N., Bradley, E. W., Ho, T. H., Rubin, B. P., et al. (2016) The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. Science 352, 1344–1348 CrossRef Medline
10. Lu, C., Jain, S. U., Hoepeler, D., Bechet, D., Molden, R. C., Ran, L., Murphy, D., Venneti, S., Hameed, M., Patel, B. R., Wunder, J. S., Dickson, B. C., Lundgren, S. M., Jani, K. S., De Jay, N., et al. (2016) Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. Science 352, 844–849 CrossRef Medline
11. Wagner, E. J., and Carpenter, P. B. (2012) Understanding the language of Lys36 methylation at histone H3. Nat. Rev. Mol. Cell Biol. 13, 115–126 CrossRef Medline
12. Mellon, I., Rajpal, D. K., Koi, M., Boland, C. R., and Champhe, G. N. (1996) Transcription-coupled repair deficiency and mutations in human mismatch repair genes. Science 272, 557–560 CrossRef Medline
13. Rochette, P. J., Bastien, N., McKay, B. C., Therrien, J. P., Drobsketsby, E. A., and Drouin, R. (2002) Human cells bearing homozygous mutations in the DNA mismatch repair genes hMLH1 and hMSH2 are fully proficient in transcription-coupled nucleotide excision repair. Oncogene 21, 5745–5752 CrossRef Medline
14. Sonneveld, E., Vrieling, H., Mullenders, L. H., and van Hoffen, A. (2001) Mouse mismatch repair gene Msh2 is not essential for transcription-coupled repair of UV-induced cyclobutane pyrimidine dimers. Oncogene 20, 538–541 CrossRef Medline
15. Ni, T. T., Marsischky, G. T., and Kolodner, R. D. (1999) MSH2 and MSH6 are required for removal of adenine misincorporated opposite 8-oxo-guanine in S. cerevisiae. Mol. Cell 4, 439–444 CrossRef Medline
16. Li, G. M. (1999) The role of mismatch repair in DNA damage-induced apoptosis. Oncol. Res. 11, 393–400 CrossRef Medline
17. Song, Q., and Smith, A. D. (2011) Identifying dispersed epigenomic domains from ChIP-Seq data. Bioinformatics 27, 870–871 CrossRef Medline
18. Vermeulen, M., Eberl, H. C., Matarrese, F., Marks, H., Denissov, S., Butler, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G., and Mann, M. (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell 142, 967–980 CrossRef Medline
19. Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837 CrossRef Medline
20. Woo, Y. H., and Li, W. H. (2012) DNA replication timing and selection shape the landscape of nucleotide variation in cancer genomes. Nat. Commun. 3, 1004 CrossRef Medline
21. Liu, L., De, S., and Michor, F. (2013) DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes. Nat. Commun. 4, 1502 CrossRef Medline
22. 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 CrossRef Medline
23. Holmes, J., Jr, Clark, S., and Modrich, P. (1990) Strand-specific mismatch correction in nuclear extracts of human and Dro sophila melanogaster cell lines. Proc. Natl. Acad. Sci. U.S.A. 87, 5837–5841 CrossRef Medline
24. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bollli, N., Borg, A., Borresen-Dale, A. L., Boyault, S., Burkhhardt, B., Butler, A. P., Caldas, C., Davies, H. R., et al. (2013) Signatures of mutational processes in human cancer. Nature 500, 415–421 CrossRef Medline
25. Jinks-Robertson, S., and Bhagwat, A. S. (2014) Transcription-associated mutagenesis. Annu. Rev. Genet. 48, 111–139 CrossRef Medline
26. Haradhvala, N. J., Polak, P., Stojanov, P., Covictong, K. R., Shinbrot, E., Hess, J. M., Rheinbay, E., Kim, J., Maruzka, Y. E., Braunstein, L. Z., Kam-burov, A., Hanawalt, P. C., Wheeler, D. A., Koren, A., Lawrence, M. S., and Getz, G. (2016) Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. Cell 164, 538–549 CrossRef Medline
27. Ikehata, H., and Ono, T. (2007) Significance of CpG methylation for solar UV-induced mutagenesis and carcinogenesis in skin. Photochem. Photobiol. 83, 196–204 CrossRef Medline
28. Pfeifer, G. P. (2006) Mutagenesis at methylated CpG sequences. Annu. Rev. Biochem. 75, 119–139 CrossRef Medline
29. Poulos, R. C., Olivier, J., and Wong, J. W. (2017) The interaction between cytosine methylation and processes of DNA replication and repair shape...
Mismatch repair preferentially protects transcribed genes

The mutational landscape of cancer genomes. *Nucleic Acids Res.* 45, 7786–7795 CrossRef Medline

30. Zheng, C. L., Wang, N. J., Chung, J., Moslehi, H., Sanborn, J. Z., Hur, J. S., Collisson, E. A., Vemula, S. S., Naujokas, A., Chriott, K. E., Cheng, J. B., Fasshi, H., Blumberg, A. J., Bailey, C. V., Fedem, G. M., et al. (2014) Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. *Cell Rep.* 9, 1228–1234 CrossRef Medline

31. Shah, K. A., McGinty, R. J., Egorova, V. I., and Mirkin, S. M. (2014) Coupling transcriptional state to large-scale repeat expansions in yeast. *Cell Rep.* 9, 1594–1602 CrossRef Medline

32. Hamperl, S., and Cimprich, K. A. (2014) The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair* 19, 84–94 CrossRef Medline

33. Costantino, L., and Koshland, D. (2015) The Yin and Yang of R-loop biology. *Curr. Opin. Cell Biol.* 34, 39–45 CrossRef Medline

34. Forbes, S. A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, I. W., Futreal, P. A., and Stratton, M. R. (2008) The catalogue of somatic mutations in cancer (COSMIC). *Crit. Proct. Hum. Genet.* 2008 Chapter 10, Unit 11 CrossRef Medline

35. Chen, K., Chen, Z., Wu, D., Zhang, L., Lin, X., Su, J., Rodriguez, B., Xi, Y., Xia, Z., Chen, X., Shi, X., Wang, Q., and Li, W. (2015) Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat. Genet.* 47, 1149–1157 CrossRef Medline

36. Supek, F., and Lehner, B. (2015) Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* 521, 81–84 CrossRef Medline

37. Supek, F., and Lehner, B. (2017) Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. *Cell* 170, 534–547 CrossRef Medline

38. Supek, F., Mišta, B., Valcárcel, J., Gabaldón, T., and Lehner, B. (2014) Synonymous mutations frequently act as driver mutations in human cancers. *Cell* 156, 1324–1335 CrossRef Medline

39. Mellon, I., Spivak, G., and Hanawalt, P. C. (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* 51, 241–249 CrossRef Medline

40. Hanawalt, P. C., and Spivak, G. (2008) Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* 9, 958–970 CrossRef Medline

41. Mazurek, A., Berardini, M., and Fischel, R. (2002) Activation of human MutS homologs by 8-oxo-guanine DNA damage. *J. Biol. Chem.* 277, 8260–8266 CrossRef Medline

42. Mu, D., Tursun, M., Deckert, D. R., Drummond, J. T., Modrich, P., and Sancar, A. (1997) Recognition and repair of compound DNA lesions (base damage and mismatch) by human mismatch repair and excision repair systems. *Mol. Cell. Biol.* 17, 760–769 CrossRef Medline

43. Wang, H., Lawrence, C. W., Li, G. M., and Hays, J. B. (1999) Specific binding of human MSH2-MSH6 mismatch-repair protein heterodimers to DNA incorporating thymine- or uracil-containing UV light photoproducts opposite mismatched bases. *J. Biol. Chem.* 274, 16894–16900 CrossRef Medline

44. Gu, L., Wu, J., Qiu, L., Jennings, C. D., and Li, G. M. (2002) Involvement of DNA mismatch repair in folate deficiency-induced apoptosis small star, filled. *J. Nutr. Biochem.* 13, 355–363 CrossRef Medline

45. Hiratani, I., Ryba, T., Itoh, M., Rathjen, J., Kulik, M., Papp, B., Fussner, E., Bazzet-Jones, D. P., Plath, K., Dalton, S., Rathjen, P. D., and Gilbert, D. M. (2010) Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome Res.* 20, 155–169 CrossRef Medline

46. Hiratani, I., Ryba, T., Itoh, M., Yok ochi, T., Schwaiger, M., Chang, C. W., Lyou, Y., Townes, T. M., Schübler, D., and Gilbert, D. M. (2008) Global reorganization of replication domains during embryonic stem cell differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 333–338 CrossRef Medline

47. Lubelsky, Y., Prinz, J. A., DeNapoli, L., Li, Y., Belsky, J. A., and MacAlpine, D. M. (2014) DNA replication and transcription programs respond to the same chromatin cues. *Genome Res.* 24, 1102–1114 CrossRef Medline