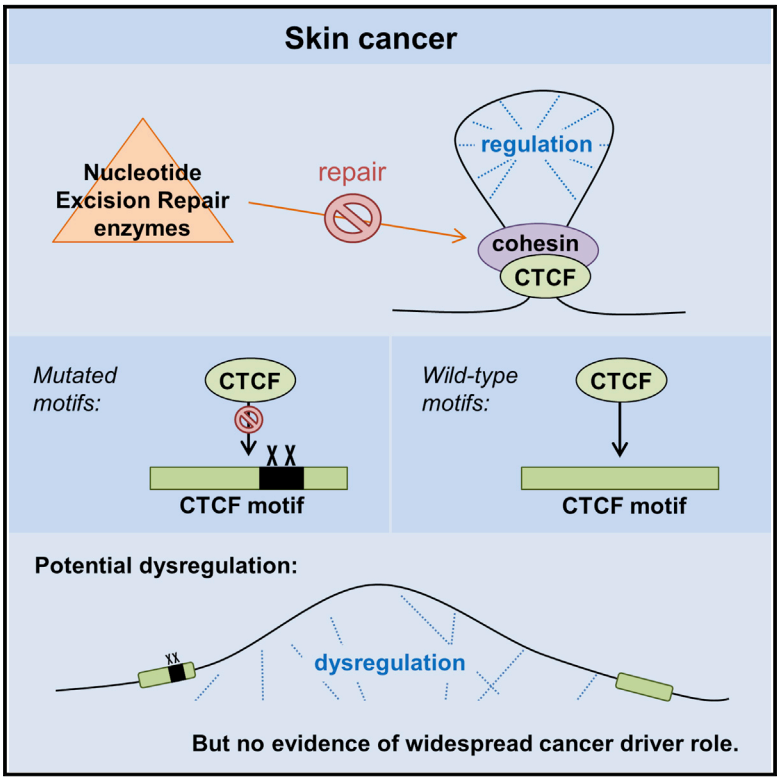
Cell Reports

Report

Functional Mutations Form at CTCF-Cohesin Binding Sites in Melanoma Due to Uneven Nucleotide Excision Repair across the Motif

Graphical Abstract



Highlights

- Mutations localize to specific nucleotides within CTCF binding sites in skin cancer
- This mutation pattern forms due to differential NER across pyrimidine pairs
- Mutations on highly conserved bases alter CTCF binding
- CTCF binding site mutations in melanoma are generally under neutral selection

Authors

Rebecca C. Poulos, Julie A.I. Thoms, Yi Fang Guan, Ashwin Unnikrishnan, John E. Pimanda, Jason W.H. Wong

Correspondence

jason.wong@unsw.edu.au

In Brief

Poulos et al. identify a specific mutation pattern within CTCF binding sites in skin cancers, which is attributable to differential nucleotide excision repair across the motif. Mutations on highly conserved bases cause allele-specific reduction of CTCF binding. Despite the frequency of these mutations, they are generally under neutral selection in melanoma.

Accession Numbers

GSE81945

Poulos et al., 2016, Cell Reports *17*, 2865–2872
December 13, 2016
http://dx.doi.org/10.1016/j.celrep.2016.11.055

CellPress

Cell Reports
Report

Functional Mutations Form at CTCF-Cohesin Binding Sites in Melanoma Due to Uneven Nucleotide Excision Repair across the Motif

Rebecca C. Poulos,[1] Julie A.I. Thoms,[1] Yi Fang Guan,[1] Ashwin Unnikrishnan,[1] John E. Pimanda,[1,2] and Jason W.H. Wong[1,3,4,*]
[1]Prince of Wales Clinical School and Lowy Cancer Research Center, UNSW Australia, Sydney, NSW 2052, Australia
[2]Department of Haematology, Prince of Wales Hospital, Sydney, NSW 2052, Australia
[3]Lead Contact
[4]Twitter: @jasonwwong
*Correspondence: jason.wong@unsw.edu.au
http://dx.doi.org/10.1016/j.celrep.2016.11.055

SUMMARY

CTCF binding sites are frequently mutated in cancer, but how these mutations accumulate and whether they broadly perturb CTCF binding are not well understood. Here, we report that skin cancers exhibit a highly specific asymmetric mutation pattern within CTCF motifs attributable to ultraviolet irradiation and differential nucleotide excision repair (NER). CTCF binding site mutations form independently of replication timing and are enriched at sites of CTCF/cohesin complex binding, suggesting a role for cohesin in stabilizing CTCF-DNA binding and impairing NER. Performing CTCF ChIP-seq in a melanoma cell line, we show CTCF binding site mutations to be functional by demonstrating allele-specific reduction of CTCF binding to mutant alleles. While topologically associating domains with mutated CTCF anchors in melanoma contain differentially expressed cancer-associated genes, CTCF motif mutations appear generally under neutral selection. However, the frequency and potential functional impact of such mutations in melanoma highlights the need to consider their impact on cellular phenotype in individual genomes.

INTRODUCTION

CCCTC-binding factor (CTCF) is an architectural protein involved in establishing the three-dimensional organization of the genome and defining genomic segments by gene expression (Holwerda and de Laat, 2013). *CTCF* is a tumor suppressor that is commonly deleted or mutated in cancer (Filippova et al., 1998; Kemp et al., 2014; Rakha et al., 2006). Recently, CTCF/cohesin binding sites have been found to be highly mutated in cancer (Ji et al., 2016; Kaiser et al., 2016; Katainen et al., 2015; Sabarinathan et al., 2016; Umer et al., 2016). The majority of research to date has been on CTCF binding site mutations in gastrointestinal cancers, with associations established between motif mutations and late replication, presence of cohesin, and signature 17 (Kaiser et al., 2016; Katainen et al., 2015; Umer et al., 2016).

Mutations in melanoma and other UV-light-induced skin cancers are driven by unique mutational processes associated with UV light, and many mutations form at transcription factor binding sites in these cancers because protein-binding impairs access by nucleotide excision repair (NER) enzymes (Perera et al., 2016; Sabarinathan et al., 2016). Despite CTCF binding sites being highly mutated in melanoma (Sabarinathan et al., 2016), the mutation pattern at the nucleotide level, together with additional factors contributing to mutation accumulation at CTCF binding sites have yet to be examined. Importantly, the functional consequences of these highly prevalent mutations in skin cancer remain unknown.

In this study, we examined three types of skin cancer: melanoma (n = 36), skin squamous cell carcinoma (SCC; NER-proficient $XPC^{wild\text{-}type}$: n = 8; NER-deficient $XPC^{-/-}$: n = 5), and skin basal cell carcinoma (BCC; n = 3), to comprehensively evaluate CTCF binding site mutations across UV-light-induced cancers. We examined the mutation profile within the CTCF binding motif, identifying factors underlying mutation accumulation, which associate with NER impairment. We also assessed the functional impact of such mutations on CTCF binding in cells through allele-specific chromatin immunoprecipitation sequencing (ChIP-seq) and, using computational analyses, show that CTCF binding site mutations are generally under neutral selection in melanoma.

RESULTS AND DISCUSSION

Mutation Rates at CTCF Motifs in UV-Light-Induced Cancers

We defined a 13-bp CTCF motif for our analyses as follows: CCA(C|G)(C|T)AG(A|G)(G|T)GGC(A|G) (shown in Figure 1C). This strict motif allowed us to minimize false-positive binding sites which may introduce noise, and to exclude the three low information-content nucleotides at the 5′ and 3′ ends of the consensus CTCF motif from the JASPAR database (Mathelier

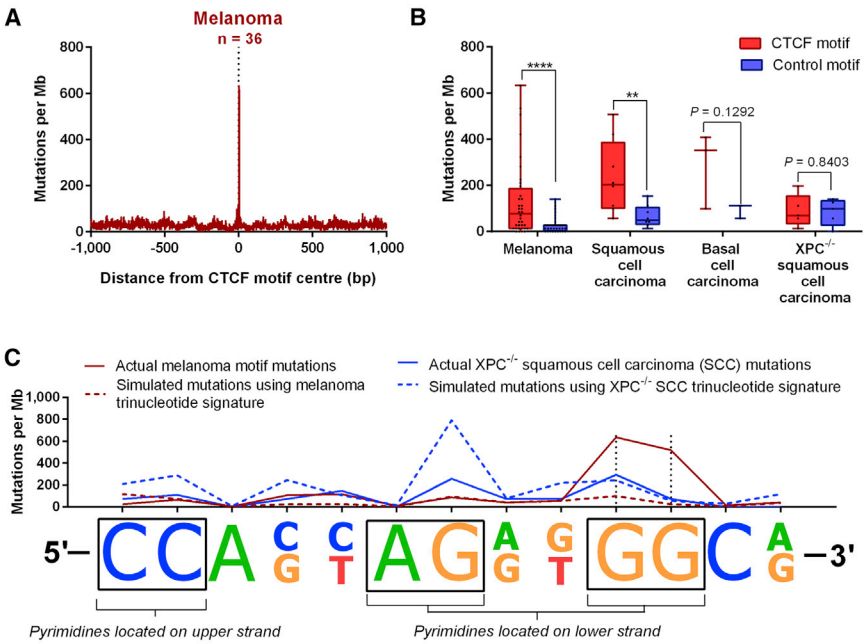


Figure 1. Prevalence of Mutations at CTCF Motifs in UV-Light-Induced Cancers

(A) Mutation density per melanoma sample (n = 36) for region ±1 kb of CTCF motif center.

(B) Mutation rate at CTCF and control motifs in melanoma and skin squamous, basal, and nucleotide excision repair (NER)-deficient XPC$^{-/-}$ squamous cell carcinomas (SCC). Boxplot shows median, quartiles, maximum, and minimum. Black dots indicate individual cancer samples. Significance is by paired t test for each sample.

(C) Mutation density per sample in melanoma and XPC$^{-/-}$ SCC at CTCF motifs and mutations simulated using the trinucleotide signature of melanoma or XPC$^{-/-}$ SCC. Locations of pyrimidine pairs present in all motifs are indicated.

(D and E) Normalized repair reads for (D) cyclobutane pyrimidine dimer (CPD) and (E) (6-4)pyrimidine-pyrimidone photoproduct ((6-4)PP) XR-seq across pyrimidine pairs.

Significance is by paired t test between pyrimidine pairs, with only significant associations shown. ****p < 0.0001, ***p < 0.001, **p < 0.01, and *p < 0.05. See also Figure S1.

et al., 2015). To select for bound motifs, we chose motifs within CTCF ChIP-seq peaks from normal human epidermal keratinocytes (NHEK) from ENCODE (ENCODE Project Consortium, 2012). Constructing mutation profiles centered upon these motifs (n = 5,470; Table S1), we observed an over 20-fold increase in mutation rate at CTCF motifs in melanoma (Figure 1A), with mutations similarly increased in other UV-light-induced cancers including NER-proficient SCC (Figure S1A) and BCC (Figure S1B). In contrast, NER-deficient XPC$^{-/-}$ SCCs showed no enrichment for mutations at CTCF motifs (Figure S1C), confirming previous findings that differential NER efficiency leads to CTCF mutation hotspots (Sabarinathan et al., 2016). (As XPC$^{-/-}$ SCCs showed no clear reduction of mutations at CTCF motifs (Figure S1C), this suggests that CTCF binding is unlikely to significantly inhibit DNA lesion formation in response to UV light.)

Across all melanoma, NER-proficient SCCs and BCCs, we found 91% (43/47) of samples to have at least one mutated CTCF motif (72.1% mean ±4.9% SD when mutations randomly shuffled). Of all CTCF motifs examined (n = 5,470), we found 8.6% of motifs (n = 468) mutated at least once across the 47 samples, with 52 of those motifs mutated twice (11.1%), and seven mutated three times (1.5%). Examination of recurrence at an individual nucleotide level revealed that, of all CTCF motif mutations (n = 534), 13 nucleotides were mutated in two samples, and one nucleotide was mutated in three samples. These rates of motif mutation recurrence were substantially higher than what would be expected by chance alone (125.7 motifs mutated at least once ±11.1 SD, 1.4 motifs mutated twice ±1.2 SD and 0.008 motifs mutated three times ±0.10 SD). This suggests that other factors such as trinucleotide context or access by NER machinery underlie the increased CTCF binding site mutation rate observed.

Differential NER Underlies Asymmetric Mutation Accumulation within CTCF Motifs in UV-Light-Induced Cancers

A specific conserved adenine central to the CTCF motif (that is, position 6; Figure 1C) becomes highly mutated in many cancer types (Kaiser et al., 2016; Katainen et al., 2015; Umer et al., 2016). These mutations are associated with a genome-wide mutation signature within the cancer sample consisting of high numbers of T > C/G mutations (Kaiser et al., 2016; Katainen et al., 2015). However, such a mutation signature does not feature prevalently in the genomes of UV-light-induced cancers (Alexandrov et al., 2013). Instead, we found a highly conserved guanine pair (positions 10 and 11) to harbor almost 70% (373/534) of CTCF motif mutations across melanoma (Figure 1C), SCC (Figure S1D), and BCC (Figure S1E).

Mutations simulated to account for trinucleotide mutation preferences showed the mutation pattern at CTCF motifs not to be due to sequence context alone (Figures 1C, S1D, and S1E). To account for any sequence-dependent mutation rate variation beyond trinucleotide context, we defined a control

motif (see Figure S1F) within NHEK CTCF ChIP-seq peaks. We observed only a small increase in mutation rate at the 3′ guanine pair in replication timing-matched control motif (Table S1; Figures S2D–S2J), likely attributable to significantly lower CTCF binding at control motifs than actual CTCF motifs ($p < 0.0001$, unpaired t test; Figure S1G). Further, the mutation rate was significantly lower in the control motif than the CTCF motif in melanoma ($p < 0.0001$; paired t test) and SCC ($p < 0.01$; paired t test) but not NER-deficient XPC$^{-/-}$ SCCs ($p = 0.8403$; Figure 1B), suggesting that CTCF binding rather than sequence context underlies the highly localized mutation pattern at CTCF motifs.

UV-light-induced cancers accumulate many mutations at CC/CT/TC dinucleotides (Cleaver and Crowley, 2002), and the CTCF motif consists of three conserved pyrimidine pairs (see Figure 1C). Given that XPC$^{-/-}$ SCC mutation density is broadly similar at each pyrimidine pair (Figure 1C), differential cyclobutane pyrimidine dimer (CPD) and (6-4)pyrimidine-pyrimidone photoproduct (6-4PP) formation is unlikely to account for the increased mutation load at the 3′ guanine pair. In contrast, using strand-specific CPD and (6-4)PP NER maps (excision repair sequencing [XR-seq]) (Adar et al., 2016), we found that over the 48 hr CPD (Figure 1D) and 4 hr (6-4)PP (Figure 1E) time courses, repair was consistently and significantly lower at the 3′ guanine pair than either the 5′ cytosine pair (CPD: $p < 0.05$ and (6-4)PP: $p < 0.01$, by paired t test) or the central AG (CPD and (6-4)PP: $p < 0.001$, by paired t test). While XR-seq data do not resolve the specific DNA lesion targeted by NER, nevertheless, these findings indicate that while NER is generally inhibited across entire regions of CTCF binding (Sabarinathan et al., 2016), differential rates of repair at specific pyrimidines within the CTCF motif may account for the asymmetric mutation accumulation observed at CTCF motifs in skin cancer.

Impairment of NER Is Dependent upon Cohesin at CTCF Binding Sites

In gastrointestinal cancers and leukemia, most CTCF motif mutations occur at sites of CTCF/cohesin binding (Ji et al., 2016; Katainen et al., 2015; Umer et al., 2016). However, the relationship between CTCF motif mutations and the cohesin-complex has not been established in UV-light-induced cancers. By defining a list of cohesin- and non-cohesin-bound CTCF motifs using ChIP-seq for RAD21 (a component of the cohesin complex) across eight cell types (Table S2), we found that, consistent with other cancer types, in melanoma, mutation density is only increased within cohesin-bound CTCF motifs (Figure 2A).

We next investigated the relationship between NER and the presence of the CTCF/cohesin complex. Given that the strength of binding of a transcription factor can impact the degree of NER impairment at a given locus (Perera et al., 2016; Sabarinathan et al., 2016), we investigated the influence of cohesin on the relationship between mutations and CTCF binding. Mutation density was significantly correlated with CTCF binding strength only at CTCF/cohesin binding sites ($p < 0.01$ for motifs with RAD21, $p = 0.8056$ for motifs without RAD21, Spearman's correlation; Figure 2B), though it is worth noting the low numbers of mutations in non-cohesin-bound motifs which may reduce the robustness of the latter correlation. Next, using XR-seq data (Hu et al., 2015), we constructed profiles of CPD repair surrounding CTCF motifs. We observed levels of NER to be high at sites immediately adjacent to CTCF binding, consistent with nucleosome depletion in these regions, making DNA more accessible to NER enzymes (Sabarinathan et al., 2016). Compared with these immediate flanks, we found NER to be significantly impaired only at CTCF motifs with cohesin binding (Figure 2C). NER was also more evidently abrogated at cohesin-bound CTCF motifs with high rather than low CTCF read coverage (Figure 2D). Our findings suggest that multifactorial impairment of NER, in response to both CTCF binding strength and the presence of a CTCF/cohesin protein complex, underlies mutation accumulation at CTCF motifs in UV-light-induced cancers.

CTCF Binding Site Mutations Form Independent of Replication Timing in Skin Cancers

In gastrointestinal cancers, CTCF binding site mutations have been associated with late replication (Kaiser et al., 2016; Katainen et al., 2015). However, in skin cancers, such mutations accumulate when CPDs and 6-4PPs are not repaired by NER prior to DNA replication (Perera et al., 2016; Sabarinathan et al., 2016), and so we hypothesized that replication timing may not greatly impact upon mutation accumulation at CTCF motifs. Consistent with our hypothesis, we found mutation accumulation to significantly correlate with replication timing only in motifs with low CTCF read coverage, but not in motifs with high coverage ($p < 0.05$ and $p = 0.9239$, respectively, Spearman's correlation; Figure 2E). Since accumulation of CTCF motif mutations is correlated with CTCF binding strength (Figure 2B), mutations with high CTCF binding would be more likely than those with low CTCF binding to be driven by NER impairment from CTCF binding than by the influences of late replication.

Impact of CTCF Motif Mutations on CTCF Binding in COLO829

Having identified the specific mutation pattern and a mechanism underlying its formation in UV-light-induced skin cancers, we sought to determine whether CTCF motif mutations affect CTCF binding. We performed CTCF ChIP-seq in the melanoma cell line COLO829. From replicate independent experiments, we identified 44,050 common CTCF peaks (Figure S2A) and, within these, 48,603 CTCF motifs (motifs accorded to the 19-bp consensus motif from the JASPAR database [Mathelier et al., 2015] [shown in Figure 3C]). We found most motifs to be positioned near the centers of CTCF peaks (Figure S2B), with some peaks containing multiple motifs.

Using mutations from COLO829 whole-genome sequencing (WGS) data (Craig et al., 2016), we identified 124 COLO829 mutations in CTCF peaks, of which 16 were within the CTCF motif (Figure 3A; Table S3). Mutations in CTCF peaks were generally distributed across the genome without a distinct pattern, consistent with the apparent random distribution of mutations in CTCF peaks in the WGS The Cancer Genome Atlas (TCGA) melanoma samples (Figure S2C). Of the 15 heterozygous COLO829 mutations in CTCF motifs, eight were at the 3′ guanine pair (Figures 3A and 3C; Table S3). Two sites harbored dinucleotide mutations within the same motif (Table S2).

To identify whether the heterozygous COLO829 CTCF motif mutations (Table S2) caused allele-specific loss of CTCF

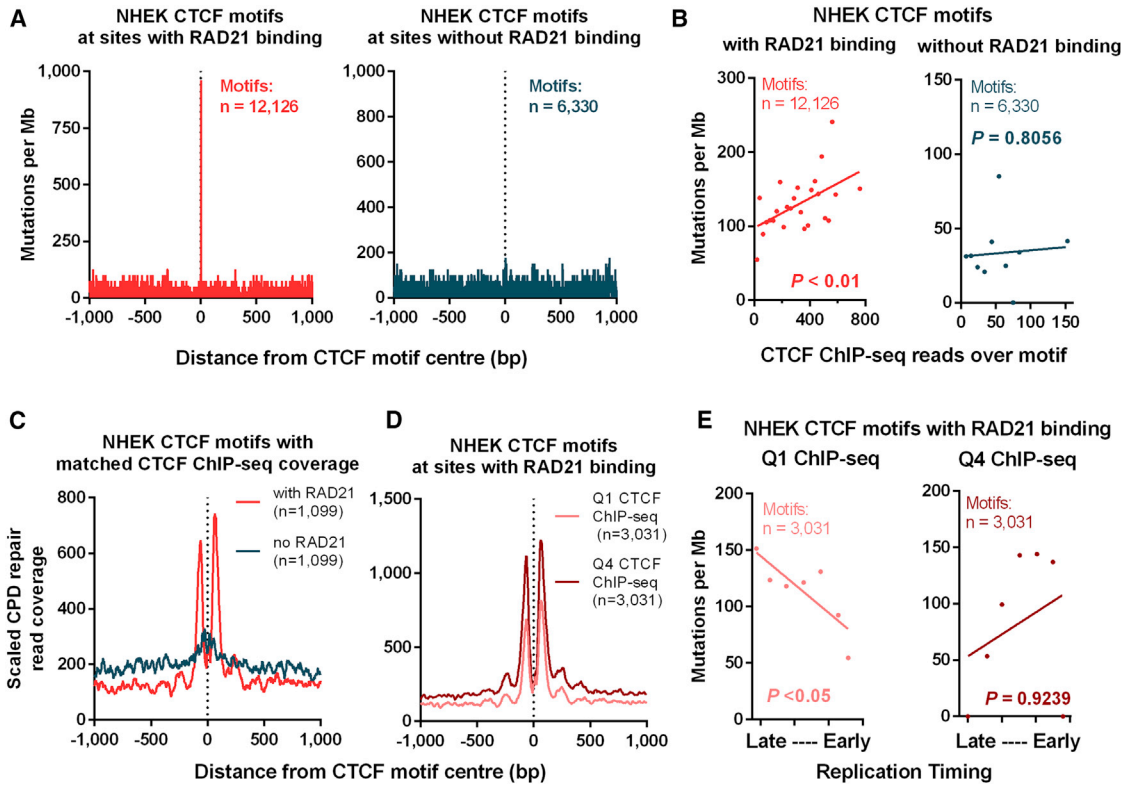


Figure 2. Mechanism Underlying CTCF Motif Mutation Accumulation in Melanoma

(A) Mutation density per sample for region ±1 kb of CTCF motif center in melanoma, for NHEK CTCF motifs with (left) and without (right) RAD21 binding.

(B) Correlation of mutation density and CTCF ChIP-seq read coverage over NHEK CTCF motifs with (left) and without (right) RAD21 binding.

(C and D) Nucleotide excision repair cyclobutane pyrimidine dimer (CPD) repair profiles using XR-seq (Hu et al., 2015) data for region ±1 kb of NHEK CTCF motifs for (C) motifs with matched CTCF reads with and without RAD21 binding and (D) motifs with RAD21 binding, separated into bottom (Q1) and top (Q4) quartiles of NHEK CTCF read coverage across the motif.

(E) Correlation of mutation rate and replication timing for bottom (Q1; left) and top (Q4; right) quartiles of NHEK CTCF read coverage, for CTCF motifs with RAD21 binding. Data points show binned data; significance is by Spearman's correlation on un-binned data.

binding, we examined the normalized variant allele frequency (normVAF) of mutant alleles from our ChIP-seq data (see Figure 3B and Experimental Procedures). We found all mutations at the guanine pair to correspond to an allele-specific decrease in CTCF binding (<50% normVAF), with six out of eight mutations with <10% normVAF, indicating a dramatic reduction of CTCF binding to mutant alleles (Figure 3C). We observed no allele-specific loss of CTCF binding in mutations beyond the most highly conserved 13-bp center of the motif (normVAF >50%; Figure 3C). Considering all mutations in CTCF peaks in COLO829, mutations within the conserved center 13 bp of the motif had significantly lower normVAFs than mutations only within CTCF ChIP-seq peaks (median of 7% versus 48%; $p < 0.0001$, unpaired t test; Figure 3D).

To further support our finding that CTCF motif mutations reduce CTCF binding, we performed in vitro DNA-affinity experiments using a wild-type and mutant CTCF motif (Figure S2D; see Experimental Procedures). Incubating DNA baits with nuclear protein extract from COLO829, we observed loss of CTCF binding to mutant motifs via western blot (Figure S2E).

Taken together, our experimental analyses demonstrate single point mutations at highly conserved sites within CTCF motifs, especially at the 3′ guanine pair, to be sufficient to reduce, or in some cases entirely abrogate, CTCF binding.

CTCF Motif Mutations as Potential Cancer Driver Events

Having established that CTCF motif mutations can alter CTCF binding, we next sought to determine whether such mutations may be potential melanoma drivers. Using topologically associating domain (TAD) anchor motifs from NHEK cells (Rao et al., 2014), we identified DNA loops with mutated anchor motifs (n = 343), and found cancer-associated genes within these loops (see Experimental Procedures). Of the 16 cancer-associated genes we identified, using matched TCGA melanoma RNA sequencing data, we found seven with significantly different expression in mutant samples (significance by $p < 0.05$; unpaired t test or one-sample t test, with Bonferroni method to correct for multiple testing; Figure S3A). Though it is worth noting the small melanoma sample size in our analyses (n = 36), making accurate determinations of significance difficult.

To investigate whether our findings would be expected by chance alone, we performed a bootstrapping analysis on significantly differentially expressed anchor-gene-loop groupings (see Experimental Procedures). The actual fraction of unique

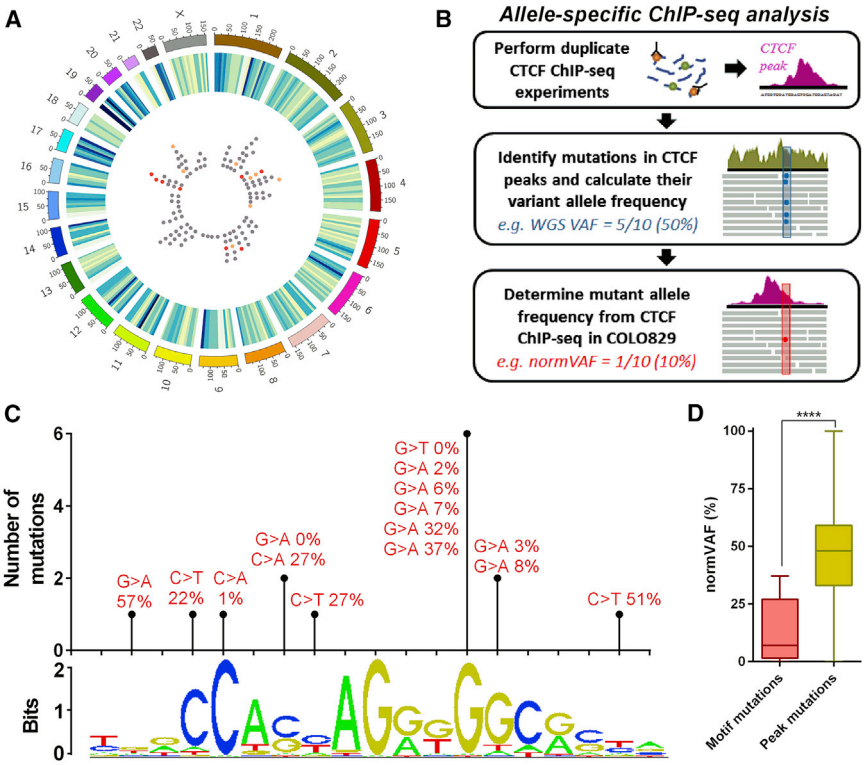


Figure 3. CTCF Motif Mutations and CTCF Binding in COLO829

(A) Circos plot (Krzywinski et al., 2009) of CTCF peak and motif mutations in COLO829. Outer track (heatmap) denotes density of bases of CTCF peaks per 10 Mb (low, yellow; high, blue). Inner track (dots) shows CTCF peak mutations (gray), motif mutations on the 3′ guanine pair (red) and other motif mutations (orange).

(B) Flow chart of allele-specific ChIP-seq analysis calculating normalized variant allele frequency (normVAF) for COLO829 mutations in CTCF peaks.

(C) Mutation count, type and normVAF from COLO829 ChIP-seq for heterozygous COLO829 mutations in CTCF consensus motifs (shown, per JASPAR database [Mathelier et al., 2015]).

(D) normVAF of CTCF motif mutations from COLO829 ChIP-seq, for all mutations within the center 13 bp of CTCF motifs and peak-only mutations. Boxplot shows median, quartiles, maximum, and minimum. Significance by unpaired t test. ****$p < 0.0001$.

See also Figure S2.

anchor-gene-loop groupings in melanoma with significantly different gene expression, for both cancer-associated and non-cancer-associated genes, fell well within the distribution from randomly shuffled mutations (Figures S3B and S3C). Further, when we compared the proportion of cancer-associated genes contained in TADs with either wild-type or mutant anchor motifs, we found no enrichment for cancer-associated genes in mutant loops in melanoma ($p = 0.8950$, Fisher's exact test), nor was the fraction of cancer-associated genes (2.84%; 16/564 genes) significantly different from the mean of the distribution of randomly shuffled mutations (2.90% mean ±0.59% SD; Figure 4A). We also found no significant difference in melanoma between the distance from mutated anchors of cancer-associated and non-cancer-associated genes ($p = 0.4132$, unpaired t test, Figure S3D).

Mutations at higher variant allele frequencies (VAFs) are present in greater proportions of cancer cells (Sidow and Spies, 2015) and may be more likely than those at lower VAFs to confer a growth advantage. Hence, we examined the VAFs of melanoma mutations in CTCF motifs and peaks, and the VAFs in our samples of the known melanoma driver mutation *BRAF* V600E. We found no significant difference between the VAFs of CTCF motif and peak mutations (Figure 4B), nor between CTCF anchor and non-anchor motif mutations (Figure S3E). However, these mutations generally had significantly lower VAFs than *BRAF* V600E mutations ($p < 0.0001$ or $p < 0.001$, unpaired t test; Figures 4B and S3E). It is yet possible that some CTCF motif mutations (perhaps those at higher VAFs) may act as cancer drivers in individual samples, but our results suggest that, in general, somatic mutations in CTCF motifs are under more neutral selection, and so the majority of such mutations are less likely to be cancer drivers.

While there is evidence that CTCF motif mutations in gastrointestinal cancers may be enriched near cancer-associated genes (Umer et al., 2016), our findings indicate that this is not generally the case in melanoma. Perhaps UV-light-induced cancers differ in this way from gastrointestinal cancers due to the different mechanism which underlies their formation. Specifically, the mechanism of differential NER operating in UV-light-induced cancers would not rely on selective pressure to drive mutation propagation within cells.

Even though we have shown that CTCF binding site mutations in melanoma are in general under neither positive nor negative selection, single non-recurrent somatic point mutations may still have oncogenic effects within an individual cell or cancer sample (Kim et al., 2016). For example, one significantly differentially expressed cancer-associated gene in a TAD with a mutated CTCF anchor motif is the tumor suppressor *APC*. *APC* had significantly lower expression in the mutant sample than wild-type samples (adjusted $p < 0.0001$, one-sample t test; Figure S3A). *APC* is a component of the Wnt signaling transduction pathway, and epigenetic silencing of *APC* may contribute to melanoma development by promoting cell proliferation (Worm et al., 2004). The observed low *APC* expression does not seem associated with copy number loss or nonsense-mediated decay, as the mutant melanoma sample identified has no copy number alterations at the *APC* locus (segment mean = −0.1306, approximating a copy number of two) and no coding mutations in the gene body. It is possible that, in this sample, the low *APC* expression may be due to dysregulation caused by disruption of loop formation by the mutated anchor motif (illustrated in Figure 4C), though this hypothesis will require further investigation. Our findings

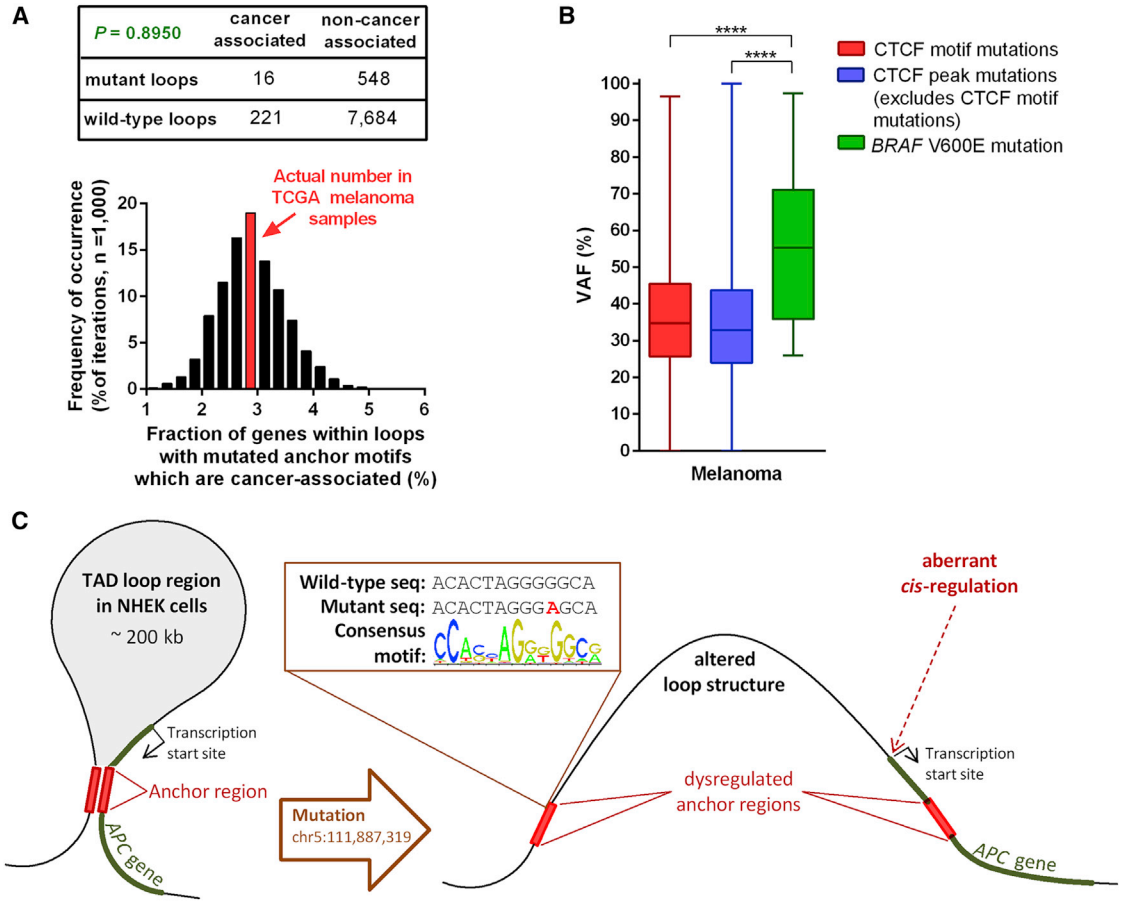


Figure 4. CTCF Motif Mutations and Study of Cancer Association in Melanoma

(A) Contingency table and p value (Fisher's exact test) for cancer-associated and non-cancer-associated genes within loops with mutant and wild-type topologically associated domain (TAD) anchor motifs. Histogram shows results from bootstrap (n = 1,000 iterations; data bins: 0.25% intervals) of proportion of cancer-associated genes in loops with mutated anchors when mutations randomly assigned across TAD anchor motifs. Actual proportion from The Cancer Genome Atlas (TCGA) melanomas is indicated.

(B) Variant allele frequency (VAF) of melanoma mutations within strict 13-bp CTCF motifs, CTCF ChIP-seq peaks from NHEK cells and *BRAF* V600E mutations. Boxplot shows median, quartiles, maximum, and minimum. Significance is by unpaired t test. ****$p < 0.0001$; only significant associations are shown.

(C) Schematic of potential dysregulation of *APC* gene resulting from mutation of anchor motif for the TAD harboring the *APC* transcription start site. Motif mutation from TCGA melanoma sample TCGA-D9-A148 shown, with CTCF motif from JASPAR database (Mathelier et al., 2015). Diagram not to scale.

See also Figure S3.

have important implications for how we define the function of non-coding mutations in cancer genomes. As individual UV-light-induced cancer genomes contain comparatively high numbers of CTCF motif mutations, due to their ability to disrupt CTCF binding, our study indicates the need for careful consideration of these mutational events in individual cancer genome analyses.

EXPERIMENTAL PROCEDURES

For further methodology, refer to Supplemental Experimental Procedures.

COLO829 Cell Culture

COLO-829 cells (cell line validation per Poulos et al., 2015) were cultured in RPMI medium (Life Technologies) supplemented with 10% fetal bovine serum, glutamax, and penicillin/streptomycin. ChIP-seq experiments were performed as previously described (Wilson et al., 2010).

Whole-Genome Sequencing Data Analysis

Somatic mutation data for melanoma were obtained from TCGA, with calls made from BAM files from CGHub (Wilks et al., 2014) using Strelka (Saunders et al., 2012) with default parameters. Somatic mutation data for SCC and BCC were obtained from published datasets (Zheng et al., 2014). COLO829 single nucleotide variant (SNV) mutation calls were also obtained from published datasets (Craig et al., 2016).

75-bp paired-end Illumina raw sequencing reads for the COLO829 melanoma cell line were obtained with permission from the EGA (https://www.ebi.ac.uk/ega/). Reads were aligned as previously described (Poulos et al., 2015).

ChIP-Sequencing Data Analysis

CTCF ChIP-seq data from NHEK cells were downloaded from ENCODE (ENCODE Project Consortium, 2012) (GEO: GSM749707 and GSM749747). RAD21 processed ChIP-seq "narrow peak" BED files were also downloaded from ENCODE (ENCODE Project Consortium, 2012) for eight human cell types (accession numbers in Table S2). Within each cell type, overlapping RAD21 peaks were merged.

COLO829 ChIP-seq data were aligned using Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009), with peaks called using the Homer Suite (Heinz et al., 2010).

Additional Datasets

XR-seq CPD repair data from UV-light-irradiated normal human skin fibroblast (NHF) cells (Hu et al., 2015) were obtained in Sequence Read Archive (SRA) format (GEO: GSM1659156) and processed as previously described (Hu et al., 2015). Time-course CPD and (6-4)PP XR-seq data for NHF cells (Adar et al., 2016) were obtained in SRA format (GEO: GSE76391) and processed per Supplemental Experimental Procedures. Repair reads over pyrimidine pairs (Figures 1D and 1E) were normalized by dividing the total strand-specific repair reads at each pyrimidine pair across all motifs, by the XR-seq reads at that time point and by total motifs (n = 5,470) and then multiplying by 10^9.

Replication timing data for NHEK cells were obtained from ENCODE (ENCODE Project Consortium, 2012) (GEO: GSM923445) as a bigwig file, with data extracted using "bigWigAverageOverBed" tool.

Melanoma gene expression was determined from processed (level 3) RNA sequencing (RNA-seq) data downloaded from TCGA data portal. Genes were designated cancer associated if listed on COSMIC's Cancer Gene Census (retrieved April 14, 2016, containing 571 unique cancer-associated genes). Transcription start sites (TSSs) for each gene were obtained from the UCSC table browser.

TAD anchors, anchor motifs, and loop regions were identified from published data of Hi-C experiments in NHEK cells (Rao et al., 2014) (GEO: GSE63525).

CTCF Motifs and Data Analysis

Mutation profiles were generated by counting mutations at each base within ±1 kb of a motif. Counts were then normalized to mutations per megabase per sample for each cancer type. Repair profiles counted the number of NER CPD reads in NHF cells (non-time-course data [Hu et al., 2015]) overlapping each base within ±1 kb of a motif. Repair read counts were scaled by multiplying total read count at each base by 1,000 (Figures 2C and 2D). All profiles were orientated so that the motif runs from 5′ to 3′.

Statistical Analysis

Significance in Figures 1B, 1D, and 1E is by paired t test between samples (Figure 1B) or pyrimidine pairs (Figures 1D and 1E). Significance in Figures 2B and 2E is by Spearman's correlation on un-binned data (graph shows binned data points). Figure 4A shows significance by Fisher's exact test. All other determinations of significance were by unpaired t test. For analyses of significantly differentially expressed cancer- and non-cancer-associated genes (Figures S3A–S3C), where appropriate, a one-sample t test was performed. Also for these analyses, a Bonferroni correction for multiple testing was applied to p values to determine significance.

ACCESSION NUMBERS

The accession number for the COLO829 CTCF ChIP raw sequencing and processed peak data are GEO: GSE81945.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.11.055.

AUTHOR CONTRIBUTIONS

R.C.P. performed experiments and bioinformatic analyses. J.A.I.T., Y.F.G., A.U., and J.E.P. were involved in experimental design. J.W.H.W. undertook project design and conceptualized the study. R.C.P. and J.W.H.W. wrote the manuscript with contributions from co-authors.

ACKNOWLEDGMENTS

The authors thank TCGA and other groups who have made data publically available and also thank Intersect Pty Ltd for high-performance computing resources and data storage used in this study. This work was funded by the National Health and Medical Research Council (APP1119932), Cancer Institute NSW (13/DATA/1-02), and the Cure Cancer Foundation Australia with the assistance of Cancer Australia through the Priority-driven Collaborative Cancer Research Scheme (APP1057921) to J.W., H.W., and R.C.P. is supported by an Australian Postgraduate Award and Y.F.G. by a Prince of Wales Clinical School postgraduate research scholarship. J.E.P. is funded by the NHMRC. J.W.H.W. is supported by an Australian Research Council Future Fellowship (FT130100096).

Received: September 7, 2016
Revised: November 14, 2016
Accepted: November 17, 2016
Published: December 13, 2016

REFERENCES

Adar, S., Hu, J., Lieb, J.D., and Sancar, A. (2016). Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. Proc. Natl. Acad. Sci. USA *113*, E2124–E2133.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Cleaver, J.E., and Crowley, E. (2002). UV damage, DNA repair and skin carcinogenesis. Front. Biosci. *7*, d1024–d1043.

Craig, D.W., Nasser, S., Corbett, R., Chan, S.K., Murray, L., Legendre, C., Tembe, W., Adkins, J., Kim, N., Wong, S., et al. (2016). A somatic reference standard for cancer genome sequencing. Sci. Rep. *6*, 24607.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Filippova, G.N., Lindblom, A., Meincke, L.J., Klenova, E.M., Neiman, P.E., Collins, S.J., Doggett, N.A., and Lobanenkov, V.V. (1998). A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. Genes Chromosomes Cancer *22*, 26–36.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Holwerda, S.J.B., and de Laat, W. (2013). CTCF: The protein, the binding partners, the binding sites and their chromatin loops. Philos. Trans. R. Soc. Lond. B Biol. Sci. *368*, 20120369.

Hu, J., Adar, S., Selby, C.P., Lieb, J.D., and Sancar, A. (2015). Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. Genes Dev. *29*, 948–960.

Ji, X., Dadon, D.B., Powell, B.E., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G., Lee, T.I., et al. (2016). 3D chromosome regulatory landscape of human pluripotent cells. Cell Stem Cell *18*, 262–275.

Kaiser, V.B., Taylor, M.S., and Semple, C.A. (2016). Mutational biases drive elevated rates of substitution at regulatory sites across cancer types. PLoS Genet. *12*, e1006207.

Katainen, R., Dave, K., Pitkänen, E., Palin, K., Kivioja, T., Välimäki, N., Gylfe, A.E., Ristolainen, H., Hänninen, U.A., Cajuso, T., et al. (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. Nat. Genet. *47*, 818–821.

Kemp, C.J., Moore, J.M., Moser, R., Bernard, B., Teater, M., Smith, L.E., Rabaia, N.A., Gurley, K.E., Guinney, J., Busch, S.E., et al. (2014). CTCF

haploinsufficiency destabilizes DNA methylation and predisposes to cancer. Cell Rep. *7*, 1020–1029.

Kim, E., Ilic, N., Shrestha, Y., Zou, L., Kamburov, A., Zhu, C., Yang, X., Lubonja, R., Tran, N., Nguyen, C., et al. (2016). Systematic functional interrogation of rare cancer variants identifies oncogenic alleles. Cancer Discov. *6*, 714–726.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.-y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2015). JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *44*, D110–D115.

Perera, D., Poulos, R.C., Shah, A., Beck, D., Pimanda, J.E., and Wong, J.W.H. (2016). Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. Nature *532*, 259–263.

Poulos, R.C., Thoms, J.A., Shah, A., Beck, D., Pimanda, J.E., and Wong, J.W. (2015). Systematic screening of promoter regions pinpoints functional cis-regulatory mutations in a cutaneous melanoma genome. Mol. Cancer Res. *13*, 1218–1226.

Rakha, E.A., Green, A.R., Powe, D.G., Roylance, R., and Ellis, I.O. (2006). Chromosome 16 tumor-suppressor genes in breast cancer. Genes Chromosomes Cancer *45*, 527–535.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A., and López-Bigas, N. (2016). Nucleotide excision repair is impaired by binding of transcription factors to DNA. Nature *532*, 264–267.

Saunders, C.T., Wong, W.S., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics *28*, 1811–1817.

Sidow, A., and Spies, N. (2015). Concepts in solid tumor evolution. Trends Genet. *31*, 208–214.

Umer, H.M., Cavalli, M., Dabrowski, M.J., Diamanti, K., Kruczyk, M., Pan, G., Komorowski, J., and Wadelius, C. (2016). A significant regulatory mutation burden at a high-affinity position of the CTCF motif in gastrointestinal cancers. Hum. Mutat. *37*, 904–913.

Wilks, C., Cline, M.S., Weiler, E., Diehkans, M., Craft, B., Martin, C., Murphy, D., Pierce, H., Black, J., Nelson, D., et al. (2014). The Cancer Genomics Hub (CGHub): Overcoming cancer through the power of torrential data. Database (Oxford) *2014*, bau093.

Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E., et al. (2010). Combinatorial transcriptional control in blood stem/progenitor cells: Genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell *7*, 532–544.

Worm, J., Christensen, C., Grønbaek, K., Tulchinsky, E., and Guldberg, P. (2004). Genetic and epigenetic alterations of the APC gene in malignant melanoma. Oncogene *23*, 5215–5226.

Zheng, C.L., Wang, N.J., Chung, J., Moslehi, H., Sanborn, J.Z., Hur, J.S., Collisson, E.A., Vemula, S.S., Naujokas, A., Chiotti, K.E., et al. (2014). Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. Cell Rep. *9*, 1228–1234.