Published in final edited form as:

Nat Biotechnol. 2014 January ; 32(1): 71–75. doi:10.1038/nbt.2778.

Reduced local mutation density in regulatory DNA of cancer genomes is linked to DNA repair

Paz Polak[1,2,3,*], **Michael S. Lawrence**[3,*], **Eric Haugen**[4], **Nina Stoletzki**[1,2,3], **Petar Stojanov**[3], **Robert E Thurman**[4], **Levi A. Garraway**[2,3,5,6], **Sergei Mirkin**[7], **Gad Getz**[3], **John A. Stamatoyannopoulos**[4,#], and **Shamil R. Sunyaev**[1,2,3,#]

[1]Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Boston, MA, 02115

[2]Harvard Medical School, Boston, MA 02115, USA

[3]The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[4]Departments of Genome Sciences and Medicine (Oncology), University of Washington, Seattle, WA 98195, USA

[5]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

[6]Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA 02115, USA

[7]Department of Biology, Tufts University, Medford, MA 02155, USA

Abstract

Carcinogenesis and neoplastic progression are mediated by the accumulation of somatic mutations. Here we report that the local density of somatic mutations in cancer genomes is highly reduced specifically in accessible regulatory DNA defined by DNase I hypersensitive sites. This reduction is independent of any known factors influencing somatic mutation density and is observed in diverse cancer types, suggesting a general mechanism. By analyzing individual cancer genomes[1], we show that the reduced local mutation density within regulatory DNA is linked to intact global genome repair machinery, with nearly complete abrogation of the hypomutation phenomenon in individual cancers that possess mutations in multiple nucleotide excision repair components. Together, our results connect chromatin structure, gene regulation and cancer-associated somatic mutation.

Somatic mutations are a major contributor to cancer development and progression. In cancer cells, the density of somatic mutations is highly heterogeneous along the genome[2,3]. However, mechanisms governing the genomic distribution of somatic mutations are poorly understood. Recently, cancer genomics efforts have accumulated data on somatic mutations

[#]Correspondence: ssunyaev@rics.bwh.harvard.edu or jstam@u.washington.edu.
[#]These authors contributed equally to this work.

Author contribution

P.P. M.S.L., N.S., G.G. J.S, and S.R.S conceived the study. P.P. led the analysis of sequencing data. M.S.L. and P.S. contributed to the analysis of sequencing data. R.E.T and E.H. analyzed DHS data. L.A.G. provided melanoma sequencing data. S.M., G.G., J.A.S. and S.R.S. supervised the analysis. P.P., S.M., J.A.S. and S.R.S. wrote the manuscript.

in tumors[4], revealing that the relative density of somatic mutations in protein coding genes (including both introns and exons) is lower than the genome average[5]. This effect has been posited to result from transcription-coupled DNA repair (TCR)[2,3], which is mediated by the recruitment of the nucleotide excision repair (NER) system by Pol II RNA polymerase stalled at pre-mutation lesions[6,7]. The existence of such an effect raises the question whether other similarly specialized repair mechanisms operate on other functionally important genomic regions.

Regulatory DNA (promoters, enhancers, insulators, etc.) active within a given cell type is characterized by hypersensitivity to DNase I[8], resulting in DNase I hypersensitive sites (DHSs) that quantitatively reflect regulatory factor binding in place of canonical nucleosomes[9,10]. It has long been posited that the accessibility of DNA within regulatory regions may render such regions more susceptible to DNA damage-induced mutation[11]. Evolutionary rates of sequence divergence within DHS found in cancer genomes and primitive cells are higher than normal differentiated cells[8], and density of somatic variants detected in a cancer sample that underwent cell culture was shown to be reduced in DHS more than density of common SNPs[12]. However, particularly in view of the variability in somatic mutation rates along cancer genomes, a quantitative understanding of mutation within regulatory DNA, together with insight into the underlying biological mechanisms, has not been explored.

Results

Reduced local density of somatic mutations in DHSs

To examine mutation frequencies in regulatory DNA, we mapped DHSs genome-wide in 12 cancer cell lines, as well as normal cellular counterparts of major malignancies (see Methods). We then analyzed whole-genome sequencing data from 34 tumor/normal pairs from seven distinct datasets: small-cell lung cancer[3], melanoma[2], 23 multiple myeloma[5] (MM) samples, and 9 colon cancers[13]. We used published mutation data for small-cell lung cancer[3] and melanoma cell lines[2] (http://icgc.org) and re-analyzed primary tumor data on multiple myeloma and colon cancer using MuTect [14] (http://www.broadinstitute.org/cancer/cga/mutect). These 34 cancer genomes contained 364,226 somatic point mutations in about 2.6 Gbp of sequence that could be uniquely mapped in the DHSs assay, *i.e.* density of 0.000139 per base-pair (bp).

We observed a substantial reduction in the frequency of somatic nucleotide substitutions in DHSs compared to the genome average (Fig. 1 and Supplementary Fig. 1). This reduction is highly significant and consistent across all tumors ($P < 10^{-36}$, chi-square test). The reduction was most prominent in the core TF binding regions of DHSs marked by the maxima of DNase I cleavage intensity (Fig. 1).

We next confirmed that the reduction of frequency of somatic mutations in DHSs was not the result of confounding factors influencing local variation in cancer mutation density, nor the result of sequencing and mapping biases[15]. Confounding factors may include differences between intergenic regions and genes (including both exons and introns), distance from transcription start sites[2] (Supplementary Fig. 2), time of DNA replication during the Sphase[16], distances to telomeres and centromeres, and local G+C content[15]. Relative density of somatic mutations also depends on sequence context, especially flanking nucleotides, and different tumors exhibit different context dependencies[2,3,13] (Supplementary Fig. 3). The relative density of mutations expected from the sequence context is higher in DHSs, magnifying our observation ($P < 5*10^{-181}$).

We observed significant relative reduction of somatic mutations in DHSs located in both intronic and intergenic regions, and in DHSs proximal (<1 kb) to transcription start sites versus more distal DHSs (Supplementary Fig. 4). More notably, the reduction was evident in comparison to immediately flanking 1kb regions ($P < 2.2*10^{-16}$, chi-square test in lung, MM and colon; $P=7.21*10^{-13}$ in melanoma). As such, the observed reduction in the density of somatic mutations cannot be explained by a regional factor acting over long ranges[17] such as transcription or DNA replication timing.

To rule out biases related to sequencing and mapping, for two of the cancer types (colon[13] and MM[5]) with available raw sequencing data we repeated the analysis restricting it to nucleotide positions with above 80% detection power based on sequencing coverage. This analysis confirmed that the density of somatic sequence alterations is significantly reduced in DHSs compared to 1kb flanking sequences.

To account collectively for all of the above potential confounding factors, we applied Poisson regression model[18]. DHSs remained a significant and substantial contributor to the local somatic mutation frequency on top of other factors, including DNA replication timing[19], distance from transcription start sites, distance from the DHS itself, CpG islands status, G+C content, and region type (exonic, intergenic, intronic) (Supplementary Tables 1 and 2). Because our regression analysis included neighboring windows, short-range regional dependencies could potentially inflate statistical significance. We repeated the analysis using a small subset (20%) of spatially separated windows and confirmed that DHSs remain a highly significant contributor even if only 20% of data are used (Supplementary Tables 1 and 2).

Notably, the effect of chromatin accessibility is monotonic and continuous and thus does not depend on the specific thresholds used to define DHSs (Supplementary Fig. 5). Finally, DHSs mapped in potential cells or tissues of origin (e.g., lung tissue for lung carcinoma, etc; see Methods) substantially contribute to the regression model that already includes pooled DHSs from multiple cell types. This demonstrates that cell-selective chromatin architecture and not simply genomic location is the driving feature.

The observed reduction in the frequency of somatic sequence changes within DHSs might be explained by either reduced occurrence of somatic mutation or by the action of purifying selection. At present, purifying selection in cancers has not been carefully studied, so we lack information that would support or contradict the action of purifying selection. In general, population genetics and comparative genomics studies in a variety of organisms suggest that purifying selection is usually stronger in coding regions than in regulatory regions[20,21]. To investigate the possible action of purifying selection, we compared relative mutation densities in regulatory and protein-coding regions. The average reduction (across cancers) in frequency of somatic mutations in coding sequences relative to flanking sequences is smaller than analogous reduction in DHSs (Supplementary Fig. 6). Furthermore, the observed reduction of mutation frequency in exons may not necessarily represent the action of purifying selection. The frequency of missense mutations is not lower (and is even apparently higher) than frequency of synonymous changes (Supplementary Fig. 6). Thus, although it is possible that cancers may differ from evolving populations and cannot rule out the action of purifying selection in regulatory regions, we suggest that mutation attenuation plays a more important role.

Association with nucleotide excision repair

Relative mutation density depends on replication fidelity, levels of DNA damage or efficiency of DNA repair. The fact that the observed relative reduction of mutation density was highly limited to DHSs, makes it difficult to explain it by an increase in global replication fidelity. It is also unlikely that more accessible DNA at DHSs would be less prone to damage than less accessible DNA elsewhere. In fact, mutation frequencies observed in model organisms are reduced by positioned nucleosomes[22,23], while the effects of nucleosome positioning on somatic mutations in cancer are relatively small and differ in directions between various cancer types[15].

Chromatin accessibility plays a major role in targeting nuclear proteins to regulatory DNA, and may provide a mechanism for preferential access by the repair machinery. Preferential activity of DNA repair proteins in accessible regulatory DNA may thus offer an explanation for the observed effect, analogous to the action of transcription-coupled repair in protein coding genes. We hypothesized specifically that potentiation of nucleotide excision repair (NER) and base excision repair (BER) by chromatin accessibility could be responsible for the observed relative reduction of mutation density at DHSs. The level of oxidative stress and the subsequent accumulation of lesions targeted by BER[24] is higher in malignant vs. normal cells[25]. BER is an evolutionary conserved DNA repair pathway, which starts from the recognition and excision of various base lesions by specific DNA glycosylases, followed by the processing of the resulting AP sites and then DNA repair synthesis and ligation. Direct access of glycosylases to DNA lesions is pivotal for this repair process[26]. Not surprisingly, therefore, BER complexes preferentially assemble in non-nucleosomal regions in response to oxidative stress[27], which naturally targets them to DHSs.

The NER pathway consists of two converging branches: global genome repair and transcription-coupled repair (TCR). DNA damage is first recognized by the XPC complex, DNA duplex is opened by XPD and XPB helicases, followed by the incision of the damaged strand XPF-ERCC1 and XPG nucleases, then gap filling by the replication polymerases and, finally DNA ligation[28]. A priori, NER machinery shall be able to correct damaged DNA regardless of its chromatin state. A fully assembled NER complex, however, has a footprint of ~100 bps in DNA, which is significantly longer than the length of an internucleosomal linker. As a result, chromatin structure inhibits functional NER complex assembly and function[29-31]. In case of TCR, this problem is circumvented by the fact that DNA damage is first sensed by the RNA polymerase followed by NER recruitment to an already unraveled chromatin requires by the CSB and CSA proteins (reviewed in ref. 6). This is not the case for global genome repair, and the problem is additionally acerbated by the fact that the damage sensor, the XPC complex, cannot bind to DNA adducts embedded in nucleosomes[32]. Thus, both lesion recognition and repair complex assembly may be potentiated in accessible chromatin[29,33]. The fact that roughly half of DHSs lie in intergenic regions suggests GGR as the more likely candidate. Notably, nucleosomal chromatin appears to inhibit global genome repair function[29-31]. Moreover, the XPC complex involved in damage recognition is inhibited from binding to lesions in nucleosomal DNA[32]. Thus, both lesion recognition and repair complex assembly may be potentiated in accessible chromatin[29,33].

Failure of NER predisposes to cancer. This is best illustrated by the extreme frequency of cancers caused by exposure to sunlight in Xeroderma pigmentosum (XP) patients, evidently due to their inability to repair UV photoproducts in DNA. Mutations in NER are commonly detected in melanomas. Similarly to XP, most somatic mutations observed in melanoma cells are C:G→T:A transitions caused by ultraviolet (UV)[34] damage, which are primarily repaired by NER.

BER and NER are high fidelity pathways as suggested by studies showing that deactivation of these pathways leads to increased chance of mutation from cryptic lesions or exogenous DNA damage[34-37]. As was shown recently, NER defects lead to increase in density of C:G→T:A mutations under chronic low-dose UV radiation, conditions specifically relevant to melanoma[38].

We, therefore, reasoned that the relative reduction of mutation density in DHSs in individual melanoma genomes should parallel the integrity of NER pathway components. To test this, we analyzed 29 individual melanoma genomes sequenced at high coverage[1]. Figure 2 and Supplementary Table 3 show the continuous dependence of C:G→T:A mutation densities on chromatin accessibility in melanocytes in both intergenic regions, introns and exons. We note that the observed effect is inconsistent with the recently discovered action of APOBEC proteins[39-41]. First, APOBEC acts on single strand DNA and it is unlikely that inaccessible and untranscribed DNA would more readily adopt single strand conformation than DNA in DHSs and transcribed regions. Second, action of APOBEC would preferentially increase rates of C→T transitions and C→G transversions within TCA/TCT motif. Our observation is not confined to this motif, and we do not detect a parallel effect for C→G changes (Supplementary Fig. 7). Quantitatively, dependence of mutation density within TCA/TCT motif on number of DNase I cleavages is slightly lower (rather than higher) than the dependence of mutations within TCG/TCC motif (*p*-value for the interaction term in the regression analysis in Supplementary Figure 7 is 2.11×10^{-5}). This is inconsistent with the hypothesis that APOBEC action induces the observed dependency on chromatin accessibility.

This effect was far more pronounced in melanocyte chromatin vs. that of other cell. Furthermore, the signature of TCR activity is also observed by the higher density of C→T over G→A in the non-template strand[2] of exons and introns (Supplementary Fig. 8). This demonstrates that NER provide genes a multi layer protection against UV-light damage due to activity of TCR and the accessibility of chromatin to global genome repair.

Overall, 9 out of 29 melanoma genomes harbored non-synonymous mutations in NER genes. Notably, four melanoma genomes with the lowest levels of mutation reduction at DHSs all harbor mutations in NER genes[42-44] ($P < 0.0237$, Wilcoxon-Mann-Whitney test; Fig. 3). In three of these samples, mutation frequencies in DHSs return close to the genomic baseline. The presence of genomes with mutations in NER genes and reduced mutation frequencies in DHSs is not surprising because NER mutations may appear late in cancer development and some of the missense mutations may be functionally benign. This result implicates NER into observed reduction of mutation frequency in DHSs. It also provides an additional argument against selection explanation because purifying selection would not be expected to differentially impact melanoma samples.

Eight out of the nine samples with mutations in NER pathway genes harbored mutations in major components of the NER machinery (*XPG/ERCC5*, *XPF/ERCC4* and *LIG1*). These lesions would be expected to compromise both NER and TCR, and therefore should affect mutation density in both DHSs and transcribed regions. Intriguingly, one of these samples also harbored a mutation in *CETN2*, which recognizes DNA distortions and therefore preferentially impacts GGR function over TCR[45]. In agreement with this reasoning, three genomes carrying mutations in core subunits had markedly reduced or negligible reduction of mutation density in transcribed regions compatible with defective TCR (Fig. 4). Concordantly, in the genome carrying a mutation in the *CETN2* gene, strong suppression of mutations in transcribed regions remained, implying that TCR function was not significantly compromised.

Discussion

Taken together, our results suggest that relative density of somatic mutations in cancer genomes is substantially suppressed in regulatory DNA, and that mutation frequency closely tracks chromatin accessibility. The hypomutational effect is highly localized and is statistically associated with intact global genome repair. The analysis of individual melanoma samples suggests that relationship between relative mutation density and chromatin accessibility may be mediated by DNA repair. Our analysis could not completely rule out alternative explanations such as selection for regulatory function or increased C→T deamination through enzymatic activities getting abnormal access to DNA. However, these alternatives would require to invoke yet unknown mechanism[46] to explain association with NER in melanoma.

Our results link fine-scale chromatin accessibility with the cancer mutation accumulation. Given the growing interest in the role of regulatory sequences in cancer progression[47], these results will help providing a necessary baseline for cancer genomics projects targeting non-coding regions, similarly to computational approaches used in the analysis of protein-coding genes[48].

With the increasing amount of whole genome sequencing data, our approach can also be formalized and extended to associate mutational patterns with specific pathways, including DNA repair, DNA replication and chromatin remodeling. Mutational patterns can be treated as traits of individual tumor samples. With the large number of tumor samples available, these mutational traits can be associated with recurrent mutations in specific genes controlling mutagenesis, potentially identifying important players shaping somatic mutational landscapes.

Online Methods

DNaseI hypersensitivity mapping

DNaseI mapping was conducted on cultured cancer cell lines, primary ex vivo hematopoietic cells, cultured primary cells, and isolated fetal tissues using appropriate nuclei isolation protocols (below), followed by a standard processing pipeline. Data from lines A549, HepG2, LNCap, CACO2, PANC1, CLL, K562, CMK, NB4, and MCF7 derive from Reference 8. Generation of new data from M059J (Glioblastoma), RPMI_7951 (melanoma), CD19 (B-cell), CD20 (B-cell), melanocytes, fetal lung and fetal intestine are described below.

Isolation of nuclei from cultured cancer cell lines

Cells were cultured in accordance with the detailed protocols provided at http://www.uwencode.org/protocols. To prepare nuclei, freshly grown cells were centrifuged at 500g for 5 minutes (4°C) in an Eppendorf Centrifuge 5810R, and washed in cold PBS (Cellgro/Mediatech Inc.). Cell pellets were resuspended in Buffer A (15 mM Tris-Cl pH 8.0, 15 mM NaCl, 60 mM KCl, 1 mM EDTA (Ambion/Life Technologies Corp) pH 8.0, 0.5 mM EGTA (Boston BioProducts) pH 8.0, 0.5 mM spermidine (MP Biomedicals, LLC) and 0.15 mM spermine (MP Biomedicals, LLC) to a final concentration of 2 × 106 cells/mL. Nuclei were obtained by dropwise addition of an equal volume of Buffer A containing 0.04% IGEPAL CA-630 (Sigma-Aldrich) to the cells, followed by incubation on ice for 10 min. Nuclei were centrifuged at 1,000g for 5 min and then resuspended and washed with 25 mL of cold Buffer A. Nuclei were resuspended in 2 mL of Buffer A at a final concentration of 1 × 107 nuclei/mL.

Isolation of nuclei from hematopoietic cells

CD19+ and CD20+ cells (separately) were isolated by immunomagnetic separation by the Large Scale Cell Processing Facility at the Fred Hutchinson Cancer Research Center from normal volunteer donors under an IRB-approved protocol. Cells were pelleted by centrifugation for 5 minutes at 500g at 4 °C. Cells were washed in ice-cold PBS, then resuspended to 5 million cells per mL in Buffer A. An equal volume of ice-cold 2X IGEPAL CA-630 solution (ranging from 0.02%-0.06%) was added and the tube was incubated for 5 - 6 minutes on ice to lyse the cells. Nuclei were pelleted by centrifugation for 5 minutes at 500g at 4 °C, resuspended in Buffer A and counted with a hemocytometer.

Isolation of nuclei from fetal tissues

Fetal lung and intestine tissues were obtained from morphologically normal fetuses by the Birth Defects Research Laboratory in the Dept. of Pediatrics at University of Washington, collected under an IRB-approved protocol. Tissue was minced, resuspended in cold 250 mM sucrose, 1 mM MgCl2, 10 mM Tris-Cl pH 7.5, with added EDTA Protease Inhibitor Cocktail (Roche Applied Science Corp.). Resuspended tissue from fetal brain, fetal lung, fetal kidney, and fetal adrenal was dissociated by slowly homogenizing with a Dounce homogenizer. Resuspended tissue from fetal heart or fetal intestine was dissociated in a gentleMACS Dissociator (Miltenyi Biotech Inc.). Following dissociation, all fetal tissues were filtered through a 100 uM filter, and nuclei pelleted by centrifugation 600g for 10 minutes. Pelleted nuclei were washed with Buffer A, resuspended in Buffer A and counted in a hemocytometer.

DNaseI mapping from isolated nuclei

Briefly, DNaseI digestion was performed as described in Reference 8, with minor modifications. Isolated nuclei (2 × 106) from suspension cells or dissociated tissue were washed with 15 mM Tris-Cl pH 8.0, 15 mM NaCl, 60 mM KCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.5 mM spermidine and 0.15 mM spermine then subjected to DNaseI digestion for 3 min at 37 °C in 13.5 mM Tris-HCl pH 8.0, 87 mM NaCl, 54 mM KCl, 6 mM CaCl2, 0.9 mM EDTA, 0.45 mM EGTA, 0.45 mM Spermidine. Digestion was stopped by addition of 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.1% SDS, 100 mM EDTA pH 8.0, 1 mM spermidine, 0.3 mM spermine. A range of DNaseI (Sigma-Aldrich), 10–80 U/mL) concentrations was used for each preparation of nuclei and the sample giving the optimum difference between DNaseI treated and untreated was used for sequencing library construction. DNaseI double-hit fragments were collected by ultra-centrifugation and gel-purified. Adaptors were ligated to the ends of purified fragments, and the resulting libraries sequenced on an Illumina Genome Analyzer IIx according to a standard protocol.Processing of DNaseI-seq data. 36-base reads with up to two mismatches were mapped to the human genome (GRCh37/hg19) using the sequence aligner BOWTIE. DHSs were identified using the Hotspot algorithm (8) at a false discovery rate (FDR) threshold of 5%. Genomic feature overlaps and distance calculations were performed using the BEDOPS suite of software tools available at http://code.google.com/p/bedops/.

Data availability

All DNaseI data used in this study have been released to the ENCODE Project repository or to the Roadmap Epigenomics Mapping Consortium data coordination center. These data have been deposited in GEO under accession numbers GSE29692 and GSE18927. Data are also available for download through www.uwencode.org/data and through the data links at www.epigenomebrowser.org.

Cancer datasets

We restricted our study to cancer genomes sequenced by Sanger Institute and by Broad Institute Whole genomes of COLO-829 and NCI-H209 cell lines that have been sequenced by Sanger Institute[2,3]. COLO-829 cell line has been derived from metastatic tissue. NCI-H209, an immortal cell line of a small cell lung cancer derived from a bone marrow metastasis. Nine colon cancer genome[13], 23 multiple myelomas[5] and 29 melanoma genomes have been sequenced by Broad Institute. For the 4 Broad institute datasets we identified sites in the genome were we have over 80% power to detect mutations (this defined by at least 14 reads the covered his position in the tumor sample and 8 reads that cover this position in normal). About 81% of the bases in colon and MM genomes are sufficiently covered and about 86% of the melanoma genomes are sufficiently covered.

Annotations

gene and exon coordinates were retrieved for hg19 from UCSC genome browser. For the flank analysis we took 1000 bps windows around the DHSs, since some of flank regions overlapped with DHSs we removed the DHS defined sites from the original set of flanks.

Density of mutations

We calculated the number of mutations per bps that can be mapped uniquely by the DHS essays.

Poisson regression

In order to test whether DHS regions have an additional impact on rates on top of other cofounding factors we used multivariate Poisson regression[18]. We divided the genomes into non-overlapping 400 bp windows. Every bin was classified as intergenic, intronic and exonic regions. The windows were also classified as DHS regions when overlap at least 80bp of any DHSs. For each window we calculated the following quantities: GC content, CpG content, distance from the nearest transcription start site, distance from nearest CpG islands, distance from nearest DHSs, and mutation counts and coverage (i.e. the number of bases which we have 80% power to detect mutation) for Colon, multiple myeloma, and similarly the number of GC bases in a window that are covered in Broad Institute melanoma samples. Then using glm function in R we calculated the estimated rates[18]. We repeated the analysis for spatially separated windows comprising 20% of the data to ensure that possible short-range interdependence between neighboring windows does not artificially inflate statistical significance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by US National Institutes of Health grants U54CA143874 and RO1MH101244. We thank Dmitry Gordenin for his valuable feedback.

References

1. Berger MF, et al. Melanoma genome sequencing reveals frequent PREX2 mutations. Nature. 2012; 485:502–6. [PubMed: 22622578]
2. Pleasance ED, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature. 2010; 463:191–6. [PubMed: 20016485]
3. Pleasance ED, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature. 2010; 463:184–90. [PubMed: 20016488]
4. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet. 2010; 11:685–96. [PubMed: 20847746]
5. Chapman MA, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011; 471:467–72. [PubMed: 21430775]
6. Hanawalt PC, Spivak G. Transcription-coupled DNA repair: two decades of progress and surprises. Nat Rev Mol Cell Biol. 2008; 9:958–70. [PubMed: 19023283]

7. Lainé J, Egly J. Initiation of DNA repair mediated by a stalled RNA polymerase IIO. EMBO J. 2006; 25:387–397. [PubMed: 16407975]

8. Thurman RE, et al. The accessible chromatin landscape of the human genome. Nature. 2012; 489:75–82. [PubMed: 22955617]

9. Gross DS, Garrard WT. Nuclease hypersensitive sites in chromatin. Annu Rev Biochem. 1988; 57:159–97. [PubMed: 3052270]

10. Neph S, et al. An expansive human regulatory lexicon encoded in transcription factor footprints. Nature. 2012; 489:83–90. [PubMed: 22955618]

11. Legault J, Tremblay A, Ramotar D, Mirault ME. Clusters of S1 nuclease-hypersensitive sites induced in vivo by DNA damage. Mol Cell Biol. 1997; 17:5437–52. [PubMed: 9271420]

12. Parker SC, et al. Mutational signatures of de-differentiation in functional non-coding regions of melanoma genomes. PLoS Genet. 2012; 8:e1002871. [PubMed: 22912592]

13. Bass AJ, et al. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. Nat Genet. 2011; 43:964–8. [PubMed: 21892161]

14. Cibulskis K, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013; 31:213–9. [PubMed: 23396013]

15. Hodgkinson A, Chen Y, Eyre-Walker A. The large-scale distribution of somatic mutations in cancer genomes. Hum Mutat. 2012; 33:136–43. [PubMed: 21953857]

16. Stamatoyannopoulos JA, et al. Human mutation rate associated with DNA replication timing. Nat Genet. 2009; 41:393–5. [PubMed: 19287383]

17. Schuster-Bockler B, Lehner B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature. 2012; 488:504–7. [PubMed: 22820252]

18. Faraway, JJ. Extending the linear model with R : generalized linear, mixed effects and nonparametric regression models. Vol. ix. Chapman & Hall/CRC; Boca Raton: 2006. p. 301

19. Hansen RS, et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A. 2010; 107:139–44. [PubMed: 19966280]

20. King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. Science. 1975; 188:107–16. [PubMed: 1090005]

21. Lindblad-Toh K, et al. A high-resolution map of human evolutionary constraint using 29 mammals. Nature. 2011; 478:476–82. [PubMed: 21993624]

22. Chen X, et al. Nucleosomes suppress spontaneous mutations base-specifically in eukaryotes. Science. 2012; 335:1235–8. [PubMed: 22403392]

23. Sasaki S, et al. Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. Science. 2009; 323:401–4. [PubMed: 19074313]

24. Cheng KC, Cahill DS, Kasai H, Nishimura S, Loeb LA. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. J Biol Chem. 1992; 267:166–72. [PubMed: 1730583]

25. Kawanishi S, Hiraku Y, Pinlaor S, Ma N. Oxidative and nitrative DNA damage in animals and patients with inflammatory diseases in relation to inflammation-related carcinogenesis. Biol Chem. 2006; 387:365–72. [PubMed: 16606333]

26. Hitomi K, Iwai S, Tainer JA. The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. DNA Repair (Amst). 2007; 6:410–28. [PubMed: 17208522]

27. Amouroux R, Campalans A, Epe B, Radicella JP. Oxidative stress triggers the preferential assembly of base excision repair complexes on open chromatin regions. Nucleic Acids Res. 2010; 38:2878–90. [PubMed: 20071746]

28. Friedberg EC, et al. DNA Repair And Mutagenesis. 2006 ASM Press.

29. Bell O, Tiwari VK, NH T, Schübeler D. Determinants and dynamics of genome accessibility. Nat Rev Genet. 2011; 12:554–564. [PubMed: 21747402]

30. Thoma F. Repair of UV lesions in nucleosomes--intrinsic properties and remodeling. DNA Repair (Amst). 2005; 4:855–69. [PubMed: 15925550]

31. Aboussekhra A, et al. Mammalian DNA nucleotide excision repair reconstituted with purified protein components. Cell. 1995; 80:859–68. [PubMed: 7697716]

32. Yasuda T, et al. Nucleosomal structure of undamaged DNA regions suppresses the non-specific DNA binding of the XPC complex. DNA Repair (Amst). 2005; 4:389–395. [PubMed: 15661662]

33. Fei J, et al. Regulation of nucleotide excision repair by UV-DDB: prioritization of damage recognition to internucleosomal DNA. PLoS Biol. 2011; 9:e1001183. [PubMed: 22039351]

34. Shuck SC, Short EA, Turchi JJ. Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. Cell Res. 2008; 18:64–72. [PubMed: 18166981]

35. Sugasawa K. Xeroderma pigmentosum genes: functions inside and outside DNA repair. Carcinogenesis. 2008; 29:455–65. [PubMed: 18174245]

36. Hanawalt PC, Ford JM, Lloyd DR. Functional characterization of global genomic DNA repair and its implications for cancer. Mutat Res. 2003; 544:107–14. [PubMed: 14644313]

37. Girard PM, Boiteux S. Repair of oxidized DNA bases in the yeast Saccharomyces cerevisiae. Biochimie. 1997; 79:559–66. [PubMed: 9466693]

38. Haruta N, Kubota Y, Hishida T. Chronic low-dose ultraviolet-induced mutagenesis in nucleotide excision repair-deficient cells. Nucleic Acids Res. 2012; 40:8406–15. [PubMed: 22743272]

39. Nik-Zainal S, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012; 149:979–93. [PubMed: 22608084]

40. Roberts SA, et al. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol Cell. 2012; 46:424–35. [PubMed: 22607975]

41. Burns MB, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature. 2013; 494:366–70. [PubMed: 23389445]

42. Wood, RD. 2011. http://sciencepark.mdanderson.org/labs/wood/dna_repair_genes.html#Human%20DNA%20Repair%20Genes

43. Lange SS, Takata K, Wood RD. DNA polymerases and cancer. Nat Rev Cancer. 2011; 11:96–110. [PubMed: 21258395]

44. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012; 40:D109–14. [PubMed: 22080510]

45. Palomera-Sanchez Z, Zurita M. Open, repair and close again: chromatin dynamics and the response to UV-induced DNA damage. DNA Repair (Amst). 2011; 10:119–25. [PubMed: 21130713]

46. Iyer LM, Zhang D, Rogozin IB, Aravind L. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic Acids Res. 2011; 39:9473–97. [PubMed: 21890906]

47. Huang FW, et al. Highly recurrent TERT promoter mutations in human melanoma. Science. 2013; 339:957–9. [PubMed: 23348506]

48. Lawrence MS, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499:214–8. [PubMed: 23770567]

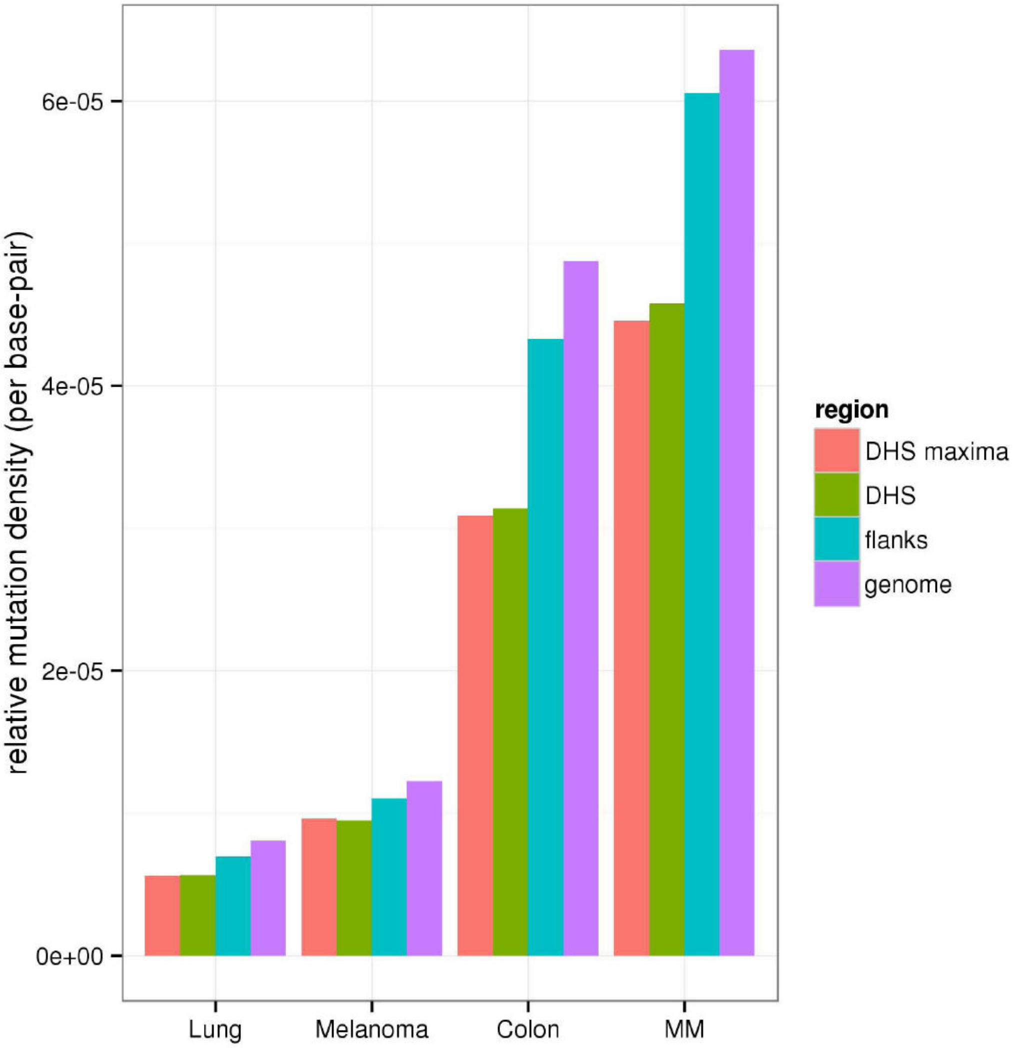

Figure 1.

Relative density of somatic mutations is reduced in DHSs of all analyzed cancer genomes (lung[3], melanoma[2], colon[13], multiple myeloma[5]). Mutation density per (uniquely mappable) bp is shown for 1) DHS maxima defined as plus or minus 75 bp around the peak of DNase I hypersensitivity (marked as DHS peaks), 2) DHSs, 3) 1000 bp flanking regions and 4) overall genome. Mutation density in DHSs is substantially lower in comparison with immediate flanking regions and genome average. The effect is stronger for DHS maxima compared to overall DHSs.

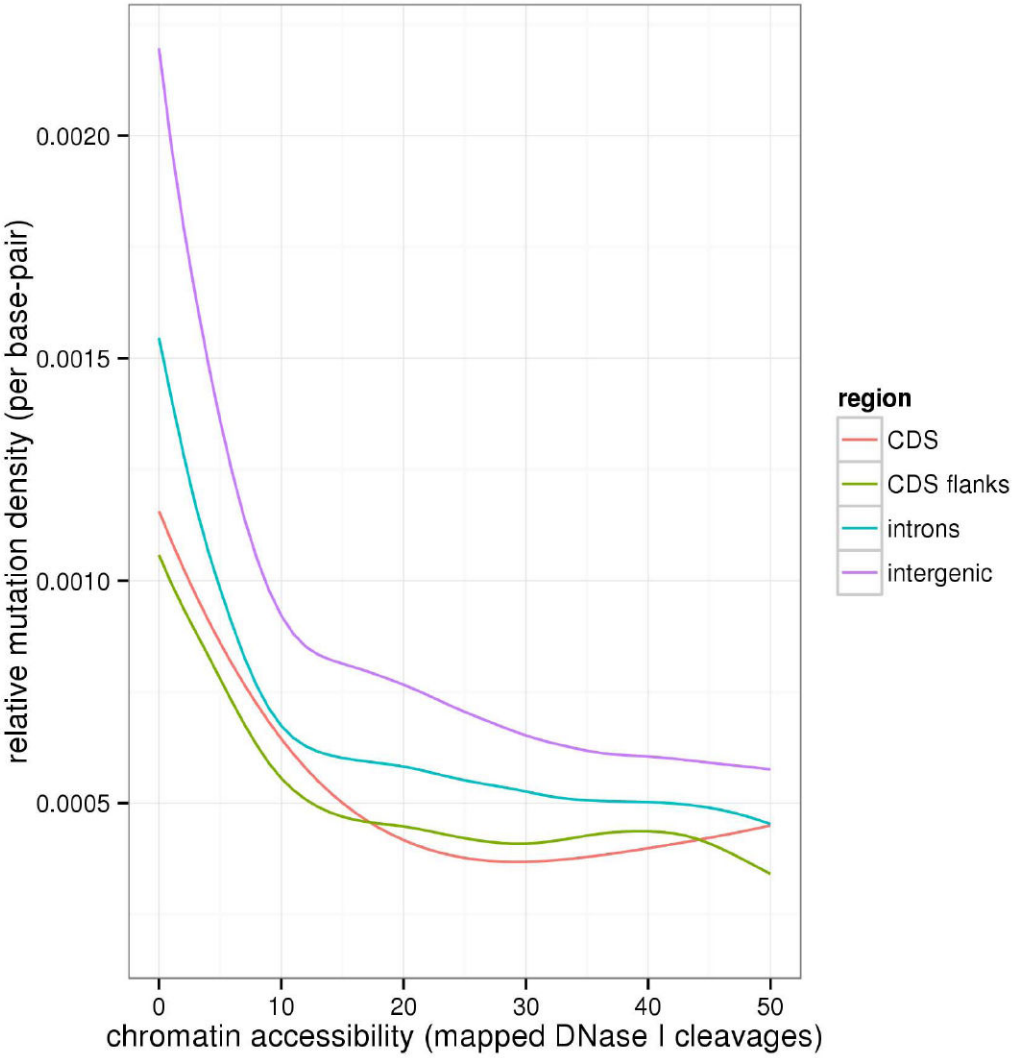

Figure 2.

Density of somatic C:G→T:A transition mutations in melanoma samples strongly depends chromatin accessibility in a monotonic and continuous fashion. Density of C:G→T:A transitions per C:G base-pair in 400bp genomic intervals is shown as function of chromatin accessibility in melanocytes measured by the density DNase I cleavages. The dependence is presented separately for introns and intergenic regions, and is equally present in both. Mutation densities are parametrically fitted to a spline function using a Generalized Additive Model Poisson regression model[18].

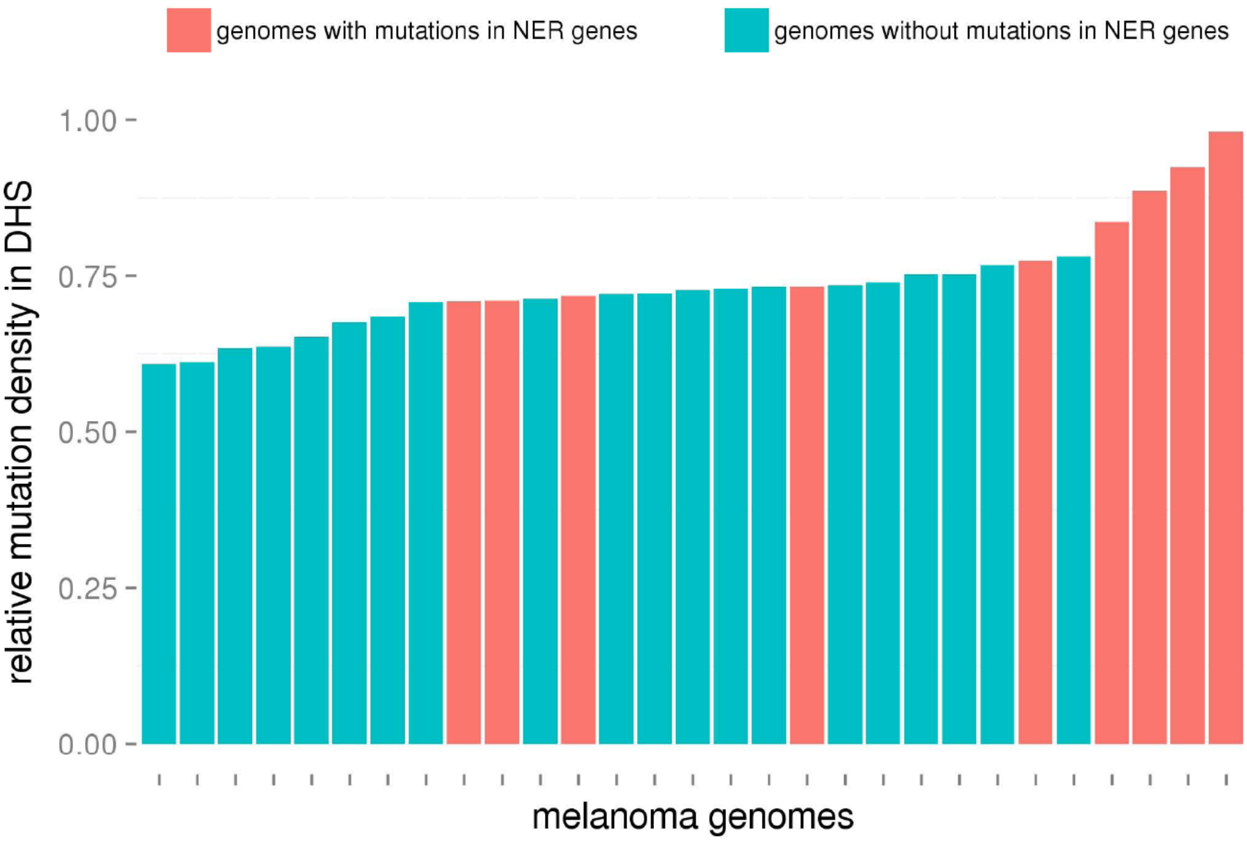

Figure 3.

Normalization of DHS hypomutation in melanoma genomes with mutated nucleotide excision repair pathway genes. Relative mutation density in DHSs of melanoma genomes is shown for samples with an intact NER system (blue) and samples with non-synonymous mutations in NER pathway genes (red). Non-synonymous changes in NER pathway genes significantly track relative mutation reduction in DHSs (P < 0.0237, Wilcoxon-Mann-Whitney test).

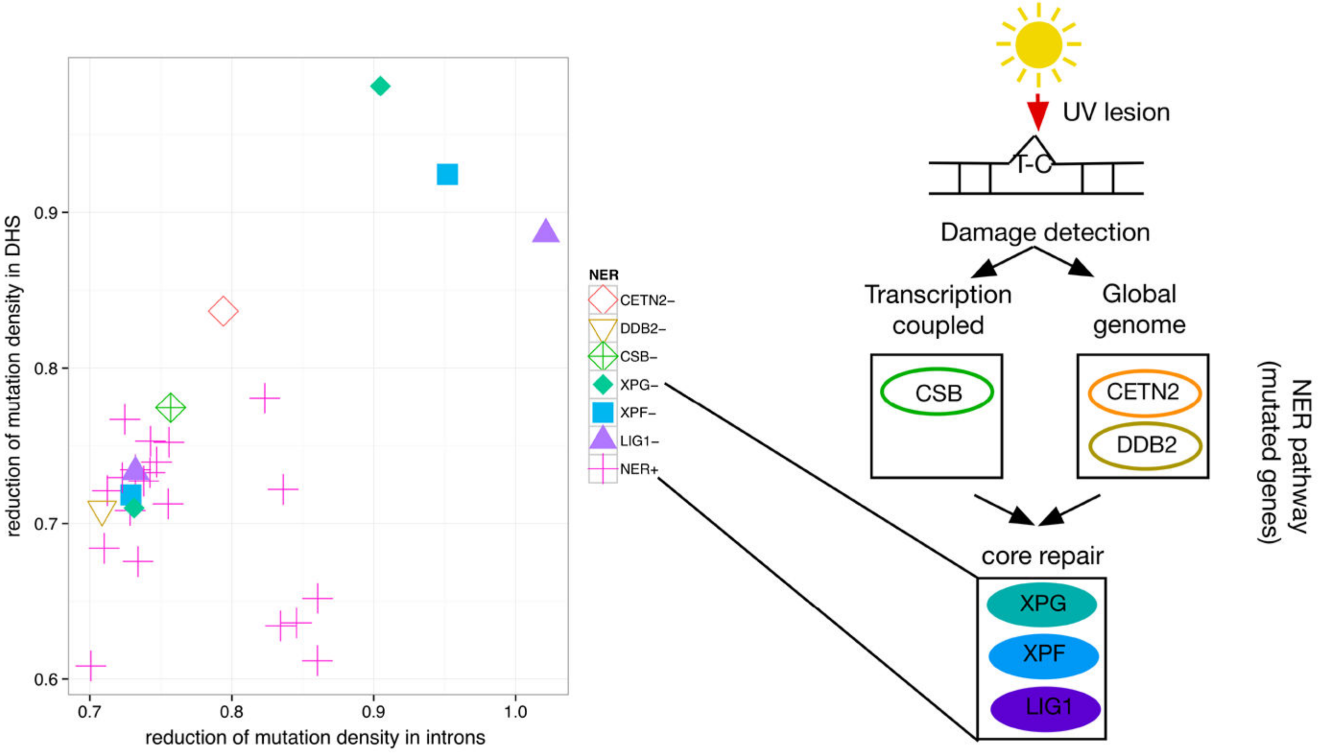

Figure 4.

Reduction of mutation density in DHSs and in transcribed regions. Shown for individual melanoma samples (scatter plot) are non-synonymous mutations in genes involved in NER (marked by shape and color corresponding to each gene). Roles of these genes within NER pathway are shown by the diagram on the right. *XPG*, *XPF* and *LIG1* are core repair components; *CETN2* and *DDB2* are specific to GGR and are involved in lesion recognition. *CSB* is specific to TCR and is involved in recruiting NER to the stalled Pol II RNA polymerase. Samples with low level (or no) reduction of somatic mutations in DHSs and carrying non-synonymous changes in genes of core NER components also show low level (or no) reduction of mutation frequency in transcribed regions, suggesting that core NER genes are responsible for both effects.