1-1-2019

Mutational Patterns in Metastatic Cutaneous Squamous Cell Carcinoma

Simon Mueller
University of Bern, Chris O'Brien Lifehouse, Garvan Institute Of Medical Research

Marie-Emilie A. Gauthier
Children's Cancer Institute, Garvan Institute Of Medical Research, Chris O'Brien Lifehouse, m.gauthier@garvan.org.au

Bruce G. Ashford
University of Wollongong, Chris O'Brien Lifehouse, Illawarra and Shoalhaven Local Health District, bruceash@uow.edu.au

Ruta Gupta
Chris O'Brien Lifehouse, University of Sydney, Royal Prince Alfred Hospital, University of Sydney, Ruta.Gupta@health.nsw.gov.au

Velimir Gayevskiy
Garvan Institute Of Medical Research

See next page for additional authors

Mutational Patterns in Metastatic Cutaneous Squamous Cell Carcinoma

Abstract

Cutaneous squamous cell carcinoma from the head and neck typically metastasize to the lymph nodes of the neck and parotid glands. When a primary is not identified, they are difficult to distinguish from metastases of mucosal origin and primary salivary gland squamous cell carcinoma. UV radiation causes a mutation pattern that predominantly features cytosine to thymine transitions at dipyrimidine sites and has been associated with cutaneous squamous cell carcinoma. In this study, we used whole genome sequencing data from 15 cutaneous squamous cell carcinoma metastases and show that a UV mutation signature is pervasive across the cohort and distinct from mucosal squamous cell carcinoma. The mutational burden was exceptionally high and concentrated in some regions of the genome, especially insulator elements (mean 162 mutations/megabase). We therefore evaluated the likely impact of UV-induced mutations on the dipyrimidine-rich binding site of the main human insulator protein, CCCTC-binding factor, and the possible implications on CCCTC-binding factor function and the spatial organization of the genome. Our findings suggest that mutation signature analysis may be useful in determining the origin of metastases in the neck and the parotid gland. Furthermore, UV-induced DNA damage to insulator binding sites may play a role in the carcinogenesis and progression of cutaneous squamous cell carcinoma.

Disciplines

Medicine and Health Sciences

Publication Details

Mueller, S. A., Gauthier, M. A., Ashford, B., Gupta, R., Gayevskiy, V., Ch'ng, S., Palme, C. E., Shannon, K., Clark, J. R., Ranson, M. & Cowley, M. J. (2019). Mutational Patterns in Metastatic Cutaneous Squamous Cell Carcinoma. Journal of Investigative Dermatology, 139 (7), 1449-1458.e1.

Authors

Simon Mueller, Marie-Emilie A. Gauthier, Bruce G. Ashford, Ruta Gupta, Velimir Gayevskiy, Sydney Ch'ng, Carsten Palme, Kerwin Shannon, Jonathan Clark, Marie Ranson, and Mark Cowley

Mutational Patterns in Metastatic Cutaneous Squamous Cell Carcinoma

Authors

Simon A Mueller, MD[1,2,3]*, ORCID 0000-0001-9318-7270
Marie-Emilie A Gauthier, PhD[1,2,13]* ORCID 0000-0002-5256-9165
Bruce Ashford, MD[1,4,5,6] ORCID 0000-0003-4547-6611
Ruta Gupta, MD[1,78] ORCID 0000-0003-2940-313X
Velimir Gayevskiy, PhD[2] ORCID 0000-0002-2457-6261
Sydney Ch'ng, MD[1,9,10]
Carsten E Palme, MD[1,7]
Kerwin Shannon, MD[1]
Jonathan R Clark, MD[1,7,9] ORCID 0000-0003-1209-772X
Marie Ranson, PhD[5,6,11], ORCID 0000-0002-5570-9645
Mark J Cowley, PhD[2,12,13] ORCID 0000-0002-9519-5714

* Equal contribution

Affiliations

[1] Sydney Head and Neck Cancer Institute, Chris O'Brien Lifehouse, Sydney, Australia

[2] Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney, Australia

[3] Department for Oto-Rhino-Laryngology, Head and Neck Surgery, Inselspital, Bern University Hospital, University of Bern, Switzerland

[4] Illawarra and Shoalhaven Local Health District (ISLHD), Wollongong, Australia

[5] School of Biological Sciences, University of Wollongong, Wollongong, Australia

[6] Illawarra Health and Medical Research Institute (IHMRI), Wollongong, Australia

[7] Central Clinical School, The University of Sydney, Sydney, Australia

[8] Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Sydney, Australia

[9] Institute of Academic Surgery, University of Sydney, Sydney, Australia

[10] Department of Plastic Surgery, Royal Prince Alfred Hospital, Sydney, Australia.

[11] Centre for Oncology Education and Research Translation (CONCERT), Liverpool, Australia

[12] St Vincent's Clinical School, UNSW Sydney, Sydney, Australia

[13] Children's Cancer Institute, Kensington, Australia

Corresponding author

Simon Andreas Mueller
Sydney Head and Neck Cancer Institute
Chris O'Brien Lifehouse
PO Box M33, Missenden Road
Camperdown NSW 2050, Australia

Tel: +61451052718
Fax: +61295199214
Email: simon.mueller@insel.ch

Study location

Sydney, New South Wales, Australia

Short title

Mutation patterns in metastatic cSCC

Abbreviations

ChIP-Seq	Chromatin immunoprecipitation sequencing
CTCF	CCTC-binding factor
CTCFbs	CTCF binding site
cSCC	Cutaneous squamous cell carcinoma
NHEK	Normal human epidermal keratinocyte
NER	Nucleotide excision repair
TAD	Topologically associated domain
UV	Ultraviolet radiation

ABSTRACT

Metastases of cutaneous squamous cell carcinoma (cSCC) from the head and neck typically occur in lymph nodes of the neck and parotid glands. When a primary is not identified, they are difficult to distinguish from metastases of mucosal origin and primary salivary gland SCC. Ultraviolet radiation causes a mutation pattern that predominantly features cytosine to thymine transitions at dipyrimidine sites and has been associated with cSCC. In this study, we used whole genome sequencing data from 15 cSCC metastases and show that a UV-signature based on the trinucleotide context of mutations, distinguishable from mucosal SCC, is present across the cohort. The mutational burden was exceptionally high and concentrated in some regions of the genome, especially insulator elements (mean 162 mutations/Mb). We therefore evaluated the likely impact of UV-induced mutations on the dipyrimidine rich binding site of the main human insulator protein, CCCTC-binding factor (CTCF), and the possible implications on CTCF function and the spatial organization of the genome. Our findings suggest that mutation signature analysis may be useful in determining the origin of metastases in the neck and the parotid gland. Furthermore, UV-induced DNA damage to insulator binding sites may play a role in the carcinogenesis and progression of cSCC.

INTRODUCTION

Cutaneous squamous cell carcinoma (cSCC) is the second most common skin cancer (Gurudutt and Genden, 2011) and most primaries arise in the skin of the face and scalp. Regional lymph node metastases to the neck and the intraparotid nodes occur in up to 5% of patients, entailing significant morbidity and mortality (D'Souza and Clark, 2011). When metastatic squamous cell carcinoma of unknown primary is diagnosed in neck lymph nodes or the parotid gland, it is sometimes impossible to determine the primary site based on clinicopathologic features alone. The potential tissues of origins include skin, mucosa, or the salivary gland. The few studies using massive parallel sequencing on cSCC demonstrate that the variability of affected genes is high and there is extensive overlap with SCC from mucosal surfaces, making gene mutation assessment alone an unreliable differentiator (Ashford et al., 2017; Lawrence et al., 2015; Pickering et al., 2014; South et al., 2014).

DNA mutation patterns may offer a solution to differentiating skin-derived SCC metastases from those originating from other primary sites. Nik-Zainal et al. (2012) demonstrated that mutation patterns can be distinguished when considering the region immediately surrounding each somatic mutation, that is, the trinucleotide context. The relative contribution of recognized mutation signatures can be quantified, providing insights into the exposure of the tumor to various mutagenic agents. Since this initial report, a growing number of signatures with known associations have been reported (Alexandrov et al., 2018). Ultraviolet radiation (UV) is the main risk factor for the development of cSCC and predominantly causes DNA mutations at dipyrimidine sites, where it induces C→T transitions (Douki and Cadet, 2001; Marteijn et al., 2014). Studies using targeted or whole-exome sequencing have observed predominance of C→T transitions in mostly primary cSCC (Durinck et al., 2011; Pickering et al., 2014; South et al., 2014; Zilberg et al., 2018). Other mutational processes also cause C→T mutations (Forbes et al., 2017), but trinucleotide mutation signature analysis is able to resolve the underlying mutagenic agents, and identify UV-specific C→T mutations.

The distribution of these mutations is uneven across the genome and affects specific regions more than others (Perera et al., 2016). Because of its high dipyrimidine content, the binding site of the main human insulator, CCCTC-binding factor (CTCF), is especially susceptible to UV-induced mutations. CTCF may play an important role in carcinogenesis, as it regulates the transcriptional activity of topologically associated domains (TADs), which represent chromatin loops harboring multiple genes (Hnisz et al., 2016; Kemp et al., 2014).
Here, we analyze genome-wide mutations in regard to their trinucleotide context, present the first signature analysis of cSCC metastases and explore its clinical utility. We further assess the distribution of UV-induced mutations across the genome, and assess the potential impact on CTCF using the approach described by Poulos et al. (2016) in melanoma.

RESULTS

Patient Characteristics

Six parotid and nine neck lymph node cSCC metastases from 15 patients were included (Table 1). The site of the cSCC primary was unknown in four patients. Two patients received immunosuppressive therapy.

Mutational Patterns Across the Genome

Average coverage of whole genome sequencing (WGS) was 83.0× in tumors and 37.9× in blood. The total number of somatic mutations per sample ranged from 136,105 to 1,423,398 (mean 638,254; Figure 1a). Mutations occurred mostly in non-coding regions (99.4%), and mutation density was 171 times higher for non-coding variants compared to coding variants (mean 206.6 vs. 1.2 mutations/Mb, median 168.3 vs. 0.9 mutations/Mb respectively; Figure 1a and 1b). Assessment of the distribution of somatic mutations across broad genome region categories indicated excessive mutation density in insulators (162 mutations/Mb; Figure 1c).

The mutations recovered were predominantly C→T transitions (average 82.5%, range 76.1% to 90.8%; Figure 1d).

Mutation Signature Analysis

De novo analysis revealed three predominant mutation signatures in our cohort (Signature I, II, III; Figure 2a), which clustered with previously reported signatures 7a and 7b (Alexandrov et al., 2018; Figure 2b). Signature 7 features predominantly C→T mutations, and the sub-signatures 7a, 7b, 7c, and 7d are driven by different mutational processes triggered by UV exposure (Alexandrov et al., 2018). Signatures 7a and 7b may reflect direct impact of photoproducts, while 7c and 7d may represent indirect processes such as erroneous repair (Alexandrov et al., 2018). Signatures 7a and 7b were observed in all samples, while 7c and 7d were present in four and two cases, respectively (Figure 2c). Signature 32 was strongly evident in a single patient (case 06). Signature 58, thought to denote a sequencing artifact, was present in seven of the 15 cases (46.7%), and its presence was not associated with tumor burden, or the sequencing laboratories.

Analysis of published whole-exome data of six metastases and 26 primary cSCC (Pickering et al., 2014), demonstrated similar dominance of signatures 7a and 7b in both primaries and metastases, while the heterogeneity of other contributing signatures was greater than in our cohort (Figure 2d). To assess whether the signature pattern allows the distinction of cSCC from mucosal SCC, signature analysis on whole-exome data of 44 oral cavity SCC primaries from The Cancer Genome Atlas (TCGA; Lawrence et al., 2015) was performed (Figure 2e). Most (41 of 44) oral cavity SCC featured a clearly distinct mutation pattern. Two samples from the lip (a UV exposed site) and another for which the anatomical location was not specified showed similar signatures to our cSCC cohort.

CTCF binding Site Mutations

The high prevalence of mutations in insulator regions prompted an investigation of the mutational distribution at DNA binding sites of CTCF, the main human insulator protein. We applied a strict 13 base pair (bp) motif as previously described (Poulos et al., 2016) (Figure 3a), excluding three low-confidence bp at both ends of the consensus CTCF binding site (CTCFbs) motif of the JASPAR database (Mathelier et al., 2016). The number of recognized CTCFbs totaled 5,470. The mutation density at CTCFbs was significantly increased compared to their flanking regions (Figure 3b) and mutations were predominantly C→T transitions (Figure 3c). The average number of mutated motifs per sample was 84.5 per Mb (range 14 to 223; median 64; IQR 30.5 to 116.5). Across the cohort, a total of 1,404 mutations were detected at the 5,470 examined sites, which showed little overlap between patients: 1,026 (18.8%) motifs were mutated in only one sample, while 219 (4.0%) were mutated in two, 42 (0.8%) in three, and 22 (0.4%) were mutated over four samples. When regarding both, the upper and the lower strand of the DNA, the conserved 13 bp CTCF motif harbors cytosine in up to 11 positions (minimum 7) and accounts for up to eight pyrimidine pairs, three of which are highly conserved (Douki and Cadet, 2001; Marteijn et al., 2014) (Figure 2a). Throughout the cohort, the highest mutation density was detected at dipyrimidine positions 10 and 11 (Figures 3d, 3e). When we repeated the analysis using control motifs that only match positions 8 to 13 of the motif (Poulos et al., 2016), the predilection for mutations in positions 10 and 11 was nearly eliminated (Figure 3d), proving that it only occurs in the context of a conserved CTCFbs.

The potential impact of the CTCFbs mutations on TADs was then assessed. From an initial 4929 TADs identified in normal human epidermal keratinocytes (NHEK) (Rao et al., 2014), 903 TADs with CTCFbs at both ends were extracted (Figure 4a). Of these, 422 (46.7%) were identified to have a mutated CTCF motif in at least one of the delimiting anchor regions (across the cohort), and 47 (5.2%) had mutations in both anchor regions. Within the affected loops, we detected 1,979 genes, including 38 oncogenes (Liu et al., 2017), 52 tumor

suppressor genes (Zhao et al., 2016), and eleven identified as potential tumor drivers (Bailey et al., 2018; Table 2). Analysis of a subset (30) of these genes located in the affected TADs did not correlate with a change of expression levels between samples that carried the CTCFbs mutations and those that did not (Figure S2).

DISCUSSION

Metastases of cSCC Exhibit UV-Induced Mutation Signatures

This is the first study to apply WGS and signature analysis on metastases of head and neck cSCC. Our results confirm that UV-associated mutation signature 7 is present in both cSCC primaries and metastases (Figure 2c, 2d), and that the mutation signature pattern clearly differs from that of mucosal human papilloma virus (HPV) negative head and neck SCC primary lesions (Figure 2e). The mutation analysis of oral cavity SCC was performed on primaries, as we are not aware of publicly available data for metastases of mucosal SCC. However, we assume that their signatures correspond to those found in metastases, as we have shown for cSCC in this study (Figure 2c, 2d). Notably, the TCGA oral cavity SCC cohort includes lip tumors (Figure 2e), which are no longer classified as oral cavity SCC by the AJCC, but as cSCC as most are solar induced (Amin et al., 2017).

Thus, algorithmic signature analysis is able to distinguish cSCC metastasis from squamous cell carcinoma of different origin. This distinction is impossible based on clinical or histo-pathological parameters alone, except in the case of HPV positive oropharyngeal SCC (Satgunaseelan et al., 2017), and is clinically relevant with regards to prognosis and treatment. Four patients in our cohort had no known primary lesion. All four patients were Anglo-celtic Australians who had acquired actinic damage to the skin of the head, and mucosal examination was unremarkable, making a cutaneous origin of their parotid metastases highly

probable. Consistently, UV-signature was strongly present in the tumor tissue of all four patients (Figure 2c). Apart from mucosal SCC, signature analysis may also help to distinguish cSCC from primary SCC of the salivary gland, an extremely rare entity that is to date a diagnosis of exclusion. Many of these presumed primaries may in fact represent metastases for which a cutaneous primary was not recognized (Chen et al., 2015), especially in the parotid gland. Due to the rarity of this entity and the subsequent lack of published sequencing data, its signature profile is unknown, but a significant UV-associated signature is unlikely. The predominance of C→T transitions is not unique to UV-induced damage, being associated with other factors (e.g. ageing processes, alkylating agents exposure), and delivers insufficient information to clearly distinguish mutation patterns. Signature analysis based on the trinucleotide context overcomes this problem. The extent of the contribution of signatures other than UV-signature is variable, and depends on the type of analyzed data. We believe that the increased heterogeneity in signatures derived from the whole-exome data of Pickering at al. (2014) compared to our WGS data (Figure 2c, 2d) is because the great majority of mutations in cSCC are located in the non-coding DNA (Figure 1a). Thus, WGS increases the power of signature analysis. The publicly available data from Pickering et al. (2014) is filtered, which may also have an impact on signatures. Furthermore, the reduced effectiveness of repair mechanisms in the non-coding DNA (Budden and Bowden, 2013; Frigola et al., 2017) may also be reflected in signature analysis of WGS data.

Consistent with recent work by Inman et al. (2018), case 06, who was under azathioprine treatment for rheumatoid arthritis, showed strong contribution of signature 32 (Figure 2c). Azathioprine increases the risk of skin cancer through inhibition of nucleotide excision repair (NER) of UV-induced DNA damage (Coghill et al., 2016). Azathioprine causes a C→A bias (Zhang et al., 2007), which occurs alongside the UV-induced C→T mutations, and is the hallmark of Signature 32. Consistently, the rate of C→A transitions was more than double in case 06 compared to all others (6.89% vs. mean 2.97%, range 1.43 to 4.1%, Figure 1d).

UV-induced Mutations at Insulator Sites

The mutational burden was 171-fold higher in the non-coding regions than in coding regions of the genome (Figures 1a, 1b). This concentration may reflect less effective mismatch repair in non-coding regions (Frigola et al., 2017) and the higher priority of the transcription coupled NER compared to global NER (Budden and Bowden, 2013). The subsequently identified concentration of mutations in insulator regions (Figure 1c) may have implications on regulatory processes. Considering that a specific tumor driving gene pattern has not yet emerged in cSCC, such alternative carcinogenic models are of special interest. We therefore assessed the main human insulator CTCF, whose DNA binding site seems prone to UV-induced damage because of its high cytosine and dipyrimidine content. CTCF simultaneously binds to specific base pair sequences at multiple DNA sites, thereby approximating distant chromatin regions and forming 3-dimensional DNA loops termed topologically associated domains (TADs; Hnisz et al., 2016; Kemp et al., 2014). These can incorporate multiple genes, whose expression is dependent on the binding status of CTCF (Holwerda and de Laat, 2013; Ong and Corces, 2014; Figure 4b). Loss of CTCF function through mutation of its binding site can disrupt TADs and alter the transcriptional activity of the associated genes. This can affect the cell phenotype and potentially plays a role in carcinogenesis (Hanssen et al., 2017; Hnisz et al., 2016; Kemp et al., 2014; Tang et al., 2015).

We observed mutated CTCFbs in all 15 cases and the prevalence was considerably higher compared to melanoma (Poulos et al., 2016) (mean 84.5 vs. 11.4 mutated motifs/patient).The mutation density in the CTCF motif was not only higher than in the flanking region, but exceeded the expected concentration based on the high cytosine and dipyrimidine content (Figure 3b, 3cc). This clustering of mutations at the CTCFbs was not only previously described in melanoma (Poulos et al., 2016), but also in colorectal cancer, which is not UV associated (Katainen et al., 2015), and is explained by the fact that bound CTCF blocks access

of NER enzymes to the CTCFbs (Sabarinathan et al., 2016). Consistently, cSCC patients with dysfunctional NER have comparable mutation density at flanking regions and CTCF motifs (Poulos et al., 2016). Interestingly, the two cases in our cohort under treatment with immunosuppressants known to inhibit NER, did not exhibit flattening of the mutational peak at the CTCFbs (Figure S1). However, this could be explained by the fact that NER is only partially deactivated by these immunosuppressants, whereas the impact of XPC-deficiency is more deleterious (Budden and Bowden, 2013; Kuschal et al., 2012). Finally, mutations were unevenly distributed within the CTCF motif. Strikingly, dypirimidine positions 10 and 11 comprised the vast majority of all mutations, whereas mutation density was comparably low at other dipyrimidine or cytosine positions (Figure 3c, 3d). This is consistent with findings by Poulos et al. (2016), who suggest that differential rates of repair at specific motif positions cause this asymmetry.

Assuming that CTCF cannot efficiently bind modified binding sites, mutations potentially result in disruption of TADs and dysregulation of their enclosed genes (Figure 4b). Even our conservative approach yielded 1,979 genes within 422 affected TADs and included tumor suppressor genes and oncogenes. (Table 2). We cannot ascertain at this stage whether the identified mutations at CTCFbs contribute to carcinogenesis or are just passenger mutations. Using Nanostring technology we did not detect significant changes in expression for genes that were located within affected TADs when compared to samples in which the same TAD was not affected (total genes tested = 30; Figure S2). However, an adequate determination of differential expression would require much larger cohorts (including UV-affected normal skin) and gene panels. Poulos et al. (2016) reported similar issues even when using RNA-Seq in 36 melanoma samples. Nevertheless, they were able to show that expression of cancer-associated genes in affected TAD loops was statistically different from wildtype TAD loops. Ultimately, the assessment of the clinical impact of CTCFbs mutations on gene expression and carcinogenesis in cSCC would require functional analysis using patient-derived primary

cell cultures.

Conclusion

Mutation signature analysis has potential future clinical use, as it allows cSCC metastases to be distinguished from other SCC when a primary cannot be identified. This is of particular value given the absence of uniquely characteristic driver gene mutations in cSCC. An alternative mechanism to driver gene mutations in cSCC may lie in UV-induced damage to regulatory elements in the DNA, such as the heavily mutated binding site of the insulator and master transcription factor CTCF assessed in this study. Functional studies are necessary to further explore such possibly carcinogenic models, and future research should therefore also be directed to non-coding and regulatory regions of the DNA.

MATERIALS AND METHODS

Sample Processing and Sequencing

Patients were prospectively enrolled between March 2015 and July 2017. Informed consent and approval from the institutional ethics committee were sought prior to the study (UOW/ISLHD HREC 14/397). Samples were snap frozen and underwent histopathology review to select areas with high neoplastic content (30-90%).

DNA was extracted using Qiagen AllPrep DNA/RNA mini Kit (Qiagen, Venlo, Netherlands). Quantitation and purity of DNA were measured using NanoDrop spectrophotometry (Thermo Fisher Scientific, North Ryde, Australia). DNA integrity was assessed by agarose gel electrophoresis and in-house analyses. WGS was performed by Genome.One (Darlinghurst, Australia) and Macrogen (Seoul, South Korea) on Illumina HiSeq X to a depth of 30-45X for normal and 65-90X for tumor samples. Mapping and variant calling was performed as per Tsoli et al. (2018).

Distribution of Mutations and Signature Analysis

The distribution of mutations was assessed across broad genome regions using the heterochromatin regions, the universal promoter and enhancer DHS datasets from Perera et al. (2016), and the CTCF chromatin immunoprecipitation sequencing (ChIPSeq) data for insulator regions of NHEK from ENCODE (The ENCODE Project Consortium, 2012). From a 96 trinucleotide mutation count matrix, we extracted signatures *de novo* applying non-negative matrix factorization (Brunet et al., 2004) using Maftools (Mayakonda and Koeffler, 2016) in R version 3.5.0 (R Development Core Team, 2008). We also applied the matrix of mutational signature weights from Alexandrov et al. (2018) to assess similarity and coherence of the mutation profiles with recognized signatures usingDeconstructSigs (Rosenthal et al., 2016). Signatures were reported when attributable mutations contributed >6% in any sample. The same approach was applied on publicly available VCFs from whole-exome sequencing data of cSCC (Pickering et al., 2014) and oral cavity SCC from TCGA (Lawrence et al., 2015).

CTCF Binding Site Mutation Assessment

To assess mutations at CTCFbs, filtering occurred in two steps to limit false positives, as reported by Poulos et al. (2016). First, we selected CTCFbs harboring a strict motif including the central 13 bp of the 19 bp consensus motif (Mathelier et al., 2016). Second, we identified the sites occurring at ChIP-Seq peaks for CTCF in NHEK available from ENCODE, thus assuring that assessed sites have bound CTCF. Binding sites were overlaid with our cohort's WGS data to quantify mutation density at every base in the motif and the neighboring 1kb. Results were normalized to mutations per megabase. We repeated the analysis using control motifs generated by Poulos et al. (2016) from NHEK ChIP-Seq peaks that only match positions 8 to 13 of the motif, while positions 1 to 7 could be any bases other than those

observed in CTCF motifs.

To identify and locate the CTCFbs at boundaries of TADs, we used chromosome conformation capture (Hi-C) TAD maps of NHEK (Rao et al., 2014) and ChIP-seq data from ENCODE. Hereby, a more permissive 20 bp motif is tolerated, and a 20 bp position weighted matrix (Kim et al., 2007) is applied to select for binding sites with high CTCF-binding probability. TADs were filtered as illustrated in Figure 4a. TADs were excluded if they had more than one CTCF binding motif in either anchor region, and if CTCF motifs were not in convergent orientation, since this arrangement has the highest association with binding of CTCF (Rao et al., 2014). Genomic coordinates delimitating a TAD were defined as the 3'-end of the upstream and the 5'-end of the downstream motif. Genes lying in TADs were identified from the UCSC genome browser (Kent et al., 2002) using BEDTools (Quinlan and Hall, 2010).

Gene expression analyses

RNA was extracted using Qiagen AllPrep DNA/RNA mini Kit (Qiagen, Venlo, Netherlands). Quantitation and purity of RNA were measured using NanoDrop spectrophotometry (Thermo Fisher Scientific, North Ryde, Australia). RNA expression was evaluated by NanoString nCounter Sprint system using the PanCancer Progression panel (Nanostring, Seattle, WA, USA) as per the manufacturer's instructions. Results were analyzed using NanoString nSolver 4.0 and Advanced Analysis Module (Nanostring, Seattle, WA, USA), which normalizes gene expression to a set of positive and negative controls genes. The expression of genes found in a mutated TAD in at least one sample in our cohort was compared to that of the other samples of the cohort using a heatmap.

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGEMENTS

We extend our thanks to Elahe Minaei and Jay Perry of Illawarra Health and Medicine Research Institute, Wollongong, Australia, for sample processing, Assistant Prof. Ludmil Alexandrov, University of California San Diego, CA, for sharing the updated mutation signature matrices, and Associate Prof. Narayanan Gopalakrishna Iyer, National Cancer Center Singapore, for conceptual support. MJC is supported by a NSW Health Early-Mid Career Fellowship. We thank the Kinghorn Centre for Clinical Genomics for assistance with production and processing of whole-genome sequencing data.

REFERENCES

Alexandrov L, Kim J, Haradhvala NJ, Huang MN, Ng AWT, Boot A, et al. The repertoire of mutational signatures in human cancer. BioRxiv 2018:322859. doi:10.1101/322859.

Amin MB, Edge SB, American Joint Committee on Cancer. AJCC cancer staging manual. Springer International Publishing; 2017.

Ashford BG, Clark J, Gupta R, Iyer NG, Yu B, Ranson M. Reviewing the genetic alterations in high-risk cutaneous squamous cell carcinoma: A search for prognostic markers and therapeutic targets. Head Neck 2017;39:1462–9. doi:10.1002/hed.24765.

Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. Cell 2018;173:371–385.e18. doi:10.1016/j.cell.2018.02.060.

Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 2004;101:4164–9. doi:10.1073/pnas.0308531101.

Budden T, Bowden NA. The role of altered nucleotide excision repair and UVB-induced DNA damage in melanomagenesis. Int J Mol Sci 2013;14:1132–51. doi:10.3390/ijms14011132.

Chen MM, Roman SA, Sosa JA, Judson BL. Prognostic factors for squamous cell cancer of the parotid gland: An analysis of 2104 patients. Head Neck 2015;37:1–7. doi:10.1002/hed.23566.

Coghill AE, Johnson LG, Berg D, Resler AJ, Leca N, Madeleine MM. Immunosuppressive medications and squamous cell skin carcinoma: nested case-control study within the skin cancer after organ transplant (SCOT) cohort. Am J Transplant 2016;16:565–73. doi:10.1111/ajt.13596.

Douki T, Cadet J. Individual determination of the yield of the main UV-induced dimeric pyrimidine photoproducts in DNA suggests a high mutagenicity of CC photolesions. Biochemisrty 2001;40:2495–501. doi:10.1021/BI0022543.

Durinck S, Ho C, Wang NJ, Liao W, Jakkula LR, Collisson EA, et al. Temporal dissection of tumorigenesis in primary cancers. Cancer Discov 2011;1:137–43. doi:10.1158/2159-8290.CD-11-0028.

D'Souza J, Clark J. Management of the neck in metastatic cutaneous squamous cell carcinoma of the head and neck. Curr Opin Otolaryngol Head Neck Surg 2011;19:99–105. doi:10.1097/MOO.0b013e328343e811.

Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res 2017;45:D777–83. doi:10.1093/nar/gkw1121.

Frigola J, Sabarinathan R, Mularoni L, Muiños F, Gonzalez-Perez A, López-Bigas N. Reduced mutation rate in exons due to differential mismatch repair. Nat Genet 2017;49:1684–92. doi:10.1038/ng.3991.

Gurudutt V V, Genden EM. Cutaneous squamous cell carcinoma of the head and neck. J Skin Cancer 2011;2011:502723. doi:10.1155/2011/502723.

Hanssen LLP, Kassouf MT, Oudelaar AM, Biggs D, Preece C, Downes DJ, et al. Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. Nat Cell Biol 2017;19:952–61. doi:10.1038/ncb3573.

Hnisz D, Weintraub AS, Day DS, Valton A-L, Bak RO, Li CH, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science 2016;351:1454–8. doi:10.1126/science.aad9024.

Holwerda SJB, de Laat W. CTCF: the protein, the binding partners, the binding sites and their

chromatin loops. Philos Trans R Soc Lond B Biol Sci 2013;368:20120369. doi:10.1098/rstb.2012.0369.

Inman GJ, Wang J, Nagano A, Alexandrov LB, Purdie KJ, Taylor RG, et al. The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature. Nat Commun 2018;9:3667. doi:10.1038/s41467-018-06027-1.

Katainen R, Dave K, Pitkänen E, Palin K, Kivioja T, Välimäki N, et al. CTCF/cohesin-binding sites are frequently mutated in cancer. Nat Genet 2015;47:818–21. doi:10.1038/ng.3335.

Kemp CJ, Moore JM, Moser R, Bernard B, Teater M, Smith LE, et al. CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. Cell Rep 2014;7:1020–9. doi:10.1016/j.celrep.2014.04.004.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res 2002;12:996–1006. doi:10.1101/gr.229102.

Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 2007;128:1231–45. doi:10.1016/j.cell.2006.12.048.

Kuschal C, Thoms K-M, Schubert S, Schäfer A, Boeckmann L, Schön MP, et al. Skin cancer in organ transplant recipients: effects of immunosuppressive medications on DNA repair. Exp Dermatol 2012;21:2–6. doi:10.1111/j.1600-0625.2011.01413.x.

Lawrence MS, Sougnez C, Lichtenstein L, Cibulskis K, Lander E, Gabriel SB, et al. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 2015;517:576–82. doi:10.1038/nature14129.

Liu Y, Sun J, Zhao M. ONGene: A literature-based database for human oncogenes. J Genet Genomics 2017;44:119–21. doi:10.1016/J.JGG.2016.12.004.

Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. Understanding nucleotide excision repair and its roles in cancer and ageing. Nat Rev Mol Cell Biol 2014;15:465–81. doi:10.1038/nrm3822.

Mathelier A, Fornes O, Arenillas DJ, Chen C, Denay G, Lee J, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 2016;44:D110–5. doi:10.1093/nar/gkv1176.

Mayakonda A, Koeffler HP. Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. BioRxiv 2016:052662. doi:10.1101/052662.

Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell 2012;149:979–93. doi:10.1016/j.cell.2012.04.024.

Ong C-T, Corces VG. CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet 2014;15:234–46. doi:10.1038/nrg3663.

Perera D, Poulos RC, Shah A, Beck D, Pimanda JE, Wong JWH. Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. Nature 2016;532:259–63. doi:10.1038/nature17437.

Pickering CR, Zhou JH, Lee JJ, Drummond JA, Peng SA, Saade RE, et al. Mutational landscape of aggressive cutaneous squamous cell carcinoma. Clin Cancer Res 2014;20:6582–92. doi:10.1158/1078-0432.CCR-14-1768.

Poulos RC, Thoms JAI, Guan YF, Unnikrishnan A, Pimanda JE, Wong JWH. Functional Mutations Form at CTCF-Cohesin Binding Sites in Melanoma Due to Uneven Nucleotide Excision Repair across the Motif. Cell Rep 2016;17:2865–72. doi:10.1016/j.celrep.2016.11.055.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2. doi:10.1093/bioinformatics/btq033.

R Development Core Team. R: A language and environment for statistical computing. R Found Stat Comput Vienna, Austria 2008:ISBN 3-900051-07-0, URL http://www.R-project.org.

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 2014;159:1665–80. doi:10.1016/j.cell.2014.11.021.

Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol 2016;17:31. doi:10.1186/s13059-016-0893-4.

Sabarinathan R, Mularoni L, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. Nature 2016;532:264–7. doi:10.1038/nature17661.

Satgunaseelan L, Chia N, Suh H, Virk S, Ashford B, Lum T, et al. p16 expression in cutaneous squamous cell carcinoma of the head and neck is not associated with integration of high risk HPV DNA or prognosis. Pathology 2017;49:494–8. doi:10.1016/j.pathol.2017.04.002.

South AP, Purdie KJ, Watt SA, Haldenby S, den Breems N, Dimon M, et al. NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. J Invest Dermatol 2014;134:2630–8. doi:10.1038/jid.2014.154.

Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell 2015;163:1611–27. doi:10.1016/j.cell.2015.11.024.

The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74. doi:10.1038/nature11247.

Tsoli M, Wadham C, Pinese M, Failes T, Joshi S, Mould E, et al. Integration of genomics, high throughput drug screening, and personalized xenograft models as a novel precision medicine paradigm for high risk pediatric cancer. Cancer Biol Ther 2018:1–10. doi:10.1080/15384047.2018.1491498.

Zhang X, Jeffs G, Ren X, O'Donovan P, Montaner B, Perrett CM, et al. Novel DNA lesions generated by the interaction between therapeutic thiopurines and UVA light. DNA Repair (Amst) 2007;6:344–54. doi:10.1016/J.DNAREP.2006.11.003.

Zhao M, Kim P, Mitra R, Zhao J, Zhao Z. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acids Res 2016;44:D1023-31. doi:10.1093/nar/gkv1268.

Zilberg C, Lee MW, Yu B, Ashford B, Kraitsek S, Ranson M, et al. Analysis of clinically relevant somatic mutations in high-risk head and neck cutaneous squamous cell carcinoma. Mod Pathol 2018;31:275–87. doi:10.1038/modpathol.2017.128.

TABLES

Table 1. Demographic and medical data of the cohort of 15 patients with cSCC lymph node metastases.

Sample	Age years	Sex	Primary location	Metastasis location	Nodal category	Tumor stage	LN ratio	ECS	Grade	Immuno-suppression
1	30	M	Left lip	Left neck	N3b	IV	3/27	Yes	1	No
2	78	M	Right ear	Right parotid	N3b	IV	2/52	Yes	3	No
3	74	M	Unknown	Right parotid	N3b	IV	2/42	Yes	3	No
4	64	M	Bilateral lip	Bilateral neck	N2c	IV	3/55	No	2	No
5	78	M	Left forehead	Left parotid	N2a	IV	unknown	Yes	3	No
6	69	M	Left cheek	Left neck	N3b	IV	4/4	Yes	3	Azathioprine
7	87	M	Unknown	Left neck	N2b	IV	2/42	No	3	No
8	87	M	Unknown	Left parotid	N3b	IV	1/16	Yes	2	No
9	66	M	Bilateral forehead	Right neck	N3b	IV	2/29	Yes	2	Cyclosporine A, Tacrolimus
10	64	M	Left scalp	Left neck	N3b	IV	3/109	Yes	3	No
11	69	M	Unknown	Right parotid	N3b	IV	2/11	Yes	3	No
12	77	M	Right nose	Right neck	N3b	IV	3/108	Yes	2	No
13	77	M	Right ear	Right parotid	N3b	IV	4/64	Yes	2	No
14	79	F	Left cheek	Left perifacial	N3b	IV	unknown	Yes	3	No
15	66	M	Left scalp	Left scalp	N2b	IV	2/2	No	2	No

ECS, extracapsular spread; LN ratio, number of lymph node metastases of total resected lymph nodes

Table 2. Genes located in TAD loops with mutated CTCF motifs across the cohort of 15 cSCC lymph node metastases which are recognized as tumor suppressor genes (TSGene, Zhao et al., 2016) oncogenes (ONGene, Liu et al., 2017), or identified as potential tumor driver genes (Bailey et al., 2018).

Tumor-suppressor genes			Oncogenes			Tumor driver genes
ASCL1	*IQGAP2*	*NOV*	*ASCL1*	*MALAT1*	*ZBTB16*	*CACNA1A*
BASP1	*KAT5*	*ONECUT1*	*BMI1*	*MAP3K8*	*ZEB1-AS1*	*CARD11*
BMP2	*KLF6*	*PARK2*	*BOC*	*MFHAS1*		*DICER1*
BMP4	*L3MBTL4*	*PAX6*	*CAD*	*MLLT3*		*ELF3*
CDO1	*LEFTY1*	*PIWIL2*	*CARD11*	*NEAT1*		*IL7R*
CKLF	*LEFTY2*	*PLAGL1*	*CCDC6*	*NEDD9*		*KEL*
COPS2	*LRIG1*	*POU6F2*	*CDC25A*	*NOV*		*PIK3CA*
CREM	*MAP3K8*	*PPARA*	*CKLF*	*NUP214*		*PLCB4*
CXCL14	*MAP4K1*	*PPP2CA*	*EPS8*	*PAK7*		*RAD21*
DCDC2	*MAT2A*	*PRDM2*	*GLI2*	*PIK3CA*		*RPS6KA3*
DCLRE1A	*MIR1226*	*PRKCE*	*HMGA1*	*PRKCA*		*SPTA1*
DICER1	*MT1F*	*RASAL1*	*HSPB1*	*PRKCE*		
EPHB3	*MT1G*	*RASAL2*	*ID2*	*RAB23*		
ESRRB	*MT1M*	*RUNX2*	*IL7R*	*SMURF1*		
FHIT	*MT2A*	*SRGAP3*	*KLF6*	*STMN1*		
FOXO3	*NCAM2*	*ST7*	*KSR2*	*TAC1*		
GDA	*NEDD4L*	*ZBTB16*	*LMO2*	*TNFRSF1B*		
HIVEP1			*MAFB*	*TWIST1*		

FIGURE LEGENDS

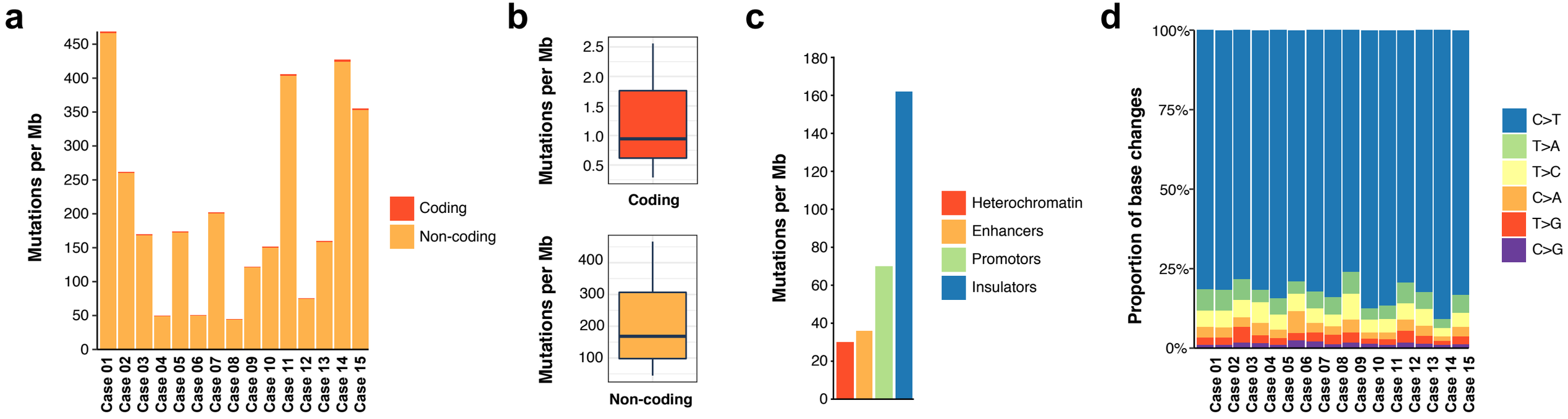
Figure 1. Mutation landscape across the cohort of 15 cSCC lymph node metastases. (a) Mutation burden per patient in coding and non-coding DNA. (b) Boxplot showing median number of mutations per megabase (Mb) in the coding and non-coding DNA. (c) Mutation density across different regions of the genome. (d) Distribution at single base level shows predominance of C→T transitions.

Figure 2. Mutation signature analysis. (a) *De novo* signature analysis of 15 cSCC lymph node metastases yielded three distinct signatures based on the contribution of trinucleotide context of each mutation. (b) Heatmap showing unsupervised clustering of *de novo* extracted signatures with recognized signatures (Alexandrov et al., 2018). All three *de novo* extracted signatures are most closely associated with UV-associated signatures 7a or 7b. (c) Signature profiles of individual samples of the cohort. Signatures contributing less than 6% were collapsed. (d) Signature profiles derived from whole-exome data of 26 cSCC primaries and 5 lymph node metastases from Pickering et al. (2014). (e) Signature profiles derived from whole-exome data of 44 oral cavity SCC from TCGA. Signatures contributing less than 8% were collapsed.

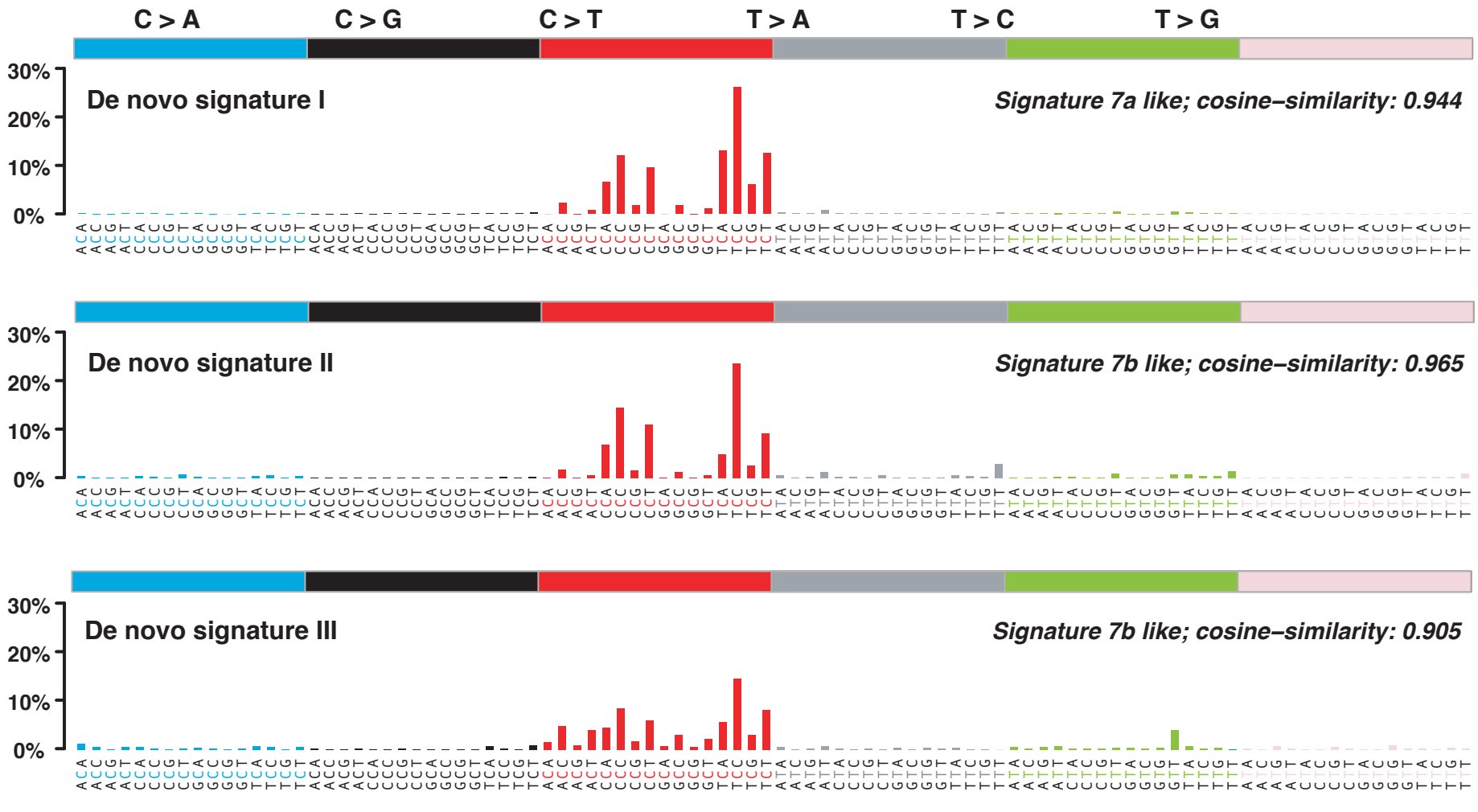
Figure 3. Mutational landscape at CTCF binding sites of 15 cSCC lymph node metastases. (a) Upper and lower strand nucleic acid sequence showing the conserved center of the CTCFbs. Large letters symbolize highly conserved positions. Potential dipyrimidine sites are indicated with brackets. (b) Overall density of mutations at merged CTCFbs and

neighboring 1000 base pairs. (c) Mutation density and mutation type in the CTCFbs motif and flanking 30 base pairs. (d) Mutation density at individual base positions of the CTCFbs derived from NHEK ChIP-seq peaks across the cohort showing concentration of mutations at position 10 and 11(blue line). The green line illustrates mutation density of a control motif, where only positions 8 to 13 match the CTCF motif (A|G)(G|T)GGC(A|G), and positions 1 to 7 can contain any nucleotide. (e) Mutation density at base positions in individual samples, demonstrating the consistency of the distribution of mutations across the cohort.

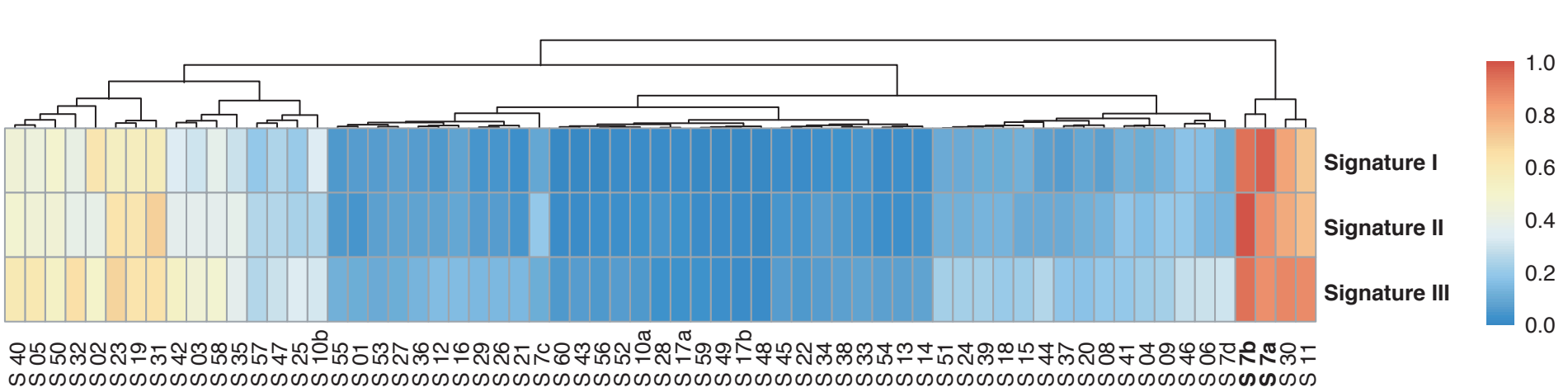
Figure 4. CTCF binding sites delimiting topologically associated domains (TADs). (a) Workflow for the identification of TADs with mutated CTCF binding sites. (b) Schematic illustration of a TAD loop regulated by CTCF. On the left, the TAD is insulated by CTCF bound to binding sites located in the anchor regions on both ends of the loop. Binding of CTCF enables the adherence of other transcription factors, such as cohesion. A proto-oncogene lies within the loop, and a downstream enhancer acting on a gene (arrow) is shown. On the right, a mutation in one of the CTCFbs (indicated with X) prevents the binding of CTCF, thus disrupting the TAD. Consequently, the downstream enhancer now acts on the proto-oncogene (dashed arrow).


a
Mutations per Mb
0
50
100
150
200
250
300
350
400
450
Case 01
Case 02
Case 03
Case 04
Case 05
Case 06
Case 07
Case 08
Case 09
Case 10
Case 11
Case 12
Case 13
Case 14
Case 15
Coding
Non-coding
b
Mutations per Mb
0.5
1.0
1.5
2.0
2.5
Coding
Mutations per Mb
100
200
300
400
Non-coding
c
Mutations per Mb
0
20
40
60
80
100
120
140
160
180
Heterochromatin
Enhancers
Promotors
Insulators
d
Proportion of base changes
0%
25%
50%
75%
100%
Case 01
Case 02
Case 03
Case 04
Case 05
Case 06
Case 07
Case 08
Case 09
Case 10
Case 11
Case 12
Case 13
Case 14
Case 15
C>T
T>A
T>C
C>A
T>G
C>G

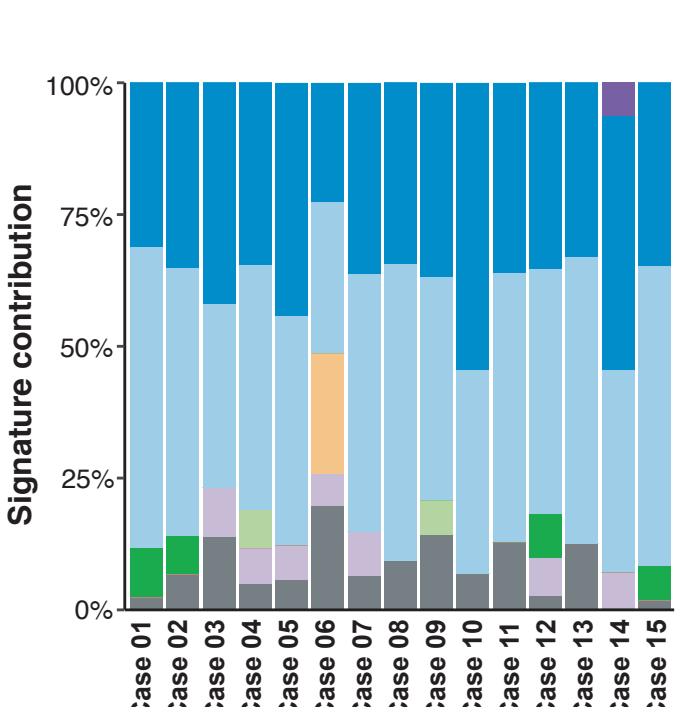
a


C > A
C > G
C > T
T > A
T > C
T > G
30%
20%
10%
0%
De novo signature I
Signature 7a like; cosine–similarity: 0.944
De novo signature II
Signature 7b like; cosine–similarity: 0.965
De novo signature III
Signature 7b like; cosine–similarity: 0.905

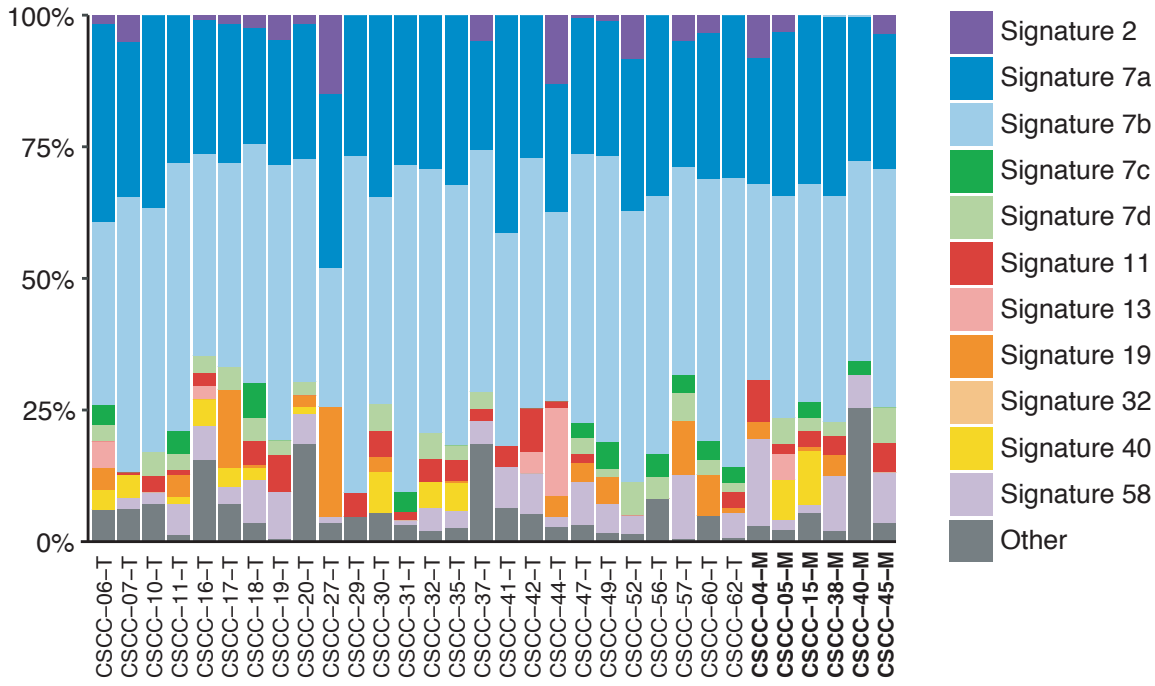
b


Signature I
Signature II
Signature III
1.0
0.8
0.6
0.4
0.2
0.0
S40 S05 S50 S32 S02 S23 S19 S31 S42 S03 S58 S35 S57 S47 S25 S10b S55 S01 S53 S27 S36 S12 S16 S29 S26 S21 S7c S60 S43 S56 S52 S10a S28 S17a S59 S49 S17b S48 S45 S22 S34 S38 S33 S54 S13 S14 S51 S24 S39 S18 S15 S44 S37 S20 S08 S41 S04 S09 S46 S06 S7d S7b S7a S30 S11

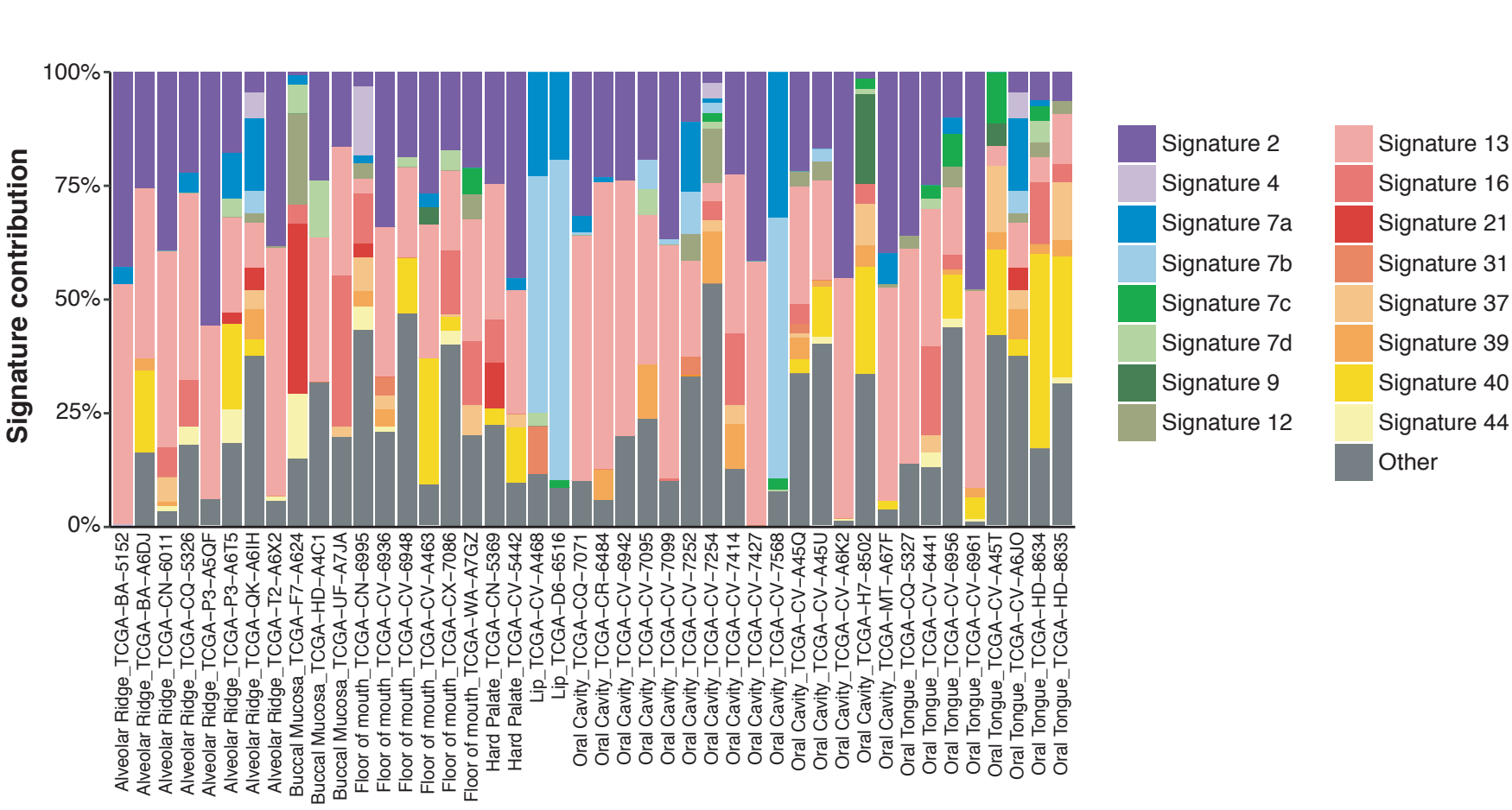
c

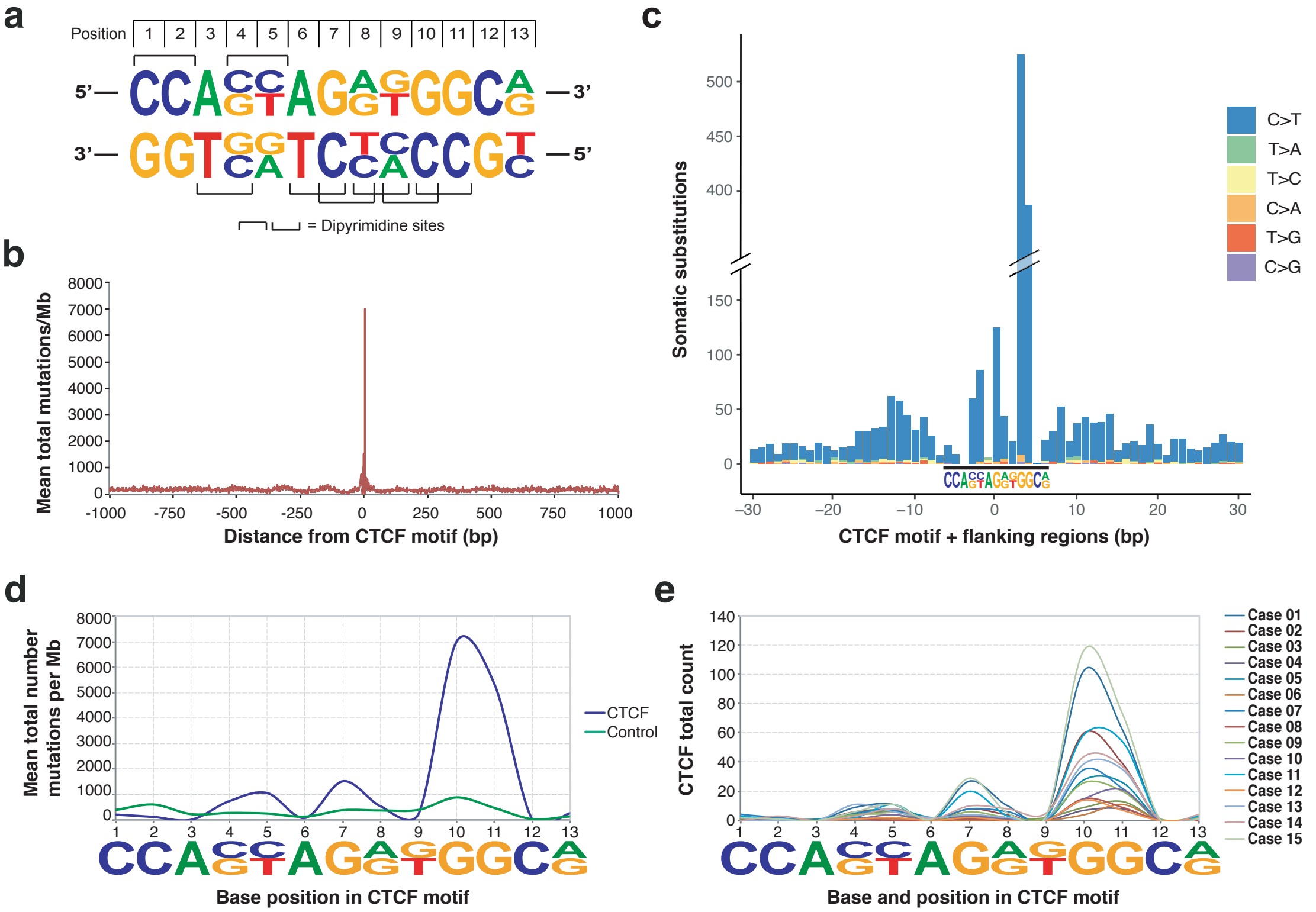

Signature contribution
100%
75%
50%
25%
0%
Case 01
Case 02
Case 03
Case 04
Case 05
Case 06
Case 07
Case 08
Case 09
Case 10
Case 11
Case 12
Case 13
Case 14
Case 15

d


100%
75%
50%
25%
0%
Signature 2
Signature 7a
Signature 7b
Signature 7c
Signature 7d
Signature 11
Signature 13
Signature 19
Signature 32
Signature 40
Signature 58
Other
CSCC–06_T
CSCC–07_T
CSCC–10_T
CSCC–11_T
CSCC–16_T
CSCC–17_T
CSCC–18_T
CSCC–19_T
CSCC–20_T
CSCC–27_T
CSCC–29_T
CSCC–30_T
CSCC–31_T
CSCC–32_T
CSCC–35_T
CSCC–37_T
CSCC–41_T
CSCC–42_T
CSCC–44_T
CSCC–47_T
CSCC–49_T
CSCC–52_T
CSCC–56_T
CSCC–57_T
CSCC–60_T
CSCC–62_T
CSCC–04–M
CSCC–05–M
CSCC–15–M
CSCC–38–M
CSCC–40–M
CSCC–45–M

e


Signature contribution
100%
75%
50%
25%
0%
Signature 2
Signature 4
Signature 7a
Signature 7b
Signature 7c
Signature 7d
Signature 9
Signature 12
Signature 13
Signature 16
Signature 21
Signature 31
Signature 37
Signature 39
Signature 40
Signature 44
Other
Alveolar Ridge_TCGA–BA–5152
Alveolar Ridge_TCGA–BA–A6DJ
Alveolar Ridge_TCGA–CN–6011
Alveolar Ridge_TCGA–CQ–5326
Alveolar Ridge_TCGA–P3–A5QF
Alveolar Ridge_TCGA–P3–A6T5
Alveolar Ridge_TCGA–QK–A6IH
Alveolar Ridge_TCGA–T2–A6X2
Buccal Mucosa_TCGA–F7–A624
Buccal Mucosa_TCGA–HD–A4C1
Buccal Mucosa_TCGA–UF–A7JA
Floor of mouth_TCGA–CN–6995
Floor of mouth_TCGA–CV–6936
Floor of mouth_TCGA–CV–6948
Floor of mouth_TCGA–CV–A463
Floor of mouth_TCGA–CX–7086
Floor of mouth_TCGA–WA–A7GZ
Hard Palate_TCGA–CN–5369
Hard Palate_TCGA–CV–5442
Lip_TCGA–CV–A468
Lip_TCGA–D6–6516
Oral Cavity_TCGA–CQ–7071
Oral Cavity_TCGA–CR–6484
Oral Cavity_TCGA–CV–6942
Oral Cavity_TCGA–CV–7095
Oral Cavity_TCGA–CV–7099
Oral Cavity_TCGA–CV–7252
Oral Cavity_TCGA–CV–7254
Oral Cavity_TCGA–CV–7414
Oral Cavity_TCGA–CV–7427
Oral Cavity_TCGA–CV–7568
Oral Cavity_TCGA–CV–A45Q
Oral Cavity_TCGA–CV–A45U
Oral Cavity_TCGA–CV–A6K2
Oral Cavity_TCGA–H7–8502
Oral Cavity_TCGA–MT–A67F
Oral Tongue_TCGA–CQ–5327
Oral Tongue_TCGA–CV–6441
Oral Tongue_TCGA–CV–6956
Oral Tongue_TCGA–CV–6961
Oral Tongue_TCGA–CV–A45T
Oral Tongue_TCGA–CV–A6JO
Oral Tongue_TCGA–HD–8634
Oral Tongue_TCGA–HD–8635


a
Position 1 2 3 4 5 6 7 8 9 10 11 12 13
5'— —3'
3'— —5'
= Dipyrimidine sites
b
Mean total mutations/Mb
8000 7000 6000 5000 4000 3000 2000 1000 0
-1000 -750 -500 -250 0 250 500 750 1000
Distance from CTCF motif (bp)
c
Somatic substitutions
500 450 400 150 100 50 0
C>T
T>A
T>C
C>A
T>G
C>G
–30 –20 –10 0 10 20 30
CTCF motif + flanking regions (bp)
d
Mean total number mutations per Mb
8000 7000 6000 5000 4000 3000 2000 1000 0
CTCF
Control
1 2 3 4 5 6 7 8 9 10 11 12 13
Base position in CTCF motif
e
CTCF total count
140 120 100 80 60 40 20 0
Case 01
Case 02
Case 03
Case 04
Case 05
Case 06
Case 07
Case 08
Case 09
Case 10
Case 11
Case 12
Case 13
Case 14
Case 15
1 2 3 4 5 6 7 8 9 10 11 12 13
Base and position in CTCF motif

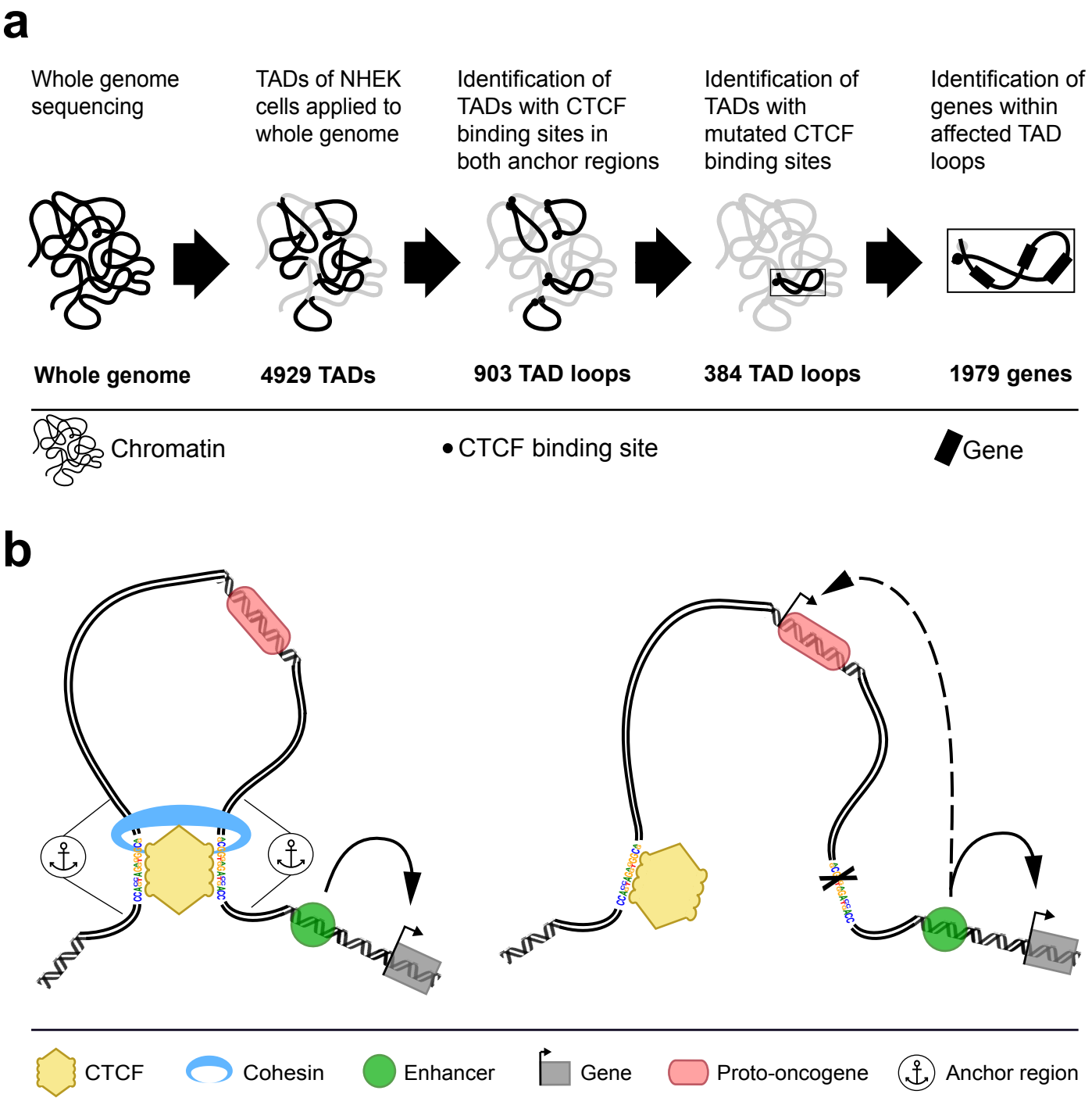
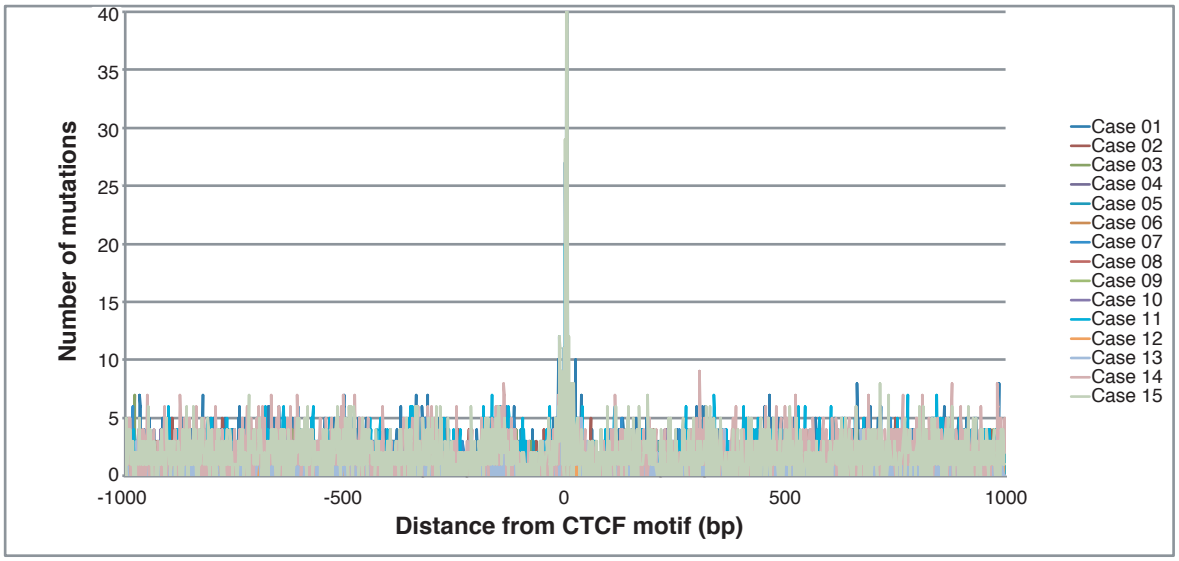

a
Whole genome sequencing
TADs of NHEK cells applied to whole genome
Identification of TADs with CTCF binding sites in both anchor regions
Identification of TADs with mutated CTCF binding sites
Identification of genes within affected TAD loops
Whole genome
4929 TADs
903 TAD loops
384 TAD loops
1979 genes
Chromatin
CTCF binding site
Gene
b
CTCF
Cohesin
Enhancer
Gene
Proto-oncogene
Anchor region

Figure S1. Overall density of mutations at CTCF binding sites and flanking regions. (a) Density of mutations in CTCF binding sites and flanking regions of in all 15 samples. (b) Density of mutations in CTCF binding sites and flanking regions in a patient treated with cyclosporine A and tacrolimus.

a



b

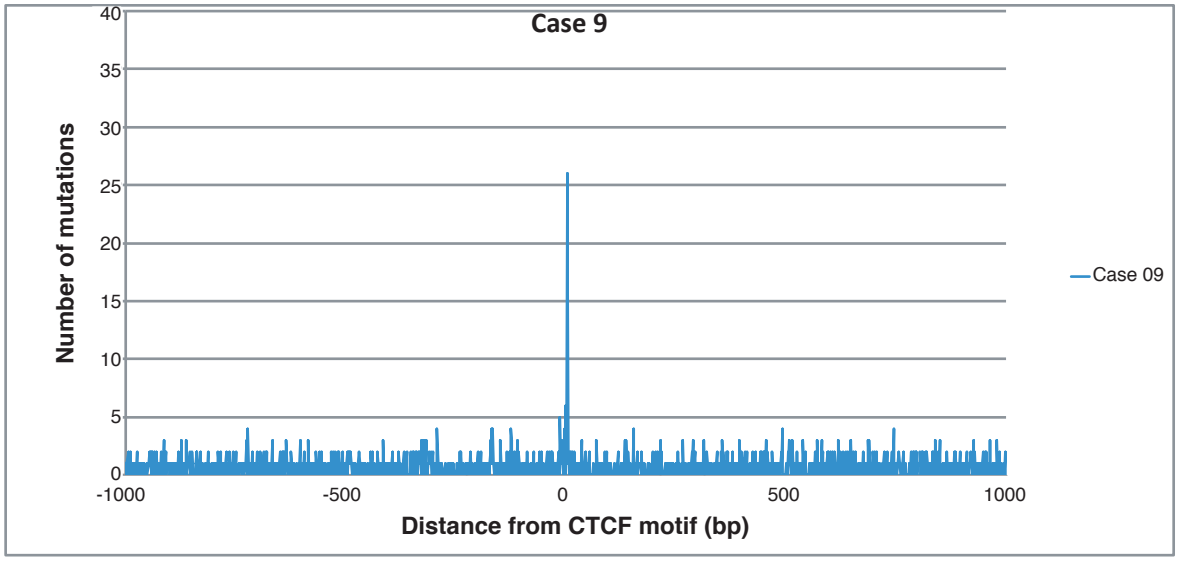


Figure S2. Expression levels of genes in TADs with mutated CTCFbs. Heatmap showing expression analysis of 30 genes lying in TADs affected by CTCFbs mutations in at least one of 15 samples of cSCC lymph node metastasis. Black dots mark samples with affected TADs.