Mutational Patterns in Metastatic Cutaneous Squamous Cell Carcinoma

Simon A. Mueller[1,2,3,14], Marie-Emilie A. Gauthier[1,2,4,14], Bruce Ashford[1,5,6,7], Ruta Gupta[1,8,9], Velimir Gayevskiy[2], Sydney Ch'ng[1,10,11], Carsten E. Palme[1,8], Kerwin Shannon[1], Jonathan R. Clark[1,8,10], Marie Ranson[6,7,12] and Mark J. Cowley[2,4,13]

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.

Journal of Investigative Dermatology (2019) ■, ■–■; doi:10.1016/j.jid.2019.01.008

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 SCC 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 predominantly include skin or mucosa, or the salivary gland. The few studies using massive parallel sequencing on cSCC have revealed a complex molecular landscape, with variation in the occurrence of alterations within cSCC on one hand, and commonality with other SCCs on the other (Ashford et al., 2017; Dotto and Rustgi, 2016; Inman et al., 2018; Lawrence et al., 2015; Pickering et al., 2014; South et al., 2014).

DNA mutation patterns may offer a solution to distinguish skin-derived 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). UV radiation 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 cSCCs (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.

[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, Bern, Switzerland;* [4]*Children's Cancer Institute, Kensington, Australia;* [5]*Illawarra and Shoalhaven Local Health District, Wollongong, Australia;* [6]*School of Biological Sciences, University of Wollongong, Wollongong, Australia;* [7]*Illawarra Health and Medical Research Institute, Wollongong, Australia;* [8]*Central Clinical School, The University of Sydney, Sydney, Australia;* [9]*Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Sydney, Australia;* [10]*Institute of Academic Surgery, University of Sydney, Sydney, Australia;* [11]*Department of Plastic Surgery, Royal Prince Alfred Hospital, Sydney, Australia;* [12]*Centre for Oncology Education and Research Translation, Liverpool, Australia; and* [13]*St Vincent's Clinical School, University of New South Wales, Sydney, Sydney, Australia*

[14]*These authors contributed equally to this work.*

Correspondence: Simon Andreas Mueller, Sydney Head and Neck Cancer Institute, Chris O'Brien Lifehouse, PO Box M33, Missenden Road, Camperdown, New South Wales, 2050, Australia. E-mail: simon.mueller@insel.ch

Abbreviations: bp, base pair; CTCFbs, CTCF binding site; cSCC, cutaneous squamous cell carcinoma; NHEK, normal human epidermal keratinocyte; NER, nucleotide excision repair; TAD, topologically associated domain; WGS, whole genome sequencing

Received 20 August 2018; revised 2 January 2019; accepted 2 January 2019; accepted manuscript published online XXX; corrected proof published online XXX

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	Immunosuppression
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

Abbreviations: ECS, extracapsular spread; F, female; LN ratio, number of lymph node metastases of total resected lymph nodes; M, male.

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 along with their trinucleotide context, and present signature analysis of cSCC metastases. We compare these signatures to those of SCC from mucosal sites to explore the clinical utility of the mutation signature analysis. We further assess the distribution of UV-induced mutations across the genome, and assess the potential impact on CTCF sites.

RESULTS

Patient characteristics

Six parotid and nine neck lymph node cSCC metastases from 15 patients were included, predominantly from males (Table 1). The primary site of the cSCC was known for 11 patients and the tumors all originated from sun-exposed face. Two patients had 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 tumor ranged from 136,105 to 1,423,398 (mean 638,254; Figure 1a). Mutations occurred mostly in noncoding regions (99.4%), and the mutation density was 171 times higher for noncoding variants compared to coding variants (mean 206.6 vs. 1.2 mutations/megabase, median 168.3 vs. 0.9 mutations/megabase, respectively; Figure 1a, 1b). Assessment of the distribution of somatic mutations across broad genome region categories indicated excessive mutation density in insulators (162 mutations/megabase; Figure 1c). The mutations recovered were predominantly C→T transitions (mean 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 subsignatures 7a, 7b, 7c, and 7d are thought to be 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 contributed a significant number of somatic mutations in a single patient (case 06). Signature 58, thought to denote a sequencing artifact, was detected in 7 of the 15 cases (46.7%), and its presence was not associated with tumor burden, or the laboratory where the tissue was sequenced.

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, we performed signature analysis on whole exome data of 44 oral cavity SCC primaries from The Cancer Genome Atlas (Lawrence et al., 2015; Figure 2e). All but three oral cavity SCCs featured a clearly distinct mutation pattern with an absence of UV-induced mutations. Two samples from the lip (a UV-exposed site) and another for which the exact anatomical location was not specified, showed patterns similar to our cSCC cohort.

CTCF binding site mutations

The high prevalence of mutations in insulator regions prompted an investigation of the mutational distribution at

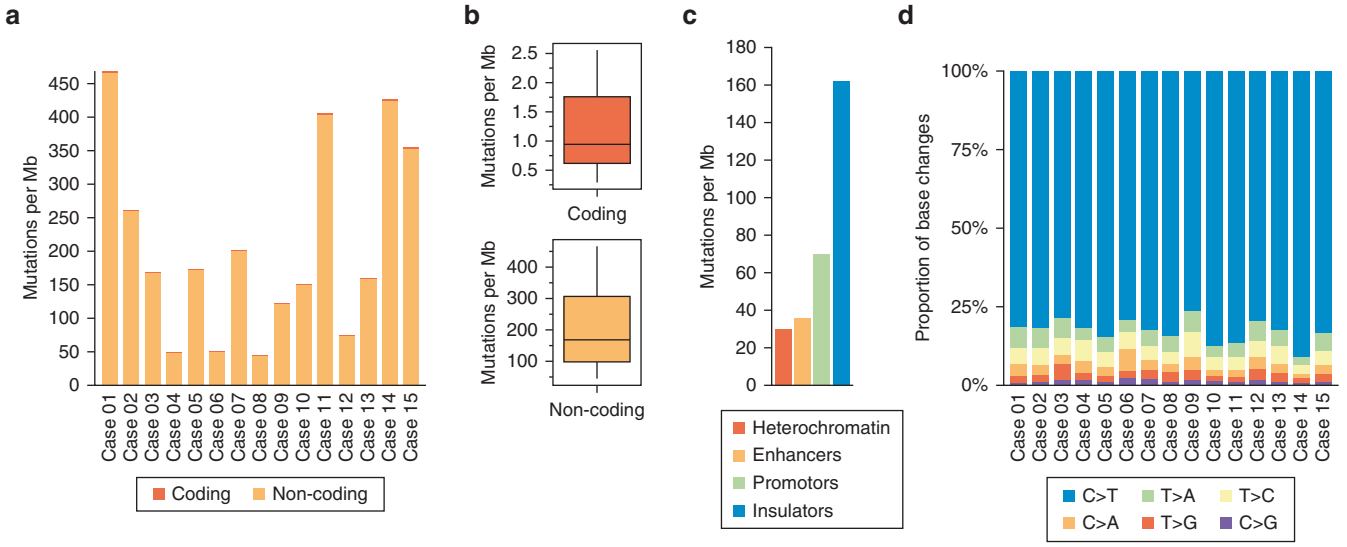


Figure 1. Mutation landscape across the cohort of 15 cutaneous squamous cell carcinoma lymph node metastases. (**a**) Mutation burden per patient in coding and noncoding DNA. (**b**) Boxplot showing median number of mutations per megabase (Mb) in the coding and noncoding DNA. (**c**) Mutation density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

DNA binding sites of CTCF, the main human insulator protein. We applied a strict 13 base pair (bp) motif as described previously (Poulos et al., 2016) (Figure 3a), excluding three low-confidence bps at both ends of the consensus CTCF binding site (CTCFbs) motif of the JASPAR database (Mathelier et al., 2016). Chromatin immunoprecipitation sequencing analysis in normal human epidermal keratinocyte cells (NHEKs) identified 5,470 CTCFbs throughout the genome. 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 CTCFbs motifs per sample was 84.5 per megabase (range 14–223; median 64; interquartile range 30.5–116.5). 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 four or more samples. When considering both, the sense and antisense 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 (Figure 3d, 3e). When we repeated the analysis using control regions in the genome that only match half of the motif, from positions 8 to 13 (Poulos et al., 2016), the predilection for mutations in positions 10 and 11 was nearly eliminated (Figure 3d), demonstrating 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 4,929 TADs identified in NHEK (Rao et al., 2014), we extracted 903 high-confidence TADs, clearly demarcated with CTCFbs at both ends with high CTCF protein binding probability (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 chromatin loops, we detected 1,979 genes, including 38 oncogenes (Liu et al., 2017), 52 tumor suppressor genes (Zhao et al., 2016), and 11 identified as potential tumor drivers (Bailey et al., 2018; Table 2).

DISCUSSION

Metastases of cSCC exhibit UV-induced mutation signatures

WGS and signature analysis on metastases of head and neck cSCCs are, to our knowledge, previously unreported. 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 primary human papilloma virus–negative mucosal SCCs (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 Cancer Genome Atlas oral cavity SCC cohort includes lip tumors (Figure 2e), which are no longer classified as oral cavity SCCs by the American Joint Committee on Cancer, but rather as cSCCs, as most are associated with UV-induced damage (Amin et al., 2017).

Thus, algorithmic signature analysis is able to distinguish cSCC metastasis from SCC of different origin. This distinction is generally not possible based on clinical or histopathological parameters alone, except in case of human papilloma virus–positive oropharyngeal SCC (Satgunaseelan et al., 2017), and is clinically relevant with regard 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 metastases highly probable. Consistently, UV signature was strongly present in the tumor

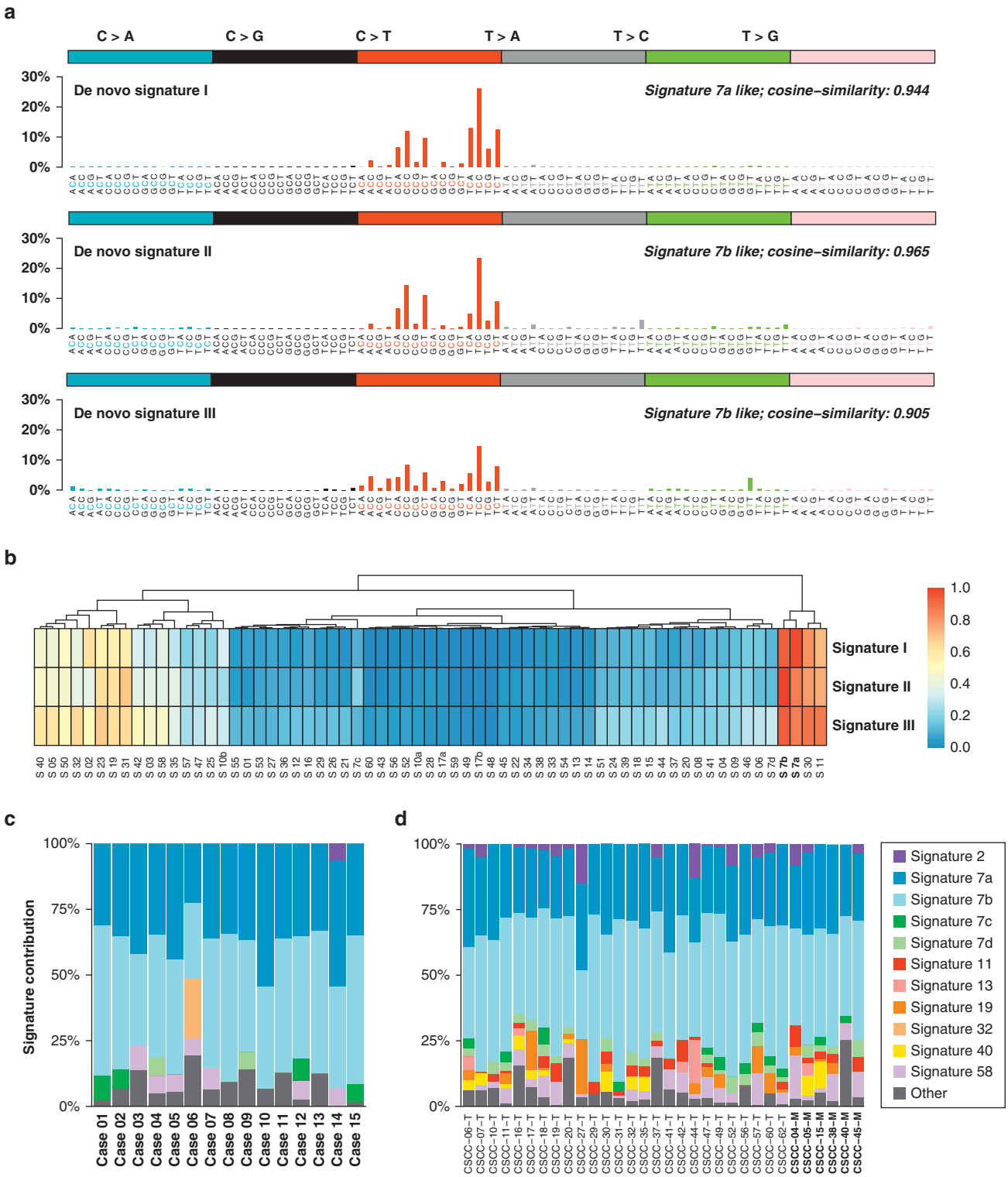


Figure 2. Mutation signature analysis. (**a**) De novo signature analysis of 15 cutaneous squamous cell carcinoma lymph node metastases yielded three distinct signatures based on the contribution of trinucleotide context of each mutation. (**b**) Heatmap showing unsupervised clustering of de novo extracted 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 <6% were collapsed. (**d**) Signature profiles derived from whole exome data of 26 cutaneous squamous cell carcinoma primaries and 5 lymph node metastases from Pickering et al. (2014). (**e**) Signature profiles derived from whole exome data of 44 oral cavity squamous cell carcinoma from The Cancer Genome Atlas. Signatures contributing <8% were collapsed.

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, diagnosed by exclusion. Many of these presumed primaries may in fact represent metastases for which a cutaneous primary was not recognized

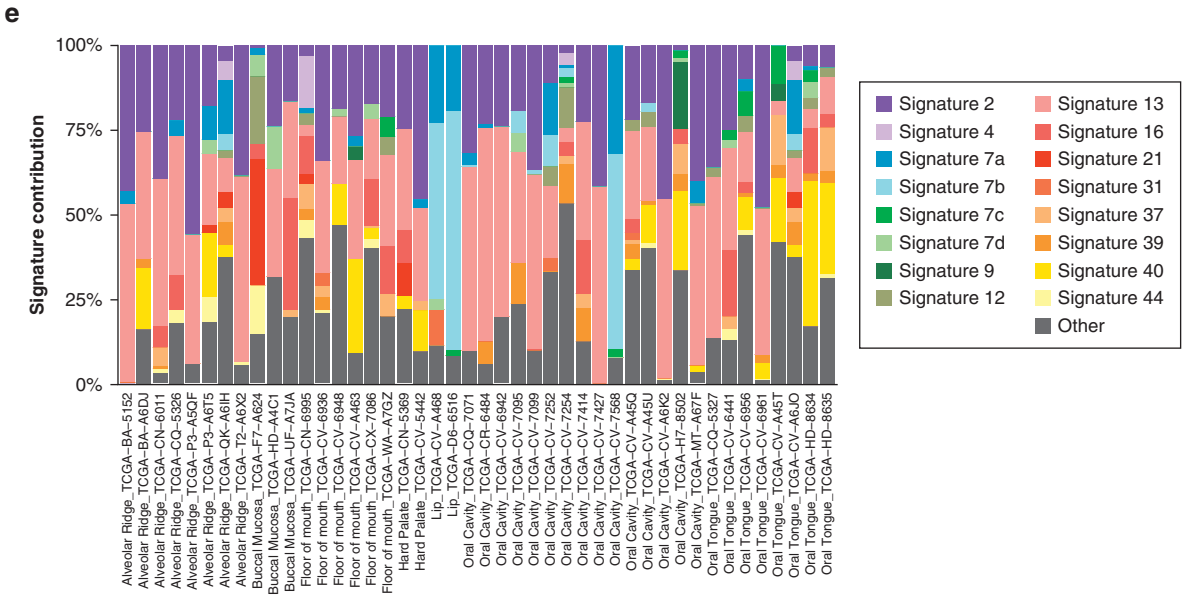


Figure 2. Continued

(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., aging 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 et al. (2014) compared to our WGS data (Figure 2c, 2d) is because the vast majority of mutations in cSCCs are located in the noncoding DNA (Figure 1a). Thus, WGS increases the power of signature analysis. Only filtered data are publicly available from Pickering et al. (2014), and filtering of variants may also have an impact on signatures. Furthermore, the reduced effectiveness of repair mechanisms in the noncoding 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 1.43–4.1%, Figure 1d).

UV-induced mutations at insulator sites

The mutational burden was 171-fold higher in the noncoding regions than in coding regions of the genome (Figures 1a, 1b). This concentration may reflect less effective mismatch repair in noncoding 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, the DNA binding site of which seems prone to UV-induced damage because of its high cytosine and dipyrimidine content. CTCF simultaneously binds to specific bp sequences at multiple DNA sites, thereby approximating distant chromatin regions and forming 3-dimensional DNA loops termed *topologically associated domains* (Hnisz et al., 2016; Kemp et al., 2014). These can incorporate multiple genes, the expression of which 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, 3c). This clustering of mutations at the CTCFbs was described in melanoma previously (Poulos et al., 2016), and also in colorectal cancer, which is not UV-associated (Katainen et al., 2015), and is thought to be due to bound CTCF blocking access of NER enzymes to the CTCFbs (Sabarinathan et al., 2016). Consistently, cSCC patients with dysfunctional NER due to deficiency of xeroderma pigmentosum group C (XPC) protein have comparable mutation density at flanking regions and CTCF motifs (Poulos et al.,

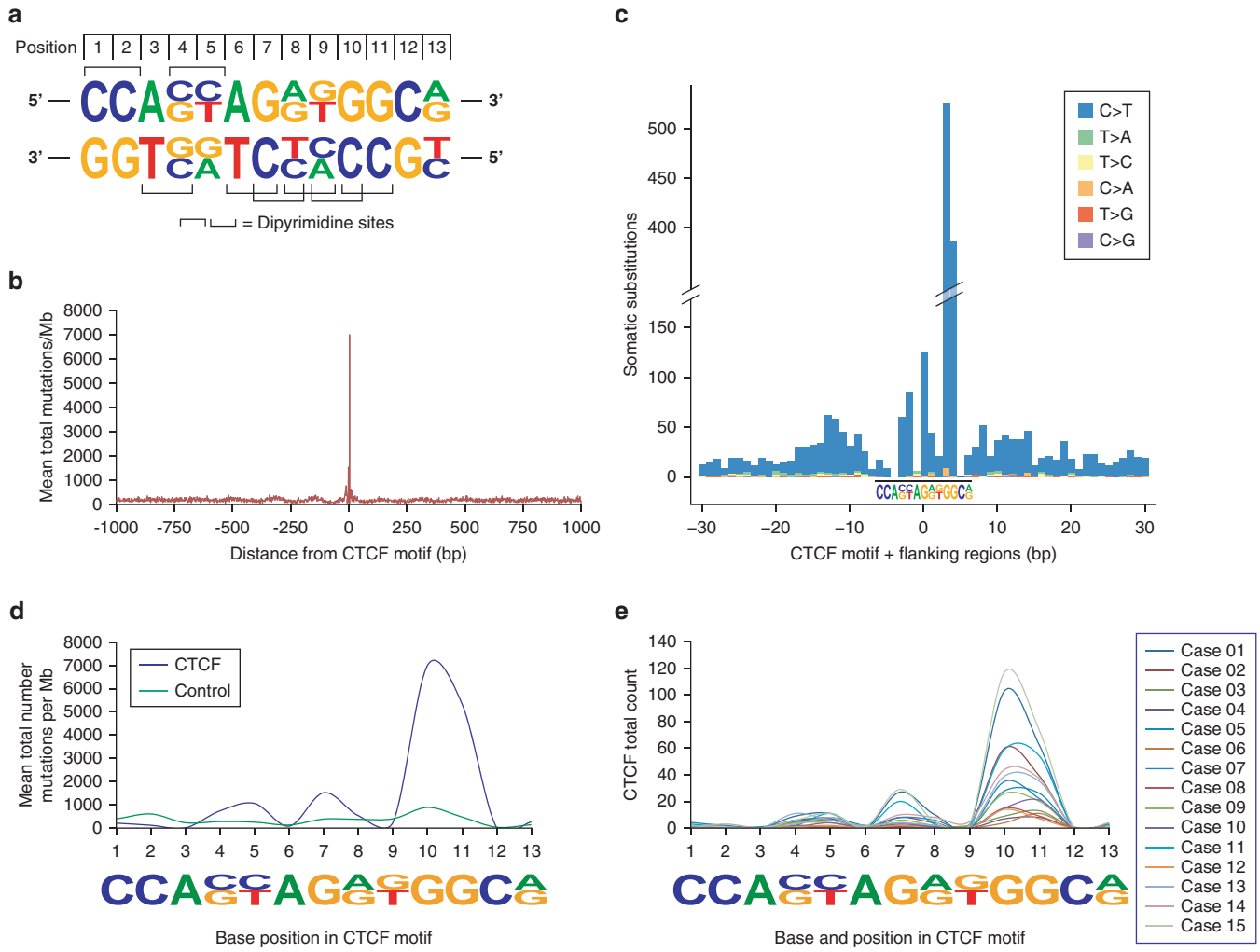


Figure 3. Mutational landscape at CTCF binding sites of 15 cutaneous squamous cell carcinoma lymph node metastases. (**a**) Upper- and lower-strand nucleic acid sequence showing the conserved center of the CTCF binding site. Large letters symbolize highly conserved positions. Potential dipyrimidine sites are indicated with brackets. (**b**) Overall density of mutations at merged CTCF binding sites and neighboring 1,000 base pairs. (**c**) Mutation density and mutation type in the CTCF binding site motif and flanking 30 base pairs. (**d**) Mutation density at individual base positions of the CTCF binding site derived from normal human epidermal keratinocyte chromatin immunoprecipitation sequencing peaks across the cohort showing concentration of mutations at positions 10 and 11 (blue line). The green line illustrates mutation density of a control motif, where only positions 8–13 match the CTCF motif (A|G)(G|T)GGC(A|G), and positions 1–7 can contain any nucleotide. (**e**) Mutation density at base positions in individual samples, demonstrating the consistency of the distribution of mutations across the cohort. Mb, megabase.

2016). Interestingly, the two cases in our cohort under treatment with immunosuppressants known to inhibit NER did not exhibit a more even distribution of mutations in and around the CTCFbs (Supplementary Figure S1 online). However, this could be explained by the fact that NER is only partially blocked 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. Poulos et al. (2016) were able to show that expression of cancer-associated genes in affected TAD loops was statistically different from wild-type TAD loops, but ultimately the assessment of the clinical impact of CTCFbs mutations on gene expression and carcinogenesis in cSCC will require additional functional analysis.

CONCLUSIONS

Mutation signature analysis has potential use in the clinic, as it allows distinguishing cSCC metastases from other non-cutaneous SCCs when a primary cannot be identified. This is

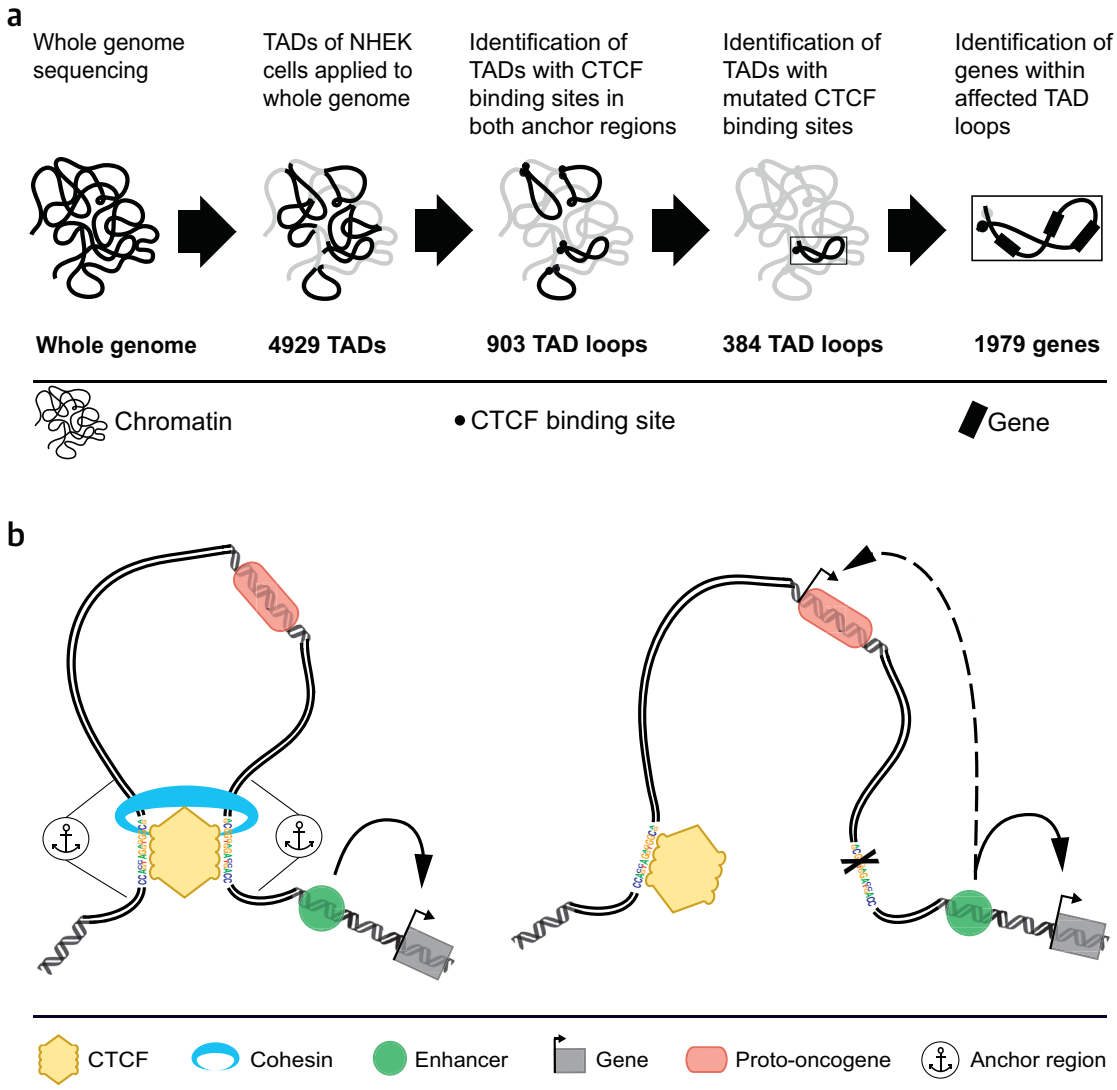


Figure 4. CTCF binding sites delimiting TADs. (a) Workflow for the identification of TADs with mutated CTCF binding sites. (**b**) Schematic illustration of a TAD loop regulated by CTCF. On the left, the TAD is insulated by CTCF bound to binding sites located in the anchor regions on both ends of the 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). NHEK, normal human epidermal keratinocyte; TAD, topologically associated domain.

an important finding, given the absence of uniquely characteristic driver gene mutations in cSCC. In addition to driver gene mutations in cSCC, gene regulation may be affected by UV-induced damage to DNA regulatory elements, notably CTCF insulator elements. Future research into drivers of cSCC carcinogenesis should consider the contribution of non-coding mutations and their impact on gene regulation.

MATERIALS AND METHODS

Sample processing and sequencing

Patients were prospectively enrolled between March 2015 and July 2017. Written 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 (ThermoFisher 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 ×30–45 for normal and ×65–90 for tumor samples. Mapping and variant calling was performed as per Tsoli et al. (2018) (Supplementary Methods online).

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 data sets from Perera et al. (2016), and the CTCF chromatin immunoprecipitation sequencing 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

Table 2. Genes located in topologically associated domain loops with mutated CTCF motifs across the cohort of 15 cutaneous squamous cell carcinoma 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*		

(Brunet et al., 2004) using Maftools (Mayakonda and Koeffler, 2016) in R, version 3.5.0 (R Development Core Team, Vienna, Austria). 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 using DeconstructSigs (Rosenthal et al., 2016). Signatures were reported when attributable mutations contributed >6% in any sample. The same approach was applied on publicly available variant call format files from whole exome sequencing data of cSCC (Pickering et al., 2014) and oral cavity SCC from The Cancer Genome Atlas (Lawrence et al., 2015).

CTCFbs 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 chromatin immunoprecipitation sequencing peaks for CTCF in NHEK available from ENCODE (The ENCODE Project Consortium, 2012), 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 1 kb. Results were normalized to mutations per megabase. We repeated the analysis using control motifs generated by Poulos et al. (2016) from NHEK chromatin immunoprecipitation sequencing peaks that only match positions 8−13 of the motif, while positions 1−7 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 chromatin immunoprecipitation sequencing data from ENCODE (The ENCODE Project Consortium, 2012). 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, because 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).

Data accession

The variant call format files have been deposited at the European Genome-Phenome Archive, which is hosted by the EMBL-European Bioinformatics Institute and the Center for Genomic Regulation, under accession number EGAS00001003370. The data set is under an 18 months embargo. Further information about European Genome-Phenome Archive can be found on: https://ega-archive.org and The European Genome-Phenome Archive of Human Data Consented for Biomedical Research (Lappalainen et al., 2015).

ORCIDs

Simon A. Mueller: http://orcid.org/0000-0001-9318-7270
Marie-Emilie A. Gauthier: http://orcid.org/0000-0002-5256-9165
Bruce Ashford: http://orcid.org/0000-0003-4547-6611
Ruta Gupta: http://orcid.org/0000-0003-2940-313X
Velimir Gayevskiy: http://orcid.org/0000-0002-2457-6261
Jonathan R. Clark: http://orcid.org/0000-0003-1209-772X
Marie Ranson: http://orcid.org/0000-0002-5570-9645
Mark J. Cowley: http://orcid.org/0000-0002-9519-5714

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

We extend our thanks to Elahe Minaei and Jay Perry of Illawarra Health and Medicine Research Institute, Wollongong, Australia, for sample processing, Ludmil Alexandrov, University of California San Diego, CA, for sharing the updated mutation signature matrices, and Narayanan Gopalakrishna Iyer, National Cancer Center Singapore, for conceptual support. MJC is supported

by a New South Wales Health Early-Mid Career Fellowship. We thank the Kinghorn Centre for Clinical Genomics for assistance with production and processing of whole genome sequencing data. This research was undertaken with support of the Illawarra Cancer Carers and the assistance of resources and services from the National Computational Infrastructure, which is supported by the Australian Government.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at https://doi.org/10.1016/j.jid.2019.01.008.

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.

Amin MB, Edge SB, American Joint Committee on Cancer. AJCC Cancer Staging Manual. Cham, Switzerland: 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.

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.

Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci USA 2004;101:4164–9.

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.

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.

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.

Dotto GP, Rustgi AK. Squamous cell cancers: a unified perspective on biology and genetics. Cancer Cell 2016;29:622–37.

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. Biochemistry 2001;40:2495–501.

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.

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.

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.

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.

Gurudutt VV, Genden EM. Cutaneous squamous cell carcinoma of the head and neck. J Skin Cancer 2011;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.

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.

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.

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.

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.

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.

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.

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.

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.

Lappalainen I, Almeida-King J, Kumanduri V, Senf A, Spalding JD, ur-Rehman S, et al. The European Genome-phenome Archive of human data consented for biomedical research. Nat Genet 2015;47:692–5.

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.

Liu Y, Sun J, Zhao M. ONGene: a literature-based database for human oncogenes. J Genet Genomics 2017;44:119–21.

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.

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.

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

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.

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

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.

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.

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.

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

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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.

SUPPLEMENTARY METHODS

Mapping and variant calling

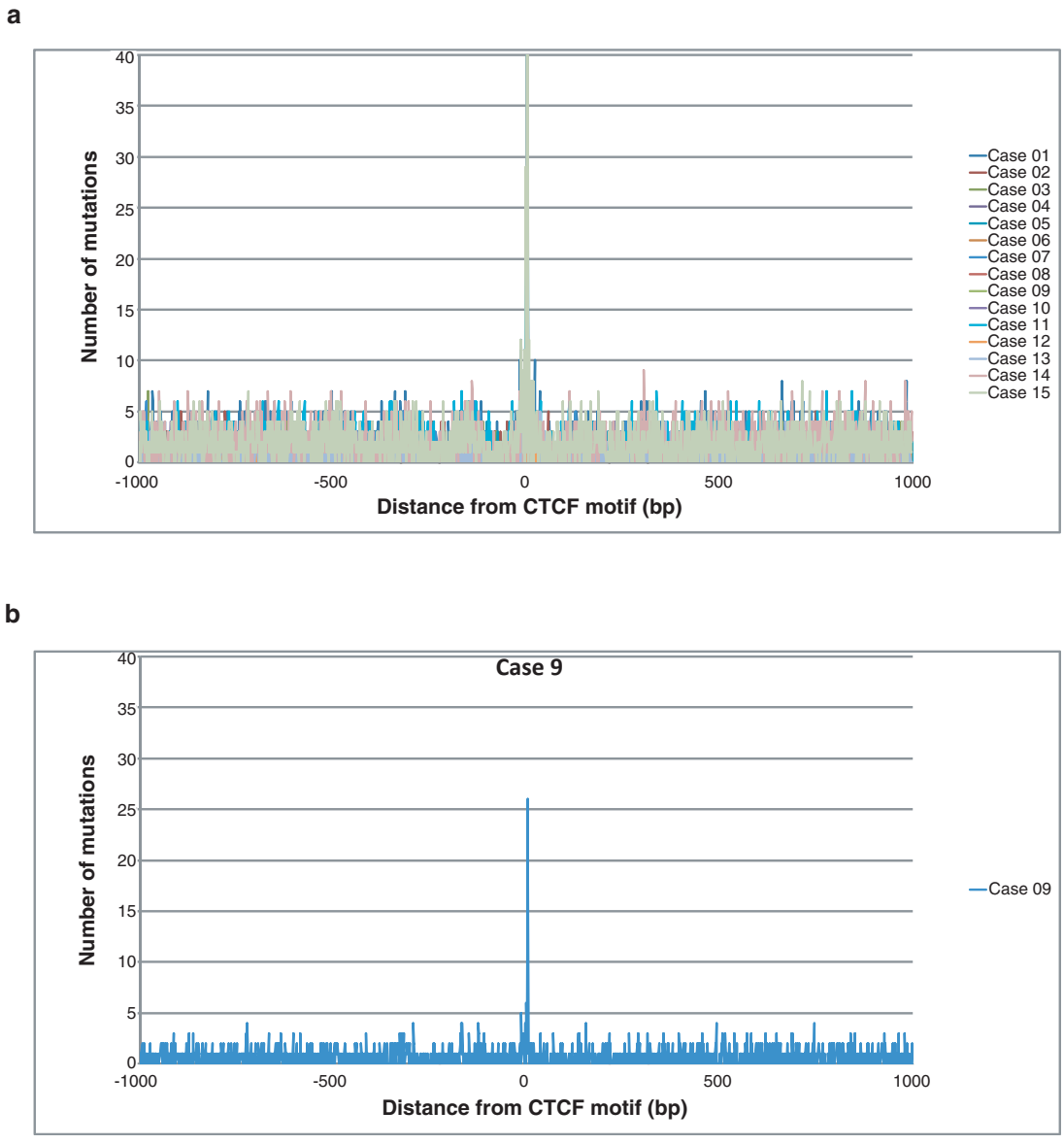
Sequencing reads were mapped to the human reference genome hs37d5 (b37 + decoy) using Burrows-Wheeler Aligner-MEM, version 0.7.10-r789 (Li et al., 2009). The resulting BAM files were sorted and duplicate reads were marked using Novosort, version 1.03.01 (Novocraft Technologies, Petaling Jaya, Malaysia). Because the tumor samples were run on multiple sequencing lanes, their BAM files were merged using Novosort Merge, version 1.03.01. Insertion and deletion realignment and base quality score recalibration were performed using GATK, version 3.3-0-g37228af (McKenna et al., 2010). The run quality was checked using Picard metrics (http://broadinstitute.github.io/picard.). Somatic single nucleotide variants, including small insertions and deletions, were called using Strelka, version 2.0.17.strelka1 (Saunders et al., 2012).

REFERENCES

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.

Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. Bioinformatics 2012;28:1811–7.



Supplementary 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.