TITLE

Whole genome analysis reveals the genomic complexity in metastatic cutaneous squamous cell carcinoma

AUTHORS AND AFFILIATIONS

Amarinder S. Thind[1,2], Bruce Ashford [1,2,3,9], Dario Strbenac[4], Ruta Gupta[5], Jonathan R Clark[4,6,7], N. Gopalakrishna Iyer[8], Jenny Mitchell[3] , Jenny Lee[9,10], Simon A Mueller [7,11], Elahe Minaei[2,12], Jay Perry[2,12] and Marie Ranson [2,12]

[1]School of Medicine, University of Wollongong, Wollongong, NSW, Australia
[2]Illawarra Health and Medical Research Institute, Wollongong, NSW, Australia
[3]Illawarra Shoalhaven Local Health District Wollongong, NSW, Australia,
[4]Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, NSW, Australia
[5]Anatomical Pathology, Royal Prince Alfred Hospital, Sydney, NSW, Australia.
[6]Royal Prince Alfred Institute of Academic Surgery, Sydney Local Health District, Sydney, NSW, Australia
[7]Department of Head and Neck Surgery, Chris O'Brien Lifehouse, Sydney, NSW, Australia
[8]National Cancer Center, Singapore
[9]Department of Medical Oncology, Chris O'Brien Lifehouse, Sydney, NSW, Australia
[10]Department of Clinical Medicine, Macquarie University, Sydney, NSW, Australia
[11]Department of Otorhinolaryngology, Head and Neck Surgery, Zurich University Hospital and University of Zurich, Zurich, Switzerland
[12]School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, NSW, Australia

CORRESPONDENCE

A/Prof Bruce Ashford
bruceash@uow.edu.au
372 Crown St, Wollongong, New South Wales, Australia 2500
+61 2 42266111

ABSTRACT

Metastatic cutaneous squamous cell carcinoma (cSCC) is associated with a high risk of recurrence and poor prognosis. There is limited published data exploring whole genome sequencing (WGS). The aim of this project was to provide the first comprehensive genomic understanding of the state of metastatic cSCC. In this study, we used WGS on matched tumor and blood DNA to detect somatic genetic alterations from 25 patients with regional metastases of head and neck cSCC. Our computational analyses interrogate clinical impacts of these genetic alterations on metastatic cSCC across the cohort for both the coding and non-coding genome.

In the non-coding genome, 3'UTR regions of *EVC* (48%), *PPP1R1A* (48%) and *LUM* (16%) were significantly functionally altered (Q-value < 0.05). Further, significant functional alterations are observed in the tumor suppressing lncRNA LINC01003 (68% of specimens, Q-value: 0.0158). In addition, significant recurrent copy number loss in tumor suppressor genes *KANSL1* and *PTPRD* and gain in *CALR, CCND1* and *FGF3* was observed for coding regions. SNV's driver analyses predicted *TP53, CDKN2A*, as potential drivers of the metastasis cSCC (using 3 different tools). Indel signature analysis highlight dominance of ID signature 13 followed by ID8 & ID9. Interestingly, ID 9 has previously been shown to have no association with skin melanoma, unlike ID 13 and 8, suggesting some point of difference between these two skin-based diseases. The overall landscape of variation in metastatic cSCC is dominated by cell cycle and DNA repair disruption.

70 **INTRODUCTION**

71 Cutaneous squamous cell carcinoma (cSCC) is the second most common malignancy, after basal
72 cell carcinoma (BCC), affecting up to 1 000 000 people in the United States (US) annually [1] .
73 Although primary cSCC is common, lymph node metastasis only occurs in 2-5% of cSCC [2, 3].
74 cSCC arising in the head and neck (HNcSCC) generally show a predictable pattern of spread,
75 predominantly metastasizing to the intraparotid, level II (upper jugular) and perifacial lymph nodes
76 [3]. cSCC that have metastasized to regional lymph nodes are associated with a worse prognosis,
77 with modest progress made in the management of regionally advanced disease over the last 15
78 years. Most patients with regional metastases from HNcSCC are managed with a multimodality
79 approach, which usually involves a surgery (parotidectomy and neck dissection) and adjuvant
80 external beam radiotherapy depending on the site and stage at the time of diagnosis [4]. More
81 recently immunotherapy has attracted great interest as a potential alternative for unresectable or
82 distant metastatic disease [5].Despite the very high incidence, relatively little is described or
83 understood about the progression from primary cSCC to metastasis.

84

85 We, and others, have reported on somatic variation seen in primary cSCC [6, 7] . Three previous
86 publications have included mutational analysis of metastatic cSCC [8-10] . These studies have
87 employed either targeted next-generation sequencing (against 321, 410 or 504 genes) or whole
88 exome sequencing to look principally at coding variants. The extent of non-coding (including
89 regulatory) regions of the genome was limited and varied in each study. Pickering et al [8] was the
90 only study employing whole exome sequencing and this study included only 7 metastatic samples.
91 There was no regulatory or non-exome regions analysed in this study. Both Li et. al [10] and Zehir
92 et. al [9] used targeted next generation sequencing, with limited non-coding analysis. Zehir et al

93 [9] specifically included the TERT promoter in their targeted panels but otherwise included no
94 regulatory elements. Li et al [10] similarly reported no regulatory of non-coding variant analysis.
95 No discussion in these reports addressed variation in either 5' or 3' untranslated regions (UTR).
96
97 We have previously analysed mutational burden in a cohort of 15 cSCC lymph node metastases
98 using whole genome sequencing (WGS) and reported the predominance of C→T transitions
99 associated with COSMIC UV mutational signatures 7a and 7b in metastatic cSCC, which were
100 particularly prevalent in non-coding DNA, and distinct from those seen in mucosal squamous cell
101 carcinoma [11]. The key findings of the study were the preservation of a UV associated signature
102 through the metastatic process and the concentration of variants within CCCTC-binding factor
103 regions. In the current report, we have used an expanded cohort of metastatic cSCC samples
104 processed by WGS for a detailed analysis of the variation in both coding and non-coding regions.
105 3'UTRs regulate crucial aspects of post-transcriptional gene regulation [12]. Mutations in these
106 regions can deregulate gene expression by disrupting miRNA-mRNA interactions, which in both
107 tumor suppressor genes and oncogenes can drive cancer progression [13, 14] . This variation in so
108 called *cis-elements* can also impact gene expression by altering translation initiation in cancer [15].
109 Despite the role of lncRNAs in cSCC not being fully understood, in general terms lncRNA are
110 thought to influence expression of proteins by pre and post translation influences on DNA/RNA
111 and proteins, chromatin function, miRNA activity and signalling pathways by an array of
112 mechanisms [16]. Their effect on miRNA might be augmented by soaking up, or sponging free
113 within lncRNA which impact this function can therefore have a potentiating effect on mRNA
114 translation hence biological effect [17].

115 The implications for functional consequences associated with both coding and these non-coding
116 genomic changes are also explored and combined to describe functional implications.
117

118 **MATERIALS AND METHODS**

119 This study had Institutional Human Research Ethics approval (UOW/ISLHD HREC14/397).
120 Thirty-two patients with resectable metastatic cSCC were identified by the treating surgeons
121 preoperatively. In addition to whole blood (for germline DNA), sections of fresh tumor from nodal
122 metastases were collected during surgery and immediately snap frozen. These sections were used
123 for both DNA/RNA extraction (Qiagen All Prep, Qiagen, Hilden, Germany) and for cellularity
124 estimates. Only samples with > 30% tumor (range 35-95%) proceeded to DNA quality control
125 (QC). QC comprised spectrophotography (Nanodrop 2000 ThermoFisher Scientific Inc), gel-
126 electrophoresis and SNP array. Twenty-five samples passed QC. The remaining seven samples
127 had insufficiently clonal tumor content (median variant reads ≤ 5 or median VAF < 0.1) or had an
128 extreme GC bias as determined by PURPLE [18]. Briefly, if more than 220 copy number segments
129 are unsupported by a corresponding structural variants at either end, the sample will be flagged as
130 Fail-Segment. The mean sequencing coverage of the 25 samples was 94.56X (range:64-143) for
131 tumor and 41.08X (range:30-56) for blood.
132

133 **Variant Calling and Functional significance of the SNVs and Indels**

134 FASTQ reads were aligned to reference genome GRChr38 using BWA-kit version 0.7.17 (BWA-
135 MEM read aligner) (more details refer to https://github.com/Sydney-Informatics-Hub/Fastq-to-
136 BAM). Further Genome Analysis Tool Kit (GATK) 4.1.2.0 and its BaseRecalibrator tool refined
137 the read alignment. Single nucleotide polymorphisms (SNPs) and insertion-deletion (indel)

138 variants were called by implementing GATK's Best Practices Workflow. These pipelines use
139 HaplotypeCaller for germline short variant discovery and Mutech2 caller for somatic short variant
140 discovery for SNVs and Indels (for details refer to https://github.com/Sydney-Informatics-
141 Hub/Somatic-ShortV). Furthermore, variants effect prediction and annotations completed using
142 OpenCravat platform [19]. MAF file was generated based with Variant Effect Predictor
143 annotations. We used three different methods for driver discovery including OncodriveFML [20]
144 , MutSigCV [21] and dNdScv [22] .
145
146 **OncodriveFML**
147 Functional significance of both coding and non-coding variants was predicted using the
148 OncodriveFML tool [20]. This tool is one of few that is designed for both the non-coding genome
149 and generally for WGS. It first determines the functional impact of the observed variants in
150 specified genomic region and later compares it with the functional impact of equal number of
151 random mutations from the same genomic region. The functional impact of the variants may be
152 computed in different manner in different genomic elements such as impact on protein function,
153 RNA structure, impact on microRNA Targets. Of particular interest within non-coding genome,
154 these regions were the 3' and 5' untranslated regions (UTR), promotor regions and long non-
155 coding RNA. Within 3'UTR regions of each of these genes, OncodriveFML predicted the SNVs
156 effect in terms of the binding and interactions of microRNA with mRNA. Others have sought to
157 analyse and interpret 5'UTR and promoters regions mutations in high mutational burden tumours
158 and reported limitations [23-25]. Increased mutational density at promoters can be linked to
159 decrease level of nucleotide excision repair (NER)[24, 25], high mutation rate at active
160 transcription factor binding sites (TFBS) due to E26 transformation-specific (ETS) TFs (which

161 induced cyclobutane pyrimidine dimers (CPDs) hotspots) [23].Rheinbay et. Al [26] reported
162 functional significance analysis issues for melanoma in their non-coding somatic driver study. We
163 encountered similar issues for promoters and 5' UTR regions in cSCC (Supplementary figure S.1).
164

165 **MutSigCV**

166 This tool identifies genes that are mutated more often than one would expect by chance. It reduces
167 the number of false positives in the generated list of significant genes, especially in tumour samples
168 with high mutation rates. Its model does this correction by incorporating various types of
169 information such as patient-specific mutation frequencies and mutation spectra, gene-specific
170 mutation rates, expression levels and replication times.

171 **dNdScv**

172 dNdScv is designed to test for positive and negative selection in cancer genomes. For the better
173 accuracy of the dNdScv model on UV originated cancer, we carefully monitored the annotation of
174 CC>TT changes (sometimes reported as C>T changes). Results report significance for missense,
175 truncating, Indel and then one global value. Global significance was considered for the
176 interpretation of outcomes.
177

178 **Copy Number Variation**

179 Using PURPLE [18], we derived copy number alterations on 25 metastatic specimens. PURPLE
180 estimates overall ploidy of the samples by using read depth ratio from COBALT and tumor B-
181 allele frequency (BAF) from AMBER. The pipeline is available at github of HMF Tools
182 (https://github.com/hartwigmedical/hmftools). Driver genes with significant amplifications and
183 deletions were then identified from PURPLE based driver gene identification outputs.

184

185 To explore the contribution of our analysis employing WGS to the understanding of the degree of
186 somatic variation in metastatic cSCC, we then compared short variants of the top 6 statistically
187 significant genes to all previously reported metastatic cSCC [8-10] using processed data available
188 on the cBioPortal [27, 28] to those within our cohort. Within cBioPortal, there are 63 samples of
189 nodal metastatic cSCC analyzed by either WES (n = 7) and targeted NGS (n = 56).

190

191 GRIDSS2 and its companion interpreter tool LINX were employed for somatic structural variant
192 analysis and Gene fusion [29]. COSMIC3 based, SNVs and Indels signatures from the whole
193 genome were built using deconstructSigs [30] and MutationalPatterns [31] software respectively.

194

195 RESULTS

196 Patient characteristics and clinicopathologic data

197 Twenty-five metastatic cSCC samples from lymph nodes in the head and neck region were
198 collected between 2015 and 2019. The median age of patients was 69 (range 30-87) and 24/25
199 (96%) were male. This sex disparity is in keeping with the disease burden we see in our practice,
200 particularly for advanced and metastatic cSCC (Ashford et al., manuscript in preparation). During
201 the period of patient recruitment for this study, 88% of patients with nodal metastasis of cSCC
202 were male. Further, due to histopathologic tumor percentage estimation and other quality control
203 measures, in this study only 1 female sample was sequenced. Others have highlighted that age,
204 male sex and immunosuppression are among risk factors for metastasis [32] while another study
205 reports sex determined immunological mechanisms as a determinant of the development of
206 metastasis in cSCC [33]. Two patients were immunocompromised; one patient was on long-term

207 azathioprine for rheumatoid arthritis and the other was on a combination of cyclophosphamide and
208 tacrolimus following solid organ transplantation.
209 The location of the index primary lesion was known in 11 patients ;lip (n = 2), ear (n = 2),
210 forehead/scalp (n = 4), cheek (n = 2) and nose (n = 1) . Nodal metastases were isolated from the
211 neck in 13 patients and in the parotid in 12 patients. The majority of patients had either moderately
212 differentiated (n = 8) or poorly differentiated (n = 12) cSCC, with evidence of extranodal extension
213 found in 20/25 (80%) nodal samples.
214
215 **Tumor mutational burden (TMB)**
216 The average TMB for SNV across the 25 cases was 238.7 mutations per megabase (range 32.52to
217 995.66 mutations/Mb) and for Indel it was indel/megabase 2.25 (range 0.63 to 5.9 mutations/Mb).
218 (Figure.1(F,G))(S.Table1) . The only female in our cohort had the second highest TMB at 499
219 mutations/Mb. here was no correlation between age, differentiation, nodal stage or extracapsular
220 spread of the metastasis and TMB. As we have previously reported, a mutation density across the
221 genome 171 times higher for non-coding variants compared to coding variants [11]. This was
222 corroborated in our expanded cohort of metastatic cSCC.
223

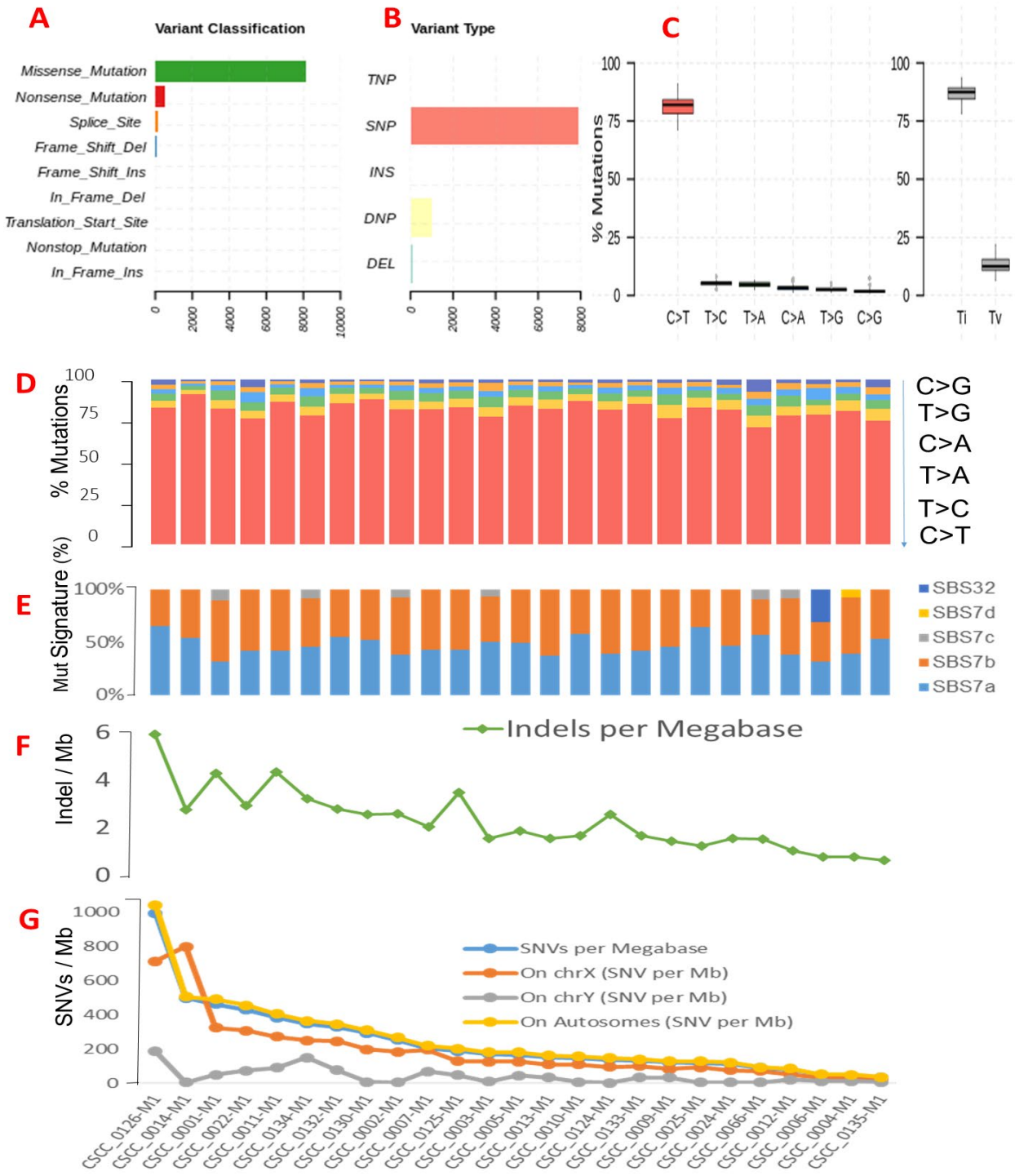
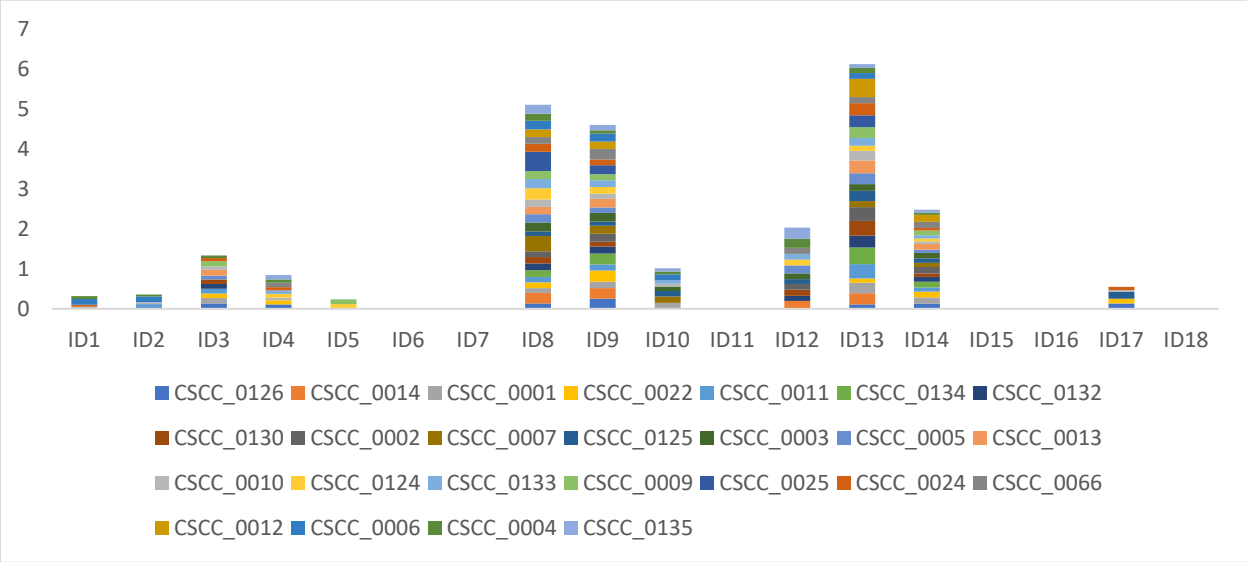


Figure1:- shows the overview of tumor mutational burden, signature and mutations types (A-D). Details of the coding mutations of the cSCC genomes are shown in figure 1(A to D), such as variants classification, variant types etc. Whereas Figure 1 (E, F, G) shows the whole genome-based Mutational Signatures, Indel and SNV Tumor mutational burden for each sample.

Mutational signatures

231 We performed mutational signature analyses based on COSMIC V.3.2
232 (https://cancer.sanger.ac.uk/signatures/). Signatures are designated as single base substitution
233 (SBS), double base substitution (DBS) or small insertion and deletion (ID) signatures. SNV
234 signature analysis of our 25 samples shows SBS signatures 7a and 7b are dominant (Figure 1E)
235 Substantial representation of SBS7c is also seen. SBS32 and SBS7d are observed in one sample.
236 Indel signature analysis of our samples showed that ID8, 9 and 13 dominate over others (Figure 2)
237 (S.Table2).
238



239

240 **Figure 2:** Indel signatures based on cosmic3 database. In the figure X-axis represent the sum of score for signature
241 among 25 samples. Y axis shows the indel signatures. In our cohort ID13, ID8 and ID9 predominate over other Indel
242 signatures. Full details are available in supplementary (S.Table2(Tab2-4)).
243

244 **Short variants**

245 **Variation in non-coding regulatory regions**

246 Variants in the 3' UTRs of key cancer associated genes with high functional impact on miRNA-
247 mRNA interactions were evident in our cohort of metastatic cSCC. Among our chort of samples,
248 within the 3'UTR region of *EVC, LUM* and *PPP1R1A*, non-coding SNVs showed significant

249 higher observed functional impact than the predicted functional impact (refer to method section)
250 (Figure 3b). We observed variation within the 3'UTR of *EVC* in 48% of samples with a high
251 predicted functional impact. Similarly, high impact mutations in the 3'UTR of *PPP1R1A* were
252 observed in the 48% of samples. We also observed one unique variant i.e *PPP1R1A*
253 Chr12:54579896 (G to A) variant with cDNA change of c.*491C>T in five samples (S.Table3)

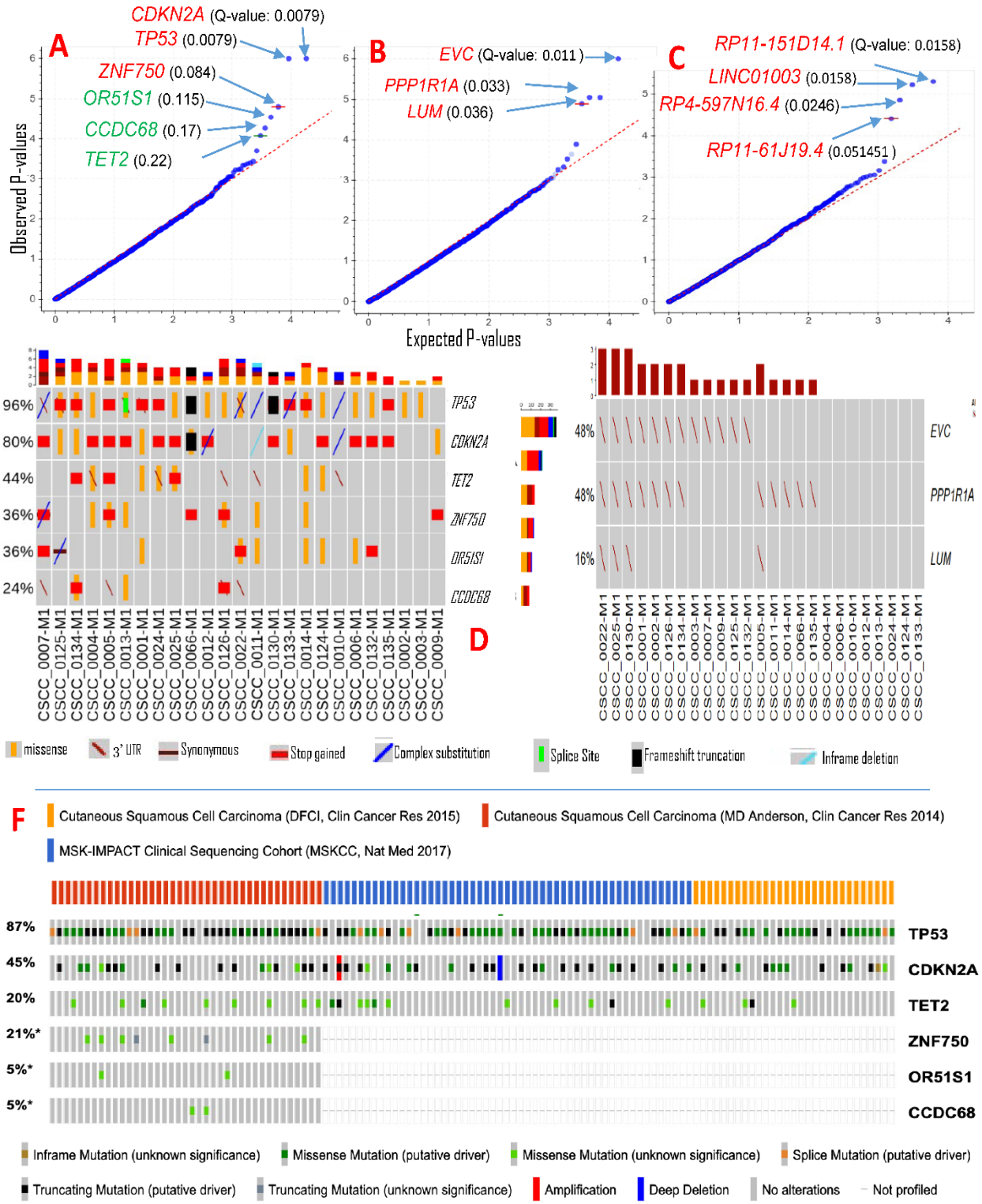


Figure 3 shows various coding and non-coding genes or regions having significant alterations in cSCC metastasis. (a –c) Plots showing the most significantly altered (in the plots above the red line Q-values are below 0.1 and between the green line & the red line Q-values are 0.25 to 0.1.) Q-values are corrected P-values using the Benjamini/Hochberg

correction (Where P-values are defined by times that the observed value is >= the expected value, divided by the number of randomizations) (a) coding region (genes), (b) 3' UTR regions of the genes, and (c) lncRNA regions in the cohort of 25 patient samples. Plots showing the frequency of observed mutations with respect to the expected frequency of the mutations in the corresponding regions. These analyses were done using OncodriveFML (2.2.0). (d) Detailed sample-level information of the SNVs along with types of variants in the top altered genes (refer to panel a). (e) Plots showing variants of significantly altered 3'UTR regions. (f). Oncplot retrieved from cBioportal to compare the variants from panel (d) with other metastatic cSCC studies, which are either whole exome or targeted panel based. Due to the use of targeted panel in some CSCC studies, not all genes from panel (d) were measured.

lncRNA were analysed to determine a potential impact on cancer associated mRNA. An example of this is in lncRNA LINC01003, which is amongst the frequently mutated lncRNA in our cohort (71% of samples). Other recurrently varied lncRNA in our cohort were RP11-151D14.1, RP4-597N16.4 and RP11-61J19.4 (Figure 3c).

Coding Short Variants

The overwhelming majority of coding SNVs were missense mutations, followed by nonsense mutation, which represented less than 5% of variants. Over 80% of SNV mutations were C>T, consistent with the dominant effect of UV radiation on pyrimidine bases. This is in keeping with our recently published findings on mutational signatures according to COSMIC V.2 (Mueller et al., 2019). Genes most frequently subject to high impact SNV included *TP53, CDKN2A and ZNF730* having Q-value <0.1, whereas Q-value of *OR51S1, CCDC68* and *TET2* is between 0.1 – 0.25 (Figure 3a).

The only sample to show no significant mutation in *TP53* was CSCC_0009. The TMB of this sample was 122/Mb, or 51% of the average across the cohort. The five samples without *CDKN2A* mutations average a TMB of 470/Mb, or 201% of the average for the cohort. In nine cases, high impact short variants of *ZNF730* was observed.

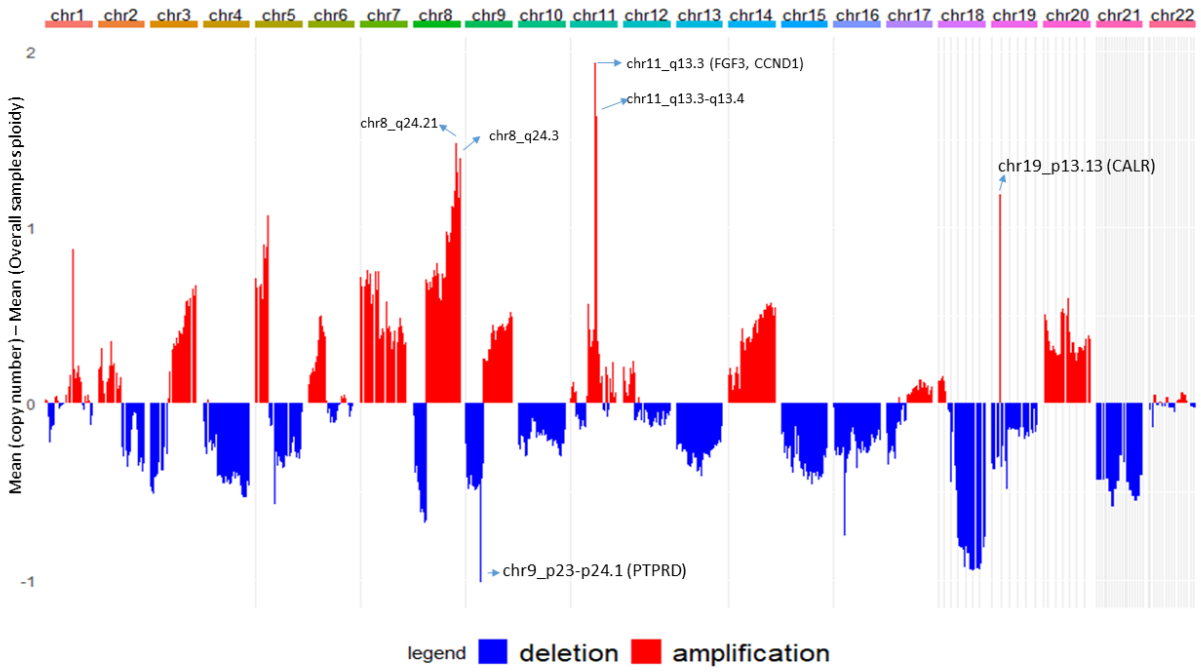
286 **Structural variation and gene fusions**

287 Structural variation analysis revealed substantial chromosomal disruption in the metastases
288 characterized by both translocation and inversion and often accompanied by allele specific copy
289 number variation. Minor allele loss, large scale deletions and amplifications were observed .
290 Despite this widespread genomic instability, only a single coding-coding gene fusion was observed
291 in our cohort – this was between *STRN* and *DLG2* in sample CSCC_0009 (*STRN*: Exon 1
292 ENST00000263918 *DLG2* Exon 7 ENST00000376104). *STRN* encodes a calcium-dependent
293 calmodulin-binding protein [34]. *DLG2* plays a role in pain signalling and deletion is seen in both
294 human and canine osteosarcoma [35]. CSCC_0009 came from a patient who had undergone liver
295 transplantation and was on immunosuppressive therapy. The primary tumor that gave rise to this
296 metastasis showed perineural involvement, but this was present in the metastatic deposit.

297

298 **Copy Number Variation**

299 The extent of chromosomal copy number gain and loss was averaged across the genome for all 25
300 tumor samples (Figure 4). Chr5p and 8q were the most frequently amplified regions, with 18q
301 being the region with the most recurrent deletion. Previously reported [36] recurrent 9p loss was
302 also seen.



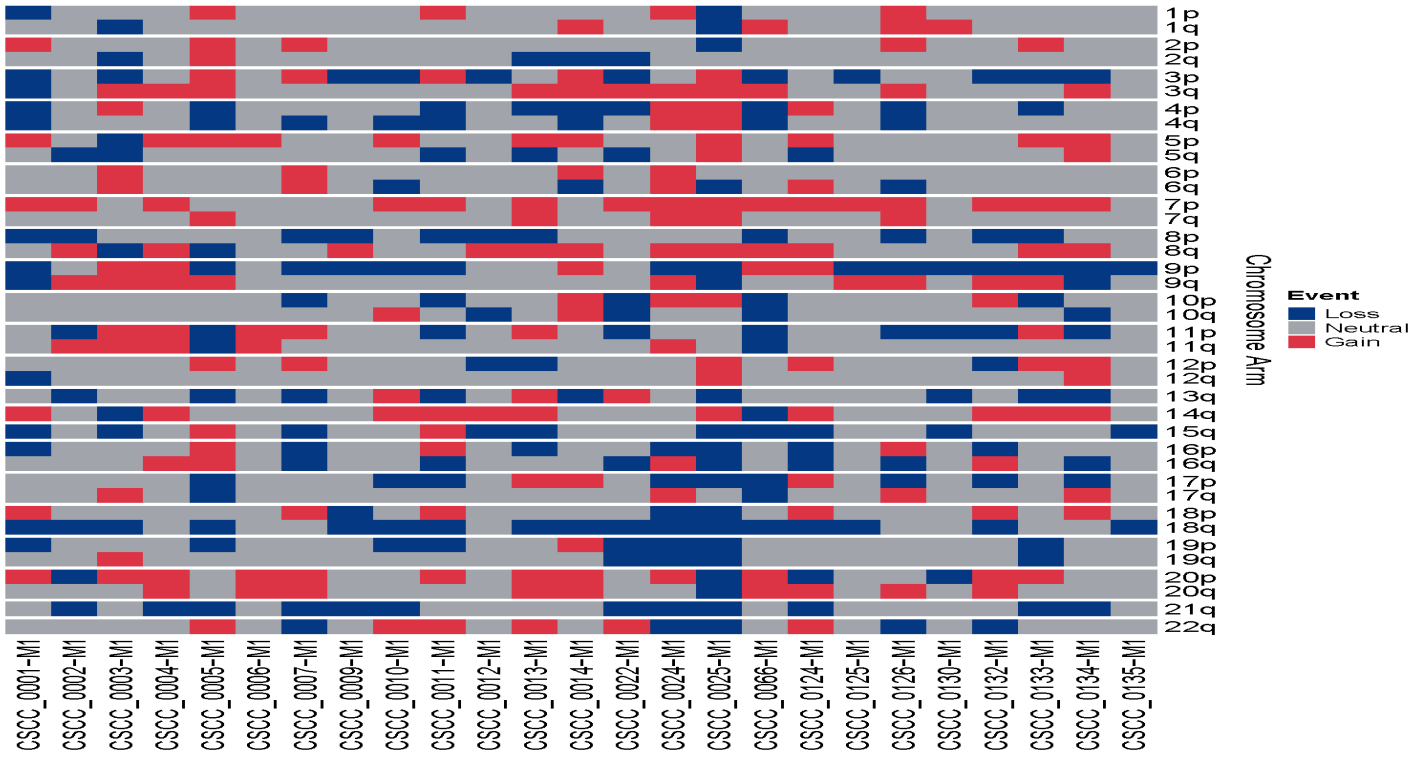


Figure 4. Chromosomal and recurrent genetic copy number variation in metastatic cSCC (a) Combined chromosomal CNV across 25 metastatic cSCC samples at chromosomal level. X-axis represents the differences of mean minimum copy number (bands) and Mean of overall samples ploidy (after adjustment for purity). For details please refer to S.Table4 (Chromosome_plot_raw_Mean_difference_file). (b) Figure show chromosome Arm loss and Gain at sample level, where red color is for Gain and Blue color is for loss. Where both arm of Chromosome 7 and 5p is gained, and 8p, 18q and 21q shows loss. (A chromosome arm is defined to be deleted if at least half of its bases are one or more

copies less than the sample ploidy. A chromosome arm is defined to be amplified if at least half of its bases are one or more copies more than the sample ploidy.)

The recurrent amplified and deleted genes across all 25 tumor samples are presented in Figure 2b. The most frequently deleted genes were *KANSL1* and *PTPRD.* Recurrent amplification was seen less commonly.

KANSL1 (Ch17q) showed deletion in 32% of samples. Deletion of *PTPRD* (Chr9p) was seen in 20% of samples. Deletion of *PTPRD* (n=6) and *CDKN2A* (Chr9p) (n=1) did not co-occur in our cohort (Figure 5), although *PTPRD* loss and significant mutation of *CDKN2A* occurred together in five samples (CSCC_9, 11, 133, 132 and 134).

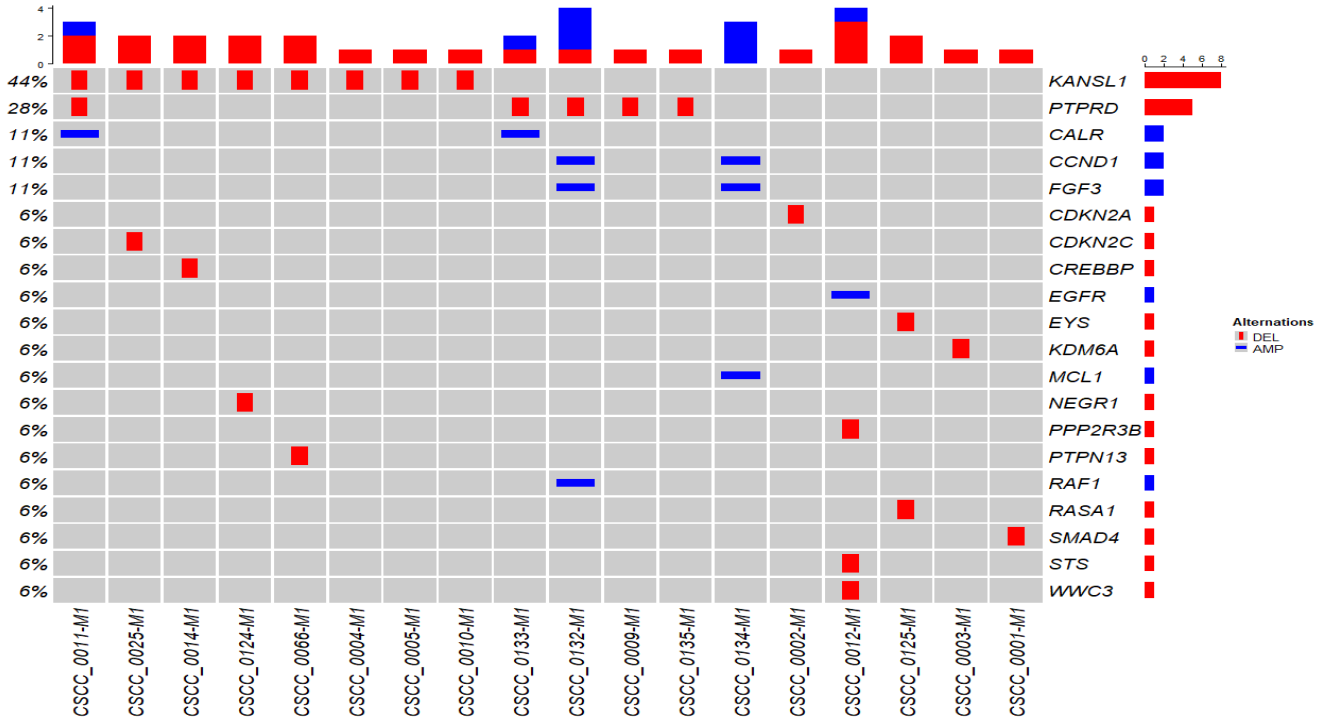


Figure 5:- Recurrent copy number gain and loss for individual genes, Deletion CN<0.5 and Amplification, minimum exonic copy number > 3 x sample ploidy >4 adjusted to cellularity and ploidy. These driver genes were derived by the use of purple-3.1. For details please refer to S.Table5.

Amplification was seen most commonly in *CALR*, *CCND1* and *FGF3* (Figure 5). Amplification of *CCDN1* and *FGF3* co-occurred in two samples (CSCC_0134 and CSCC_0132). These two

328 cases had extensive nodal involvement being the only samples with >50% of lymph nodes
329 involved by tumor.
330
331 **DISCUSSION**
332 This is the largest study to employ WGS to assess the mutational landscape of metastatic cSCC
333 and provides unique insights into the expansive array of variation. The analysis presented here
334 demonstrates the breadth of somatic variation in metastatic cSCC across non-coding and coding
335 regions, including the identification of novel Indel signature patterns, and highlights for the first
336 time the nature and depth of this variation within regulatory regions. We have previously published
337 an analysis of WGS of a smaller cohort of metastatic cSCC in which we focussed on the
338 preservation of UV associated COSMIC signature 7b and the density of mutations found in the
339 CTCF binding regions of the genome. Herein we have doubled the samples analysed and adopted
340 a more comprehensive survey of the extent of variation observed using WGS. The utility of WGS
341 has enabled an understanding of the extent of variation beyond coding regions and has defined the
342 prominence of non-coding variation, with special attention devoted to 3'UTR, lncRNA.
343 Additionally, WGS analysis has allowed us to compare the extent of variation to that previously
344 reported in metastatic cSCC, previously studies relying on WES and targeted NGS, and to update
345 and expand the understanding of mutational signature patterns in cSCC.
346
347 The rate of TMB within metastatic cSCC is greater than any tumour previously described. At 238
348 mutations/Mb, this is substantially higher than other cancers known to have high mutational
349 burden, including melanoma (49 mutations/Mb) [37] and NSCLC (10 mutations/Mb) [36]. This
350 finding is in keeping with Pickering et al [8] who found a TMB of 61.2 mutation/Mb of four times

351 that of melanoma in their high risk cSCC. Their series included just seven metastatic samples.
352 According to the formula established by Yarchoan[38] for assessing predicted response rate to PD-
353 1 inhibition, response rate = $10.8 \times \log_e(X)-0.7$, where X is the mutational burden (mutations/Mb).
354 The calculated potential response rate to PD-1 inhibition in our cohort was 39.6%, which is
355 consistent with current literature [5]. Immune checkpoint inhibitors have recently demonstrated
356 high, durable response rates and superior tolerability when compared to platinum chemotherapy
357 in unresectable locally advanced and metastatic cSCC [39]. The high TMB was associated with
358 substantial structural variation, without recurrent gene fusions.
359 As we have previously reported in a smaller cohort, the metastatic process preserves the dominant
360 UV associated variant pattern of C>T, or CC>TT in cSCC. This finding is independent of the
361 degree of differentiation or any other clinicopathologic feature.
362
363 Alexandrov et al detailed patterns of mutational signatures in 23829 tumour samples (1965 WGS)
364 from the Pan Cancer Analysis of Whole Genomes (PCAWG) datasets defining single base
365 substitution (SBS), double base substitution (DBS) and small insertion and deletion signatures (ID)
366 (Alexandrov, Kim et al., 2021). No cutaneous SCC (primary or metastatic) are included in this
367 dataset. SBS7a/SBS7b/SBS7c/SBS7d were all seen in skin melanoma in Alexandrov et al. In our
368 metastatic cSCC analysis, SBS7a and b were the most dominant signatures, with less presence of
369 SBS 7c and d. Alexandrov et al identified 17 ID mutational signatures. COSMIC version 3.2
370 (https://cancer.sanger.ac.uk) has expanded the ID signatures to 18 [40] .
371
372 Indel signature analysis on our 25 metastatic cSCC samples identified the predominance of ID
373 signatures 8, 9 and 13 (100% of samples effected). ID 8 is thought to be both related to double

374 strand DNA break repair dysfunction and to age related changes. The mechanism of ID 9 is unclear.
375 In their analysis, Alexandrov et al found ID 9 in only 2/104 melanoma samples. Thus, our finding
376 represents a departure from the prevailing dominance of ID 9 in mainly soft tissue sarcoma within
377 the PCAWG data. Alexandrov described only one cancer type with ID 13 as predominant, this
378 being skin melanoma. Our analysis revealed 100% of samples with ID 13 in metastatic cSCC. Our
379 data also provides evidence of concomitance of ID 13 with SBS 7a and 7b in keeping with a UV
380 association.
381
382 We have demonstrated that both Indel and somatic variant signatures are different for metastatic
383 cSCC than for other malignancies of the head and neck. The potential for both Indel and somatic
384 variation signatures to define the likely aetiology of nodal metastatic disease from an "unknown
385 primary SCC" would allow for discrimination between skin and mucosal origin. This would have
386 tremendous treatment impacts and would dictate the benefits of surgery, radiotherapy and systemic
387 (both chemotherapy and immunotherapy) treatments. Such analysis could be completed on
388 smaller genomic regions thus allowing for a cheaper and quicker alternative to WGS or WES.
389
390 **Copy Number Variation**
391 Loss of *KANSL1* and *PTPRD* were the most prominent CN alterations seen within our 25 samples,
392 with amplification most common in *CALR*, *CCND1* and *FGF3*. *KANSL1* is a chromatin modifier
393 gene located on the long arm of chr17 with deletion observed in 32% of samples. Germline loss
394 of this gene is associated with 17q21.31 microdeletion syndrome [41]. In contrast, epigenetic
395 analysis of epithelial ovarian cancer identified amplification of *KANSL1* as a means for potential
396 targeted therapy [42].

397 Deletion of *PTPRD* was seen in 20% of our samples. Lambert et al. [43] described deletions of
398 *PTPRD* in 37% of either primary cSCC that had metastasised or metastatic disease. In addition,
399 some of their cases also displayed a variant in the minor allele concordant with the deletion leading
400 to a loss of heterozygosity event. *PTPRD* encodes protein tyrosine phosphatase receptor D, which
401 belongs to a family of receptors whose action oppose that of the tyrosine kinases, which are central
402 to cell growth and differentiation and oncogenic transformation (genecards.org). Large scale
403 genomic events impacting *CDKN2A* can also affect *PTPRD* due to their proximity on chr9 [44].
404 *PTPRD* dephosphorylates S*TAT3*, deactivating its tumorigenic activity. In head and neck SCC,
405 *PTPRD* inactivation significantly increased *STAT3* hyperactivation, which was associated with
406 decreased survival and resistance to EGFR-targeted therapy [45]. *PTPRD* has been implicated as
407 a tumor suppressor in several cancers with inactivating somatic variants found in >50% of GBM
408 and between 10-20% of head and neck mucosal SCC (HNSCC)[46]. It is possible that *PTPRD*
409 plays a tumor suppressor role in metastatic cSCC.
410
411 Amplification was seen most commonly in *CALR*, *CCND1* and *FGF3* (Figure 5). *CALR* encodes
412 a ubiquitous endoplasmic reticulum bound calcium receptor [47]. Cellular stress can move *CALR*
413 fragments to the plasma membrane from the ER and influence immune recognition of cancer cells.
414 Recent analysis of *CALR* fragments in myeloproliferative disease suggest an immunosuppressive
415 influence of extracellular *CALR* [48]. Cyclin D1 (*CCND1*) amplification is associated with nodal
416 metastasis and worse survival in oral SCC [49]. In a review of *CCND1* copy number variation in
417 metastatic non-cutaneous melanoma, amplification of *CCDN1* was prominent in those patients
418 whose disease did not respond to immune checkpoint inhibition [50]. *FGF3* amplification is more
419 common in metastatic breast cancer than primary tumors [51]. Targetable *FGF3* amplification was

420 associated with a poorer prognosis and lung metastasis in hepatocellular carcinoma [52]. This
421 amplification was seen in only 2% of total HCC but was most common in those cancers showing
422 rapid response to sorafenib.
423
424 **Non coding short variants**
425 We identified substantial somatic variation within the 3'UTR region of *EVC, LUM* and *PPP1R1A.*
426 *EVC* effects ciliary Hedgehog (Hh) regulation. Aberrant overexpression of *EVC* (and upregulation
427 of Hh) has been reported in adult T-cell leukaemia as a result of epigenetic modulation [53]. Within
428 breast cancer, the expression of *EVC* was found to be reduced in nodal deposits of metastatic breast
429 cancer compared with primary breast cancer suggesting a role in the metastatic process [54].
430 *PPP1R1A* is a protein phosphatase inhibitor which appears to have a variable but significant role
431 in the metastatic process. In Ewing Sarcoma, it is overexpressed and has been proposed as a driver
432 of metastasis [55]. Conversely, levels of *PPP1R1A* were reduced in breast cancer when compared
433 to adjacent non-diseased breast tissue [56]. Within our cohort, we observed a recurrent missense
434 mutation in the 3'UTR of *PPP1R1A* in five samples.
435 LINC01003 was the most mutated lncRNA in our cohort (71% of samples). In multiple
436 myeloma, LINC01003 behaves as a tumor suppressor genomic element. Up-regulation
437 suppresses multiple myeloma by repressing cell viability and adhesion and promoting apoptosis.
438 This effect is via its sponge effect on miR-33a-5p and its target *PIM1 [57]*.
439
440 **Coding short variants**

441 The genes with most recurrent variation were *TP53, CDKN2A.* Of note was the absence of
442 significant or recurrent SNV affecting NOTCH1/2. Inman et al (2018) compared well

443 differentiated to moderately and poorly differentiated primary cSCC and identified *NOTCH1*,
444 *NOTCH2, TP53* and *CDKN2A* as the most commonly mutated genes, with *ATP1A1, HERC6,*
445 *MAPK1P1L, GRHL2, TRAPPC9, FLNB* and *MAP3K9* identified as common early events in
446 primary cSCC (Inman et al., 2018). Within this group, *GRHL2* was associated with less well
447 differentiated tumours including those with a worse prognosis, and the authors suggested it's role
448 needed further assessment in cSCC. We found a single example of splice variant in our 25 samples
449 in *GRHL2,* suggesting its role in metastatic disease is limited.

450 Loss of function mutations within *TP53* and *CDKN2A* adversely impact cell cycle pathway control
451 and DNA repair mechanisms and mutations in these genes has been frequently reported for cSCC
452 [58]. Kilnakis et al describe a pattern of *TP53* mutation that differed between primary and
453 metastatic disease in head and neck (mucosal) SCC [59]. They found an overall lower rate of
454 mutations in metastatic tumors, but a higher concentration of missense mutations in the DNA
455 binding regions of the gene. Additionally, they found a higher TMB with *TP53* mutation, but a
456 worse response to immunotherapy. Burtness et al [60] reported that the extent of TMB in HPV-
457 negative HNSCC is associated with loss of function mutations in both *TP53* and *CDKN2A* .
458

459 Loss of function mutations in *ZNF730* are associated with poor prognosis in cervix, lung and
460 HNSCC and its tumor suppressive role is mediated via *TINCR* and *LAMC2* to effect cell
461 differentiation, migration and death [61]. Its role in the metastatic phenotype of cSCC has not been
462 previously reported and herein we showed a high level of deleterious single nucleotide variation
463 with nine samples showing either a missense or a stop gained mutation. *ZNF730* showed no
464 mutation in the only study that included its analysis [8].

465 *OR51S1* encodes a gene that regulates olfactory receptor function via a G protein-coupled receptor
466 pathway (genecards.org). It could be false positive as reported by Lawrence et al. [21] olfactory
467 receptor genes are long protein coding genes. In our case the Q-value is 0.115 which is not too
468 low.

469 High impact short variants of TET2 were seen in 44% of sample. Deletion of *TET2* is understood
470 to play a role in haemopoietic malignancy [62]. More recently, loss of *TET2* activity was
471 associated with reduced tumor chemokine production and PD-L1 expression via IFN-
472 γ/JAK/STAT signalling pathway, conferring an increased resistance to immunotherapy in
473 melanoma cells [63]. Previous studies [8-10] have noted this variant in metastatic cSCC at a rate
474 of 20% (Figure 3f). Further interrogation of the TCGA dataset for metastatic cSCC revealed 63
475 cases [8-10] with short variant analysis (Figure 3f). Of these only seven cases had undergone
476 WES with the remainder having undergone targeted NGS. Not all genes identified as high impact
477 and recurrent variants in our cohort were included in these studies, which underscores the
478 deficiencies of targeted analyses in discovery projects.

479

480 **CONCLUSION**

481 WGS provides insight into the unparalleled burden of mutation within metastatic cSCC. In this
482 study we describe a complex interaction of inactivation of putative tumor suppressor genes and
483 amplification of oncogenes with biologically significant regulatory region variation. The
484 combination of checkpoint dysregulation and the high mutation burden underline the response to
485 immunotherapy in metastatic cSCC. The examination of Indel and somatic variant signatures
486 provide a path to more accurate determination of the origin of metastatic SCC. This analysis of

genomic variation provides an opportunity for a deeper understanding of the metastatic cSCC landscape to be the foundation for the identification of biomarkers in primary cSCC predicting metastasis.

ACKNOWLEDGMENTS

This work was funded by the Illawarra Cancer Carers, The Head and Neck Research Fund, Royal Prince Alfred Institute of Academic Surgery, The Cancer Institute NSW translational program grant, Chris O'Brien Lifehouse, and National Health and Medical Research Council Project Grant APP1181179.

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest

1. Rogers, H.W., et al., *Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012.* JAMA dermatology, 2015. **151**(10): p. 1081-1086.
2. Venables, Z.C., et al., *Nationwide incidence of metastatic cutaneous squamous cell carcinoma in England.* JAMA dermatology, 2019. **155**(3): p. 298-306.
3. Forest, V.I., et al., *N1S3: a revised staging system for head and neck cutaneous squamous cell carcinoma with lymph node metastases: results of 2 Australian Cancer Centers.* Cancer: Interdisciplinary International Journal of the American Cancer Society, 2010. **116**(5): p. 1298-1304.
4. Veness, M., *Treatment recommendations in patients diagnosed with high-risk cutaneous squamous cell carcinoma.* Australasian Radiology, 2005. **49**(5): p. 365-376.
5. Migden, M.R., et al., *PD-1 blockade with cemiplimab in advanced cutaneous squamous-cell carcinoma.* New England Journal of Medicine, 2018. **379**(4): p. 341-351.
6. Zilberg, C., et al., *Analysis of clinically relevant somatic mutations in high-risk head and neck cutaneous squamous cell carcinoma.* Modern Pathology, 2018. **31**(2): p. 275-287.
7. Inman, G.J., et al., *The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature.* Nature communications, 2018. **9**(1): p. 1-14.
8. Pickering, C.R., et al., *Mutational landscape of aggressive cutaneous squamous cell carcinoma.* Clinical cancer research, 2014. **20**(24): p. 6582-6592.
9. Zehir, A., et al., *Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients.* Nature medicine, 2017. **23**(6): p. 703-713.
10. Li, Y.Y., et al., *Genomic analysis of metastatic cutaneous squamous cell carcinoma.* Clin Cancer Res, 2015. **21**(6): p. 1447-56.
11. Mueller, S.A., et al., *Mutational patterns in metastatic cutaneous squamous cell carcinoma.* Journal of investigative dermatology, 2019. **139**(7): p. 1449-1458. e1.
12. Barrett, L.W., S. Fletcher, and S.D. Wilton, *Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements.* Cellular and molecular life sciences, 2012. **69**(21): p. 3613-3634.
13. Liu, W. and X. Wang, *Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data.* Genome biology, 2019. **20**(1): p. 1-10.
14. Schuster, S.L. and A.C. Hsieh, *The untranslated regions of mRNAs in cancer.* Trends in cancer, 2019. **5**(4): p. 245-262.
15. Kobayashi, H. and Y. Tomari, *RISC assembly: coordination between small RNAs and Argonaute proteins.* Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2016. **1859**(1): p. 71-81.
16. Statello, L., et al., *Author Correction: Gene regulation by long non-coding RNAs and its biological functions.* Nat Rev Mol Cell Biol, 2021. **22**(2): p. 159.
17. Cesana, M., et al., *A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA.* Cell, 2011. **147**(2): p. 358-69.
18. Priestley, P., et al., *Pan-cancer whole-genome analyses of metastatic solid tumours.* Nature, 2019. **575**(7781): p. 210-216.

19. Pagel, K.A., et al., *Integrated Informatics Analysis of Cancer-Related Variants.* JCO Clin Cancer Inform, 2020. **4**: p. 310-317.
20. Mularoni, L., et al., *OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations.* Genome Biol, 2016. **17**(1): p. 128.
21. Lawrence, M.S., et al., *Mutational heterogeneity in cancer and the search for new cancer-associated genes.* Nature, 2013. **499**(7457): p. 214-218.
22. Martincorena, I., et al., *Universal Patterns of Selection in Cancer and Somatic Tissues.* Cell, 2017. **171**(5): p. 1029-1041.e21.
23. Mao, P., et al., *ETS transcription factors induce a unique UV damage signature that drives recurrent mutagenesis in melanoma.* Nat Commun, 2018. **9**(1): p. 2626.
24. Perera, D., et al., *Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes.* Nature, 2016. **532**(7598): p. 259-63.
25. Sabarinathan, R., et al., *Nucleotide excision repair is impaired by binding of transcription factors to DNA.* Nature, 2016. **532**(7598): p. 264-7.
26. Rheinbay, E., et al., *Analyses of non-coding somatic drivers in 2,658 cancer whole genomes.* Nature, 2020. **578**(7793): p. 102-111.
27. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.
28. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.
29. Cameron, D.L., et al., *GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing.* Genome Biol, 2021. **22**(1): p. 202.
30. Rosenthal, R., et al., *DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution.* Genome biology, 2016. **17**(1): p. 1-11.
31. Blokzijl, F., et al., *MutationalPatterns: comprehensive genome-wide analysis of mutational processes.* Genome medicine, 2018. **10**(1): p. 1-11.
32. Tokez, S., et al., *Cumulative incidence and disease-specific survival of metastatic cutaneous squamous cell carcinoma: a nationwide cancer registry study.* J Am Acad Dermatol, 2021.
33. Budden, T., et al., *Female Immunity Protects from Cutaneous Squamous Cell Carcinoma.* Clin Cancer Res, 2021. **27**(11): p. 3215-3223.
34. Du, Q.Y., et al., *High STRN Expression Promotes HCC Invasion and Migration but Not Cell Proliferation or Apoptosis through Facilitating Epithelial-Mesenchymal Transition.* Biomed Res Int, 2020. **2020**: p. 6152925.
35. Shao, Y.W., et al., *Cross-species genomics identifies DLG2 as a tumor suppressor in osteosarcoma.* Oncogene, 2019. **38**(2): p. 291-298.
36. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415-21.
37. Hayward, N.K., et al., *Whole-genome landscapes of major melanoma subtypes.* Nature, 2017. **545**(7653): p. 175-180.
38. Yarchoan, M., A. Hopkins, and E.M. Jaffee, *Tumor Mutational Burden and Response Rate to PD-1 Inhibition.* N Engl J Med, 2017. **377**(25): p. 2500-2501.

39. Migden, M.R., et al., *Cemiplimab in locally advanced cutaneous squamous cell carcinoma: results from an open-label, phase 2, single-arm trial.* Lancet Oncol, 2020. **21**(2): p. 294-305.
40. Tate, J.G., et al., *COSMIC: the Catalogue Of Somatic Mutations In Cancer.* Nucleic Acids Res, 2019. **47**(D1): p. D941-d947.
41. Zollino, M., et al., *Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype.* Nat Genet, 2012. **44**(6): p. 636-8.
42. Fejzo, M.S., et al., *Analysis in epithelial ovarian cancer identifies KANSL1 as a biomarker and target gene for immune response and HDAC inhibition.* Gynecol Oncol, 2021. **160**(2): p. 539-546.
43. Lambert, S.R., et al., *Metastatic cutaneous squamous cell carcinoma shows frequent deletion in the protein tyrosine phosphatase receptor Type D gene.* Int J Cancer, 2012. **131**(3): p. E216-26.
44. Ortiz, B., et al., *Deletion of Ptprd and Cdkn2a cooperate to accelerate tumorigenesis.* Oncotarget, 2014. **5**(16): p. 6976-82.
45. Veeriah, S., et al., *The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9435-40.
46. Peyser, N.D., et al., *Loss-of-Function PTPRD Mutations Lead to Increased STAT3 Activation and Sensitivity to STAT3 Inhibition in Head and Neck Cancer.* PLoS One, 2015. **10**(8): p. e0135750.
47. Holmström, M.O., et al., *The calreticulin (CALR) exon 9 mutations are promising targets for cancer immune therapy.* Leukemia, 2018. **32**(2): p. 429-437.
48. Liu, P., et al., *Immunosuppression by Mutated Calreticulin Released from Malignant Cells.* Mol Cell, 2020. **77**(4): p. 748-760.e9.
49. Miyamoto, R., et al., *Prognostic significance of cyclin D1 amplification and overexpression in oral squamous cell carcinomas.* Oral Oncol, 2003. **39**(6): p. 610-8.
50. Yu, J., et al., *Genetic Aberrations in the CDK4 Pathway Are Associated with Innate Resistance to PD-1 Blockade in Chinese Patients with Non-Cutaneous Melanoma.* Clin Cancer Res, 2019. **25**(21): p. 6511-6523.
51. Rinaldi, J., et al., *The genomic landscape of metastatic breast cancer: Insights from 11,000 tumors.* PLoS One, 2020. **15**(5): p. e0231999.
52. Arao, T., et al., *FGF3/FGF4 amplification and multiple lung metastases in responders to sorafenib in hepatocellular carcinoma.* Hepatology, 2013. **57**(4): p. 1407-15.
53. Takahashi, R., et al., *Epigenetic deregulation of Ellis Van Creveld confers robust Hedgehog signaling in adult T-cell leukemia.* Cancer Sci, 2014. **105**(9): p. 1160-9.
54. Mamoor, S., *EVC is differentially expressed in lymph node metastasis in human breast cancer.* 2021.
55. Luo, W., et al., *Protein phosphatase 1 regulatory subunit 1A in ewing sarcoma tumorigenesis and metastasis.* Oncogene, 2018. **37**(6): p. 798-809.
56. Yuan, C.L., et al., *Identification of differentially expressed lncRNAs and mRNAs in luminal-B breast cancer by RNA-sequencing.* BMC Cancer, 2019. **19**(1): p. 1171.
57. Wu, L., et al., *Long non-coding RNA LINC01003 suppresses the development of multiple myeloma by targeting miR-33a-5p/PIM1 axis.* Leuk Res, 2021. **106**: p. 106565.
58. Ashford, B.G., et al., *Reviewing the genetic alterations in high-risk cutaneous squamous cell carcinoma: A search for prognostic markers and therapeutic targets.* Head Neck, 2017. **39**(7): p. 1462-1469.

59. Klinakis, A. and T. Rampias, *TP53 mutational landscape of metastatic head and neck cancer reveals patterns of mutation selection.* EBioMedicine, 2020. **58**: p. 102905.
60. Burtness, B., et al., *Correlation of tumor mutational burden (TMB) with CDKN2A and TP53 mutation in HPV-negative head and neck squamous cell carcinoma (HNSCC)*. 2020, American Society of Clinical Oncology.
61. Hazawa, M., et al., *ZNF750 is a lineage-specific tumour suppressor in squamous cell carcinoma.* Oncogene, 2017. **36**(16): p. 2243-2254.
62. Kosmider, O., et al., *TET2 gene mutation is a frequent and adverse event in chronic myelomonocytic leukemia.* Haematologica, 2009. **94**(12): p. 1676-81.
63. Xu, Y.P., et al., *Tumor suppressor TET2 promotes cancer immunity and immunotherapy efficacy.* J Clin Invest, 2019. **129**(10): p. 4316-4331.