This is a repository copy of *A novel genomic signature reclassifies an oral cancer subtype*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/86270/

Version: Accepted Version

**Article:**

https://doi.org/10.1002/ijc.29615

---

**Reuse**
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
A novel genomic signature reclassifies an oral cancer subtype

Manar Samman\textsuperscript{1,2}, Henry M Wood\textsuperscript{1}, Caroline Conway\textsuperscript{1,3}, Lucy Stead\textsuperscript{1}, Catherine Daly\textsuperscript{1}, Rebecca Chalkley\textsuperscript{1}, Stefano Berri\textsuperscript{1,4}, Burcu Senguven\textsuperscript{1,5}, Lisa Ross\textsuperscript{1}, Philip Egan\textsuperscript{1}, Preetha Chengot\textsuperscript{6}, Thian K Ong\textsuperscript{7}, Monica Pentenero\textsuperscript{6}, Sergio Gandolfo\textsuperscript{6}, Adele Cassenti\textsuperscript{9}, Paola Cassoni\textsuperscript{9}, Abdulaziz Al Ajlan\textsuperscript{10}, Alaa Samkari\textsuperscript{10}, William Barrett\textsuperscript{11}, Kenneth MacLennan\textsuperscript{1,6}, Alec High\textsuperscript{6,7}, Pamela Rabbitts\textsuperscript{1}.

* Joint first authors.

1. Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, LS9 7TF, UK.
2. King Fahad Medical City, Riyadh, Saudi Arabia.
3. Now at School of Biomedical Sciences, University of Ulster, Coleraine, Northern Ireland, BT52 1SA, UK
4. Now at Illumina UK Ltd., Chesterford Research Park, Saffron Walden, CB10 1XL, UK.
5. Now at Department of Oral Pathology, Faculty of Dentistry, Gazi University, Turkey.
6. St James’s Institute of Oncology, St James’s University Hospital, Leeds, LS9 7TF, UK.
7. Leeds Dental Institute, Leeds General Infirmary, Leeds, LS2 9LU, UK
8. Department of Oncology, Oral Medicine and Oral Oncology Unit, University of Torino, Turin, Italy.
9. Department of Medical Sciences, Pathology Unit, University of Torino, Turin, Italy.
10. National Guard Health Affairs, Saudi Arabia.
11. Department of Histopathology, Queen Victoria Hospital

* Contact – h.m.wood@leeds.ac.uk, +44 (0)113 2064070
Contact details:

Henry Wood

Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, LS9 7TF, UK.

h.m.wood@leeds.ac.uk, +44 (0)113 2064070

Running title: oral verrucous carcinoma genomic signature

Novelty and impact statement: Oral verrucous carcinoma is considered to be a variant of oral squamous cell carcinoma. We have performed the first genomic study to date on the largest cohort of cancers of this type. Complementary methodologies suggest that verrucous carcinomas have a distinct profile, separate from squamous disease, characterised by fewer changes, and lacking in common driver events. We suggest that these findings should lead to the reclassification of this disease subtype.

Keywords: Oral verrucous carcinoma, genomics, molecular signature


Category: Cancer genetics
Abstract

Verrucous carcinoma of the oral cavity (OVC) is considered a subtype of classical oral squamous cell carcinoma (OSCC). Diagnosis is problematic, and additional biomarkers are needed to better stratify patients. To investigate their molecular signature, we performed low coverage copy number sequencing on 57 OVC and exome and RNA sequencing on a subset of these and compared the data to the same OSCC parameters. Copy number results showed that OVC lacked any of the classical OSCC patterns such as gain of 3q and loss of 3p and demonstrated considerably fewer genomic rearrangements compared to the OSCC cohort. OVC and OSCC samples could be clearly differentiated. Exome sequencing showed that OVC samples lacked mutations in genes commonly associated with OSCC (TP53, NOTCH1, NOTCH2, CDKN2A and FAT1). RNA sequencing identified genes that were differentially expressed between the groups. In silico functional analysis showed that the mutated and differentially expressed genes in OVC samples were involved in cell adhesion and keratinocyte proliferation, while those in the OSCC cohort were enriched for cell death and apoptosis pathways.

This is the largest and most detailed genomic and transcriptomic analysis yet performed on this tumour type, which, as an example of non-metastatic cancer, may shed light on the nature of metastases. These three independent investigations consistently show substantial differences between the cohorts. Taken together they lead to the conclusion that OVC is not a subtype of OSCC, but should be classified as a distinct entity.
Introduction

An aim of the detailed analysis of cancer genomes is to discover genomic features that are prognostic for disease outcome and predictive of treatment response. Historically, genetic changes at specific loci have provided this information \(^1\) but more recently examination of the whole genome has been shown to have predictive power. We have illustrated the use of whole genome architecture to predict survival in squamous cell lung carcinoma patients \(^2\). Genomic analysis has been used to reclassify breast cancer tumours providing better prediction of disease outcomes \(^3\) and recently, molecular classification using multiple platforms has effectively clustered subtypes across different cancer tissue types, identifying unexpected associations that could influence choice of therapy \(^4\).

Patients with oral cancer would benefit from a biomarker that would indicate outcome, specifically regional metastatic spread. A “wait and see” approach could expose patients to the risk of under treatment of occult node metastases. Unnecessary elective neck dissection does have an associated morbidity and may remove or destroy a natural barrier to cancer spread. This is of outmost importance when considering the high risk of developing second primary tumours in case for oral cancer patients \(^5\).

OVC is a case in point. The WHO classification of oral cancer recognises verrucous carcinomas as a subtype of oral epithelial cancer \(^6\). It has a different morphological appearance being exophytic in its growth patterns and a different behaviour in that it is not associated with metastatic spread \(^7\). However some classical oral squamous carcinomas can have a verrucous appearance creating diagnostic uncertainty and subsequent dilemmas of clinical management \(^8\). We questioned whether genomic profiling could distinguish true verrucous carcinomas from classical squamous carcinomas that may have a verrucous appearance to produce a useful diagnostic biomarker. We used FFPE blocks as a source of tumour DNA for both next
generation sequencing at low coverage to generate copy number profiles, and exome sequencing. Additionally, RNA was isolated from the same blocks and transcription profiles generated using RNA-seq. Parallel analyses were undertaken for classical oral squamous cell carcinoma. Comparing the profiles, we found that true verrucous carcinomas were easily distinguishable from their classical counterparts; indeed no overlap was seen in patterns of genomic damage and transcriptional profiles could be used to clearly separate samples into their groups. Unlike classical oral tumours, TP53 was not inactivated either by mutation or HPV infection leading us to conclude that verrucous carcinoma should be considered a separate entity from classical squamous carcinoma.

Materials and methods

Sample selection

Four different sources of Formalin Fixed Paraffin-Embedded (FFPE) tissue were used in this study. Fifteen OVC FFPE blocks were retrieved from the Pathology Department, St James’s University Hospital, Leeds, UK; 15 OVC FFPE blocks were provided by the Pathology Division, Queen Victoria Hospital, West Sussex, UK; 40 OVC samples were provided by the Pathology Division, University of Torino, Italy and seven OVC FFPE blocks were provided by the Department of Pathology, National Guard Hospital, Saudi Arabia. Written informed consent was obtained from all patients for the use of their tissue in this research and the study approved by local ethics committees (REC ref no 07/Q1206/30 and 08/H1306/127). In total, 77 OVC lesions were identified and all diagnoses were confirmed by the study pathologist (AH). World health organisation (WHO) definitions and criteria were used for the histological diagnosis of OVC (Figure 1a and b). Squamous cell carcinoma lesions with verrucous appearance were classified as squamous cell carcinoma with papillary architecture (Figure 1b and c). Five of these samples were analysed by copy number sequencing (Figure 2d), following which the
samples were excluded from further analysis. For different reasons, such as low yields of
the extracted DNA and failed library preparation, 57 cases out of 77 were suitable for
NGS copy number analysis, all of which have previously been shown not to contain
Human Papilloma Virus (HPV) DNA \(^9\). Of these, 12 cases with adequate tissue were
used and successfully prepared for RNA and exome sequencing. From our ongoing
study of head and neck squamous cell carcinoma (HNSCC) \(^10\), we identified 45 HPV
negative OSCC patients for copy number (CN) analysis, 16 for RNA-seq, and 20 for
exome sequencing. Table 1S summaries the clinical data from these patients.

**DNA isolation for copy number analysis (CN)**
Tissue blocks were cut into sections and epithelial cells of interest macrodissected as
described previously using stained slides as a guide \(^9\). DNA extraction was performed on
macro-dissected FFPE tissue using Qiagen DNA extraction kits (Qiagen, Sussex, UK)
according to manufacturer’s instructions. DNA concentration and purity were quantified
using a Nanodrop UV spectrophotometer: Nanodrop-8000 (Thermo Scientific, UK).
Additionally, double-stranded DNA concentration was precisely measured using the
Quant-iT PicoGreen dsDNA BR assay (Invitrogen, UK).

**Nucleic acid isolation for exome and RNA sequencing**
DNA and RNA were simultaneously isolated from further macrodissected tissue from the
same patients using the AllPrep DNA/RNA FFPE kit (Qiagen, Sussex, UK) according to
manufacturer’s instructions. Nucleic acid concentration and purity were quantified as
before. RNA concentration was measured using The Qubit RNA Assay kit (Invitrogen,
UK).

**Copy number analysis: library preparation, sequencing and data analysis**
DNA libraries were prepared following two previously described protocols: for the Illumina
Genome analyser GAIIx sequencer \(^2\), and later, using NEBnext library preparation kits
(New England BioLabs, UK) for the Illumina HiSeq 2500 \(^9\). Samples were pooled for cluster amplification and multiplexed up to 20 samples per lane for Illumina Genome analyser GAIIx single end sequencing, and up to 40 samples per lane and paired-end sequenced (2X100bp) on an Illumina HiSeq 2500. Sample reads were split into separate files according to tags and aligned to the human reference genome. A control sample was pooled from a group of 20 British normal individuals downloaded from the 1000 Genomes Project \(^{11}\). Reads were trimmed of adaptors using cutadapt \(^{12}\), aligned to the human genome (hg19) using BWA \(^{13}\). Genomic copy number was analysed using CNAnorm \(^{14}\). The genome was split into 400Kb windows, and the ratios of tumour and normal read counts per window were counted and normalised.

Cumulative frequency karyograms were constructed for the OVC and OSCC cohorts by counting how many samples had gain and loss for each position along the genome. The copy number profiles of OVC and OSCC were also categorised using GISTIC2 \(^{15}\), identifying genes within copy number altered regions, and generating segmented copy number heat maps.

The ability to distinguish OVC and OSCC samples using their copy number profiles was also tested using a novel logistic regression technique \(^{16}\). Each sample was removed from the total data set in turn. The remaining samples were then used to build a predictive model. Each genomic window was given a score based on its ability to distinguish the two groups. The model was then applied to the test sample and a subtype prediction was made. This process was repeated for each sample, with the model being retrained on all other samples each time, so that no sample was predicted based on a training set of which it was a part.

**Exome sequencing: library preparation, sequencing and data analysis**

DNA sequencing libraries were separately prepared using the NEBNext singleplex library preparation kit (New England BioLabs, UK) and then exome DNA was
captured with Agilent exome capture kit (Agilent technologies Inc., USA). Briefly, 750ng of each library was hybridized for target enrichment, followed by washing, indexing and 10-cycle PCR amplification. Indexed samples were pooled up to four samples per lane and paired-end sequenced (2X100bp) on an Illumina HiSeq 2500. Sequencing was performed to an average of 90X coverage. Reads were trimmed using cutadapt\(^\text{12}\) to aligned to the human genome (hg19) using BWA\(^\text{13}\). PCR duplicates were removed using Picard (http://picard.sourceforge.net), and indel realignment and quality score calibration performed using GATK\(^\text{17}\). Variant calling was performed by Varscan2 in somatic mode\(^\text{18}\), and variant consequences were then predicted using the Variant Effect Predictor\(^\text{19}\). Variants were filtered using a number of criteria to enrich the mutation list for functionally important genes: Mutations had to pass a Varscan2 Phred somatic score of threshold of 15 (p-value of less than 0.05); mutations had to be present in over 50% of tumour cells, calculated taking into account tumour cell percentages and local copy number, as well as absent in the matched normal sample; mutations that have a predicted possible consequence on protein function (deleterious, splice variant, probably deleterious, possibly deleterious, stop gained).

The DAVID (Database for Annotation, Visualisation and Integrated Discovery) Bioinformatics Database 6.7 was used to produce lists of significantly enriched pathways from the mutated genes passing filters in the OVC and OSCC cohorts\(^\text{20}\).

**RNA-seq: library preparation, sequencing and data analysis**

cDNA Libraries were generated from the total RNA using ScriptSeq complete gold low input kit (Epicenter Biotechnologies, USA). All libraries were constructed according to manufacturer’s instructions and Ribo-Zero reagents were used to remove ~99% of rRNA (Table 4S).

Libraries were sequenced to an equivalence of 6 samples per lane of a HiSeq2500. Fastq files were processed using trim _galore to remove low quality bases, trim adaptors
and fix paired-end reads. Processed reads were aligned to the human genome GRCh37.p11 by Tophat2.0.7, using the gencode.v17 genome annotation as a guide. Reads could align a maximum of 5 times, with 2 mismatches. Alignment statistics (Table 4S) were ascertained using the samtools flagstat command and CollectRnaSeqMetrics program in the Picard software suite, version 1.56. Expression quantification and differential expression analysis (FDR of 0.01) were performed using cuffdiff, version 2.1.1, with multireads assigned using the –u parameter. Principal Component Analysis (PCA) was performed for all expressed protein-coding genes in OVC and OSCC samples, using the prcomp function in R. The DAVID Bioinformatics Database 6.7 web server was used to assess the functional enrichment of the significant differentially expressed genes between OVC and OSCC 20.

Results

DNA copy number analysis of OVC versus OSCC

Between one and ten million reads per sample were generated for copy number analysis, equating to one read to every 300bp-3Kb, or 0.033X-0.33X coverage. Copy number karyograms were constructed from 57 OVC and 45 OSCC cases using CNAnorm 14. Individual karyograms representative of normal epithelium, OVC, OSCC and OSCC with verrucous architecture are shown in Figure 2. In general, visual inspection of OSCC karyograms showed more whole chromosome and localized gain and loss and a higher degree of aneuploidy across the whole genome when compared to OVCs. Cumulative frequency karyograms and GISTIC heatmaps for each tumour type are shown in figure 3.

OSCC and indeed SCC from other sites have characteristic genomic features such as gains in chromosomal arms 3q, 5p, 7p, 8q and 20p, and losses in 3p, 4p, 8p, and
18q\(^{21, 22}\). All these chromosomal abnormalities were seen in our OSCC tumours but none was detected in the OVC group. However, OVC has its own characteristic features, namely: gains of 7q11.2, 7q22, 3p21, 15q15, 16q22, and 17q23. Chromosomal losses were infrequent in OVC but detected at 6p21 and 17q12. Samples of OSCC with verrucous architecture all showed karyograms consistent with classical OSCC.

Visual inspection of individual and cumulative karyograms indicated that there was a substantial difference between the copy number landscape of OVC and OSCC genomes. We tested this by applying a novel logistic regression method that blindly predicts the subtype of an ‘unknown’ sample based on the copy number of two groups of ‘known’ samples. Of the 45 OSCC samples, four were misclassified, whereas only one out of 57 OVC samples was incorrectly predicted.

**Mutation detection in OVC compared to OSCC**

Exome sequencing data was produced from the tumours and matched adjacent normal epithelium for 12 OVC and 20 OSCC patients. Two OVC patients were excluded due to evidence of severe DNA damage resulting from tissue fixation i.e. low coverage, high mismatch rate, and disproportionally high ratio of C/T mismatches. The remaining patients’ tumour DNAs were sequenced to an average of 90X coverage.

From the 12 OVC patients, a total of 224 somatically mutated genes passed filters, ranging from between nine (V-123-01) to 53 (V-116-01) genes per patient. OVC patients’ individual mutation profiles are listed in Table 2S. From the 20 OSCC patients, 1347 mutated genes passed filters, ranging from 32 (PG-105) to 143 (PG-144) per patient.

For both tumour cohorts the most frequently occurring mutations are listed in Table1.
The top five genes (*TP53, CDKN2A, NOTCH1, NOTCH2* and *FAT1*) in the OSCC cohort have previously been identified as significant HNSCC cancer genes. None of these genes was found to be mutated in the OVC cohort. However, four genes (*DPSP, MUC4, NEFH, ANP32E*) were mutated in the verrucous cohort with a frequency greater than 30% (Table 1 and Table 2S). These genes were not unique to OVC, being also detected at lower frequencies in our OSCC cohort (Table 1) although not detected in other HNSCC exome studies.

For verrucous carcinoma, no genes were mutated in over 50% of patients. Nonetheless, the use of DAVID functional analysis revealed involvement of a significant number of OVC mutated genes involved in cell adhesion and keratinocyte proliferation (Table 3S), while mutated genes in the OSCC cohort were more involved in cell death and apoptosis pathways (Table 3S).

**Gene expression in OVC compared to OSCC**

Gene transcription profiles were generated using RNA-Seq for all twelve OVC, with an average of 51412094 reads (ranging from 25602577 to 569763697), and with a median of 82% mapped reads. Ribosomal RNA ranged between 0% - 0.1% of the total reads (Table 4S). Gene expression was quantified as FPKM (Fragments per Kilobase per million Mapped) for each protein-coding and non-coding gene. The threshold of 0.1 FPKM was used to determine whether or not a gene was expressed. Gene transcription data were similarly obtained from 16 OSCC samples. Significant differential expression lists for protein-coding genes in oral verrucous samples versus OSCCs are illustrated in Table 2.

The use of DAVID functional analysis revealed that significant differentially upregulated genes in OVC versus OSCC are involved in keratinocyte differentiation and epithelium development (Table 5S), while the significant differentially upregulated genes in OSCC versus OVC were more involved in cell growth and
migration and angiogenesis (Table 5S). Principal Component Analysis (PCA) of all
expressed protein-coding genes in OVC and OSCC revealed an evident separation
of the two cohorts (Figure 4).

**Comparing genomic changes and expression levels**

The copy number and expression levels of all significantly differentially expressed
genes were correlated. Most genes showed no link between expression and copy
number. Only three genes, *SERPINE1, CDH3* and *PLA2G4D* were both significantly
overexpressed in OVC samples and found in recurrent regions of gain.

Only one gene reported as mutated in the exome data was significantly differentially
expressed. A missense variant in *CXCL5* was reported in one OVC sample. This
gene was under-expressed in OVC samples compared to OSCC.

**Discussion**

Clinically, OVC appear as exophytic masses with a verrucous surface but, less often,
you may be relatively smooth. Furthermore OSCC can have a superficial verrucous
appearance thus morphological features alone can lead to misclassification with
consequences for patient management that many include unnecessary neck
dissections.\(^{25}\). For this reason, new approaches are needed. Recognising that
genomic changes drive tumour development, we reasoned that verrucous carcinoma
with its more indolent natural history would have a simpler genotype compared to the
more aggressive classical oral squamous cell tumours. In fact this had already been
demonstrated superficially by flow cytometry studies.\(^{26}\); our intention was to define
these genomic changes more precisely, providing information that could be
developed for diagnostic biomarkers. Previously, we showed that whole genome
sequencing at low coverage could reveal differences in HPV positive and negative
head and neck cancers.\(^{10}\) and therefore we applied this approach, creating
karyograms to compare OVC and OSCC. Additionally we included exome sequencing and RNA-seq in our comparative analysis to provide further precision in defining genomic regions that distinguished the two oral cancer subtypes. This study represents the largest and ‘first’ study, to date, to inspect genomic copy number patterns, somatic gene mutations and transcriptional changes that occur in OVC and compare them with the copy number, mutational and transcriptional events found in OSCC. All earlier studies on OVC were either case reports or immunohistochemistry studies.

The rarity of these lesions makes them difficult to investigate and previous immunohistochemistry studies on oral verrucous lesions have yielded mixed results. Variations in samples, sample numbers, staining procedures and analysis methods, along with difficulties in defining ‘gold-standard’ histological criteria for diagnosis may explain the lack of concordance between these studies.

Consistent with the analyses by flow cytometry, visual examination of the 57 OVC CN traces revealed a lower level of genomic damage when compared with 45 OSCC samples. This suggests that OVC is characterised by a lower degree of chromosomal instability than OSCC. This lower level of chromosomal instability could be linked to the minimal or absent histological cytological atypia found in OVC.

We used logistic regression analysis to obtain statistical quantitation of the copy number differences, correctly identifying 56 out of 57 OVC samples and 41 out of 45 OSCCs.

Interestingly, losses were detected frequently in OSCC genomes but rarely in OVCs. Losses on chromosomal arms 3p, 4q, 9p and 18q, and gains on 3q, 5q, 8q, and 20p are chromosomal signatures commonly linked with OSCCs and were frequently identified in this OSCC cohort but were absent in OVC, suggesting that these CN alterations may be related to the more aggressive behaviour of OSCC. However, OVC karyograms revealed their own distinctive features that were consistent for the subtype, namely gains at 7q11.2, 7q22, and 17q23, as well as loss at 17q12 at a
frequency of ~50%. These changes have not been previously identified as common CN altered chromosome lesions in oral cancer. Comparison of the mutation profiles of the two sub-groups provided further evidence that they had distinguishable genotypes. Four genes are mutated in more than one OVC sample, suggesting that they may have a role in the development of OVC lesions. These genes are: DSPP gene (mutated in 40% of OVC cohort), MUC4, NEFH and ANP32E (mutated in 30% of OVC cohort). These four genes are also mutated in at least one OSCC sample and have all been seen in other cancers. Earlier studies reported up-regulation of DSPP along with other genes in histologically aggressive and poorly differentiated OSCC and in some oral epithelial dysplasia. Other studies have demonstrated a positive correlation between MUC4 expression and tumour growth and malignant progression in pancreatic lesions. NEFH was down-regulated along with other genes in metastatic lung squamous cell carcinoma samples when compared with non-metastatic samples. ANP32E expression was up-regulated in gastric cancer cells when compared with non-neoplastic epithelial cells. All three OVC samples with an ANP32E mutation had missense mutations in exactly the same position, which makes this gene a strong candidate for a role in the development of oral verrucous tumours. ANP32E has a variant that has been reported to be associated with breast cancer development. Two of the four samples showing DSPP mutations exhibited identical in-frame deletions, as did two of the three were NEFH mutations. Interestingly, analysis of the exome data for all ten OVC samples showed no mutations within the TP53, CDKN2A, NOTCH1, NOTCH2 and FAT1 genes, all of which were frequently mutated in OSCC samples in this study and previously published work. In particular, mutation of the tumour-suppressor, TP53 gene, is among the earliest identified genetic changes and the most common in HNSCC, arising in over half of all cases. No mutations were found in any of the ten OVC samples. The most recent gene expression study compared six patients with primary
OSCC and five patients with primary OVC using microarray technology. Of the 167 differentially expressed genes (DEGs) reported, 39 were shared between OSCC and OVC, and eight (HLF, TGFB1, SERPINE1, MMP1, INHBA, COL4A2, COL4A1, and ADAMTS12) were significantly differentially expressed between the two groups. Seven of the eight were up-regulated in OSCC compared to OVC, with only HLF being comparatively up-regulated in OVC. For all eight significantly differentially expressed genes between the two groups previously reported, similar significant expression changes were observed for in our study, with HLF being the only gene from that previous list significantly up-regulated in OVC compared to OSCC. HLF is thought to be tumour suppressor gene that plays a role in the detoxification processes.

Several of the genes differentially expressed between the two groups offer plausible explanations for the relatively benign nature of OVC lesions, and their correspondingly low metastatic potential, when compared to classical OSCC.

The TNFRSF12A, IGFBP6, FSTL3, SMR3B and LAMC2 protein coding genes have previously been reported as highly expressed in head and neck cancer. These were overexpressed in the OSCC cohort compared to the OVC samples. In contrast, down regulation of the DLG2, HBB, and DSC1 genes has been associated with progression. These were all expressed more highly in the OVC samples. Similarly, KRT76 and KRT2 have been reported to be down regulated in OSCC and were again more highly expressed in our OVC samples.

Amongst genes associated with metastasis, down regulation of HPGD has been shown to be a marker for metastasis. This gene is more highly expressed in the OVC cohort. In contrast, over expression of CXCL5, THBS1, MT2A and MT1X has been linked to metastasis. All of these genes were more highly expressed in the OSCC cohort.
DAVID functional enrichment and pathway analysis on RNA and exome sequencing OVC data revealed that a significant number of the differentially expressed genes and mutated genes are located in the plasma membrane, participate in cell adhesions and keratinocyte differentiation, and implemented in calcium ion ($\text{Ca}^{2+}$) binding. On the other hand, the OSCC data revealed that the majority of the up-regulated and mutated genes located in the cytoskeleton, participated in cell adhesion and migration, angiogenesis and cell death, and were implemented in nucleotide and growth factor binding.

We have undertaken to better define the genomic and transcriptomic changes associated with oral verrucous carcinoma, and compare them to the changes observed in a similar group of classical oral squamous cell carcinomas. We used the three independent, complementary techniques of copy number sequencing, exome sequencing and RNA sequencing. For all three methods, there were clear differences that distinguish the OVC and OSCC cohorts. Taken together, these three independent analyses offer persuasive evidence that OVC should not be classified as a subtype of OSCC, but should be considered a separate disease entity.

Using a number of genomic platforms, it has recently been shown that tumours are not always closely related by tissue of origin \cite{4}. We have shown here that this applies to these two tumours, verrucous and squamous, both arising in the oral cavity. Verrucous carcinoma although rarely metastatic can be locally destructive and invasive and may benefit from therapeutic approaches in addition to surgery. The genomic and transcriptomic changes described here may suggest routes to the identification of a drug target, specific for these verrucous tumours. It would be of interest to determine if verrucous carcinomas that occur at other tissue sites share the same molecular signatures as those we have described for the oral cavity and possibly grouped together for the identification of treatment targets.
Conflict of interest statement: None of the authors of this manuscript has any conflict of interest.

Acknowledgments: This work was funded by a scholarship to M. Samman from the Saudi Arabian Ministry of Higher Education; King Fahad Medical City, Saudi Arabia; Yorkshire Cancer Research (L341PG), the Betty Wolsey Endowment, Wellcome Trust and University of Leeds [personal fellowship awarded to LS: RGCALA101195]. The funders were not involved in study design, sample collection, data analysis, decision to publish, or manuscript preparation.

References


Tables

Table 1:
Frequently occurring mutations in OVC and OSCC cohorts

<table>
<thead>
<tr>
<th>Genes</th>
<th>Mutated in OSCC</th>
<th>Percentage</th>
<th>Mutated in OVC</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>Yes</td>
<td>70%</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>CDKN2A</td>
<td>Yes</td>
<td>35%</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>NOTCH1</td>
<td>Yes</td>
<td>10%</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>NOTCH2</td>
<td>Yes</td>
<td>25%</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>FAT1</td>
<td>Yes</td>
<td>20%</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>DSPP</td>
<td>Yes</td>
<td>5%</td>
<td>Yes</td>
<td>40%</td>
</tr>
<tr>
<td>MUC4</td>
<td>Yes</td>
<td>15%</td>
<td>Yes</td>
<td>30%</td>
</tr>
<tr>
<td>NEFH</td>
<td>Yes</td>
<td>20%</td>
<td>Yes</td>
<td>30%</td>
</tr>
<tr>
<td>ANP32E</td>
<td>Yes</td>
<td>5%</td>
<td>Yes</td>
<td>30%</td>
</tr>
</tbody>
</table>
Table 2:
Significant differential expression list for protein-coding genes in oral verrucous samples versus OSCCs. Log FC refers to the log of the fold change. P.adj is the adjusted p-value. Positive log fold change indicates genes overexpressed in OVC, while negative values indicate genes overexpressed in OSCC.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene Function</th>
<th>Log FC</th>
<th>P.adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNC45B</td>
<td>Myoblast fusion</td>
<td>5.73176541</td>
<td>1.02E-09</td>
</tr>
<tr>
<td>ANKRD30BL</td>
<td></td>
<td>5.068125637</td>
<td>3.00E-09</td>
</tr>
<tr>
<td>KRT2</td>
<td>Keratin gene, involved in differentiation</td>
<td>3.28887873</td>
<td>4.25E-07</td>
</tr>
<tr>
<td>DLG2</td>
<td>Guanylate kinase</td>
<td>3.142861272</td>
<td>3.25E-06</td>
</tr>
<tr>
<td>RP11-181C3,1</td>
<td></td>
<td>2.589914427</td>
<td>4.46E-06</td>
</tr>
<tr>
<td>KRT76</td>
<td>Keratin gene, structural integrity of epithelial cells</td>
<td>4.393647354</td>
<td>4.37E-05</td>
</tr>
<tr>
<td>LOR</td>
<td>Keratinocyte cell envelope protein</td>
<td>3.696791035</td>
<td>0.000182642</td>
</tr>
<tr>
<td>ELOVL4</td>
<td>Elongates fatty acids</td>
<td>1.913915467</td>
<td>0.000341653</td>
</tr>
<tr>
<td>HTRA3</td>
<td>Probable serine protease</td>
<td>-1.624779995</td>
<td>0.000341653</td>
</tr>
<tr>
<td>PDK4</td>
<td>Glucose metabolism</td>
<td>-2.442699908</td>
<td>0.000497839</td>
</tr>
<tr>
<td>FLG2</td>
<td>Filaggrin family</td>
<td>4.155257725</td>
<td>0.000682528</td>
</tr>
<tr>
<td>MT2A</td>
<td>Binds heavy metals</td>
<td>-2.067002668</td>
<td>0.000737808</td>
</tr>
<tr>
<td>HPGD</td>
<td>Prostaglandin inactivation. Inhibits proliferation in colon cancer.</td>
<td>1.75292207</td>
<td>0.00078544</td>
</tr>
<tr>
<td>TNFRSF12A</td>
<td>Induces apoptosis, promotes angiogenesis, may modulate cell adhesion to matrix proteins.</td>
<td>-1.923891955</td>
<td>0.000790342</td>
</tr>
<tr>
<td>IGFBP6</td>
<td>Affects growth-promoting effects of insulin-like growth factors.</td>
<td>-2.386076713</td>
<td>0.000966215</td>
</tr>
<tr>
<td>ATP10B</td>
<td></td>
<td>1.718836376</td>
<td>0.001069055</td>
</tr>
<tr>
<td>HBB</td>
<td>Affects blood pressure</td>
<td>3.554687453</td>
<td>0.001114319</td>
</tr>
<tr>
<td>SLC11A1</td>
<td>Iron metabolism</td>
<td>-1.677099745</td>
<td>0.001188788</td>
</tr>
<tr>
<td>Gene</td>
<td>Function</td>
<td>Fold Change (log2)</td>
<td>p-value</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>FDCSP</td>
<td>Binds to surface of B-lymphoma cells</td>
<td>3.705508687</td>
<td>0.001313018</td>
</tr>
<tr>
<td>FSTL3</td>
<td>Bone formation, differentiation of haemopoietic progenitor cells</td>
<td>-1.641022383</td>
<td>0.001395803</td>
</tr>
<tr>
<td>PTGDR2</td>
<td>Prostaglandin receptor</td>
<td>2.991133602</td>
<td>0.001395803</td>
</tr>
<tr>
<td>DSC1</td>
<td>Keratinisation of epithelial tissue.</td>
<td>2.515586707</td>
<td>0.001634268</td>
</tr>
<tr>
<td>TCAP</td>
<td>Muscle assembly</td>
<td>-4.437644385</td>
<td>0.001634268</td>
</tr>
<tr>
<td>DDIT4</td>
<td>Promotes neuronal cell death</td>
<td>-2.274816537</td>
<td>0.002274577</td>
</tr>
<tr>
<td>INPP5F</td>
<td>Decreases AKT and GSK3B phosphorylation</td>
<td>2.093068765</td>
<td>0.002500484</td>
</tr>
<tr>
<td>PID1</td>
<td></td>
<td>1.742685358</td>
<td>0.002934709</td>
</tr>
<tr>
<td>ENTPD3</td>
<td>ATD/ADP hydrolysis</td>
<td>1.037851739</td>
<td>0.002934709</td>
</tr>
<tr>
<td>PLA2G4D</td>
<td>Inflammation</td>
<td>2.117301287</td>
<td>0.003105351</td>
</tr>
<tr>
<td>CTC-236F12.4</td>
<td></td>
<td>-2.189567433</td>
<td>0.005454379</td>
</tr>
<tr>
<td>PDLIM3</td>
<td>Actin organisation</td>
<td>-2.03957167</td>
<td>0.003434804</td>
</tr>
<tr>
<td>C10orf53</td>
<td></td>
<td>2.290929173</td>
<td>0.004033844</td>
</tr>
<tr>
<td>THBS1</td>
<td>Cell-cell and cell-matrix interactions.</td>
<td>-2.189567433</td>
<td>0.005454379</td>
</tr>
<tr>
<td>CXCL5</td>
<td>Neutrophil activation</td>
<td>-2.570674984</td>
<td>0.006168571</td>
</tr>
<tr>
<td>SERPINH1</td>
<td>Collagen binding.</td>
<td>-1.43209521</td>
<td>0.007136242</td>
</tr>
<tr>
<td>MT1X</td>
<td>Binds heavy metals.</td>
<td>-1.54913047</td>
<td>0.007136242</td>
</tr>
<tr>
<td>WNT9A</td>
<td>Probable developmental protein.</td>
<td>-1.802040084</td>
<td>0.007747887</td>
</tr>
<tr>
<td>LCE1A</td>
<td>Precursor of stratum corneum</td>
<td>2.964477071</td>
<td>0.008599762</td>
</tr>
<tr>
<td>LAMC2</td>
<td>Migration, attachment and organisation of cells in embryonic development.</td>
<td>-2.28993045</td>
<td>0.008730485</td>
</tr>
<tr>
<td>CYC1</td>
<td>Mitochondrial activity</td>
<td>-1.369128527</td>
<td>0.008958146</td>
</tr>
<tr>
<td>SMR3B</td>
<td>Androgen regulated protein.</td>
<td>-3.661364151</td>
<td>0.009270752</td>
</tr>
<tr>
<td>HLF</td>
<td>Member of PAR family</td>
<td>1.349932377</td>
<td>0.009352683</td>
</tr>
<tr>
<td>CTGF</td>
<td>Cell proliferation, differentiation and adhesion.</td>
<td>-1.972202278</td>
<td>0.009409324</td>
</tr>
</tbody>
</table>
Figure legends:

Figure 1: Photomicrographs of oral SCC with papillary architecture (a and b) and OVC (c, d, e and f).
(a): H&E stain of oral SCC with papillary architecture at low power (magnification: x25 approximately), but with cytological atypia and invasion present (b) seen at magnification: x100 approximately. (c): H&E stain showing OVC classical features at low power (magnification: x40 approximately) and without cytological atypia or invasion (d) seen at magnification: x100 approximately (e) seen at magnification: x200 approximately (f) seen at magnification: x250 approximately.

Figure 2:
Examples of the genomic profiles (karyograms) of a. histologically normal oral epithelium (a patient with an ulcerative lesion), b. OVC, c. OSCC, d. OSCC with papillary architecture. Each data point represents one window of approximately 400 reads. Genomic position is on the x-axis and tumour:normal ratio is on the y-axis. The black lines are regions of common copy number between breakpoints. Windows of gain and loss are red and blue respectively. Purple and orange arrows respectively point to chromosomal gains and losses common to that tumour type.

Figure 3:
Frequency of genomic gain and loss for OVC (a) and OSCC (b). Genomic position is on the x-axis, frequency (%) of gains (red) and losses (blue) are shown on the y-axis. Heat map images of OVC (c) and OSCC (d) based on total segmented DNA copy number variation profiles. Heat map images were analysed using (GISTIC2.0).
In each heat map, the samples are arranged from left to right, and chromosomes are arranged vertically from top to bottom. Red represents CN gain and blue represents CN loss.

**Figure 4:**
Principal Component Analysis (PCA) biplot of PCs 1 and 2 using all expressed genes. This biplot best separates the two oral tumour groups: V – Verrucous carcinoma, and S – Squamous cell carcinoma.

**Supplementary:**

**Table 1S:**
Patients’ clinical information. (-) stands for unknown information, (oral cavity) stands for unknown precise location within the oral cavity.

**Table 2S:**
OVC patients’ individual mutation profiles from exome sequencing.

**Table 3S:**
David functional analysis for the mutated genes harbouring putatively functional variants in OVC and OSCC cases.

**Table 4S:**

**Table 5S:**
David functional analysis for differentially overexpressed genes in OVC and OSCC cohorts.