IJC
International Journal of Cancer

A novel genomic signature reclassifies an oral cancer subtype

Manar Samman[1,2], Henry M. Wood[1], Caroline Conway[1], Lucy Stead[1], Catherine Daly[1], Rebecca Chalkley[1], Stefano Berri[1], Burcu Senguven[1], Lisa Ross[1], Philip Egan[1], Preetha Chengot[3], Thian K. Ong[4], Monica Pentenero[5], Sergio Gandolfo[5], Adele Cassenti[6], Paola Cassoni[6], Abdulaziz Al Ajlan[7], Alaa Samkari[7], William Barrett[8], Kenneth MacLennan[1,3], Alec High[3,4] and Pamela Rabbitts[1]

[1] Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, United Kingdom
[2] Pathology and Clinical Laboratory Department, King Fahad Medical City, Riyadh, Saudi Arabia
[3] St James's Institute of Oncology, St James's University Hospital, Leeds, United Kingdom
[4] Leeds Dental Institute, Leeds General Infirmary, Leeds, United Kingdom
[5] Oral Medicine and Oral Oncology Unit, Department of Oncology, University of Torino, Turin, Italy
[6] Pathology Unit, Department of Medical Sciences, University of Torino, Turin, Italy
[7] National Guard Health Affairs, Riyadh, Saudi Arabia
[8] Department of Histopathology, Queen Victoria Hospital

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 (CN) sequencing on 57 OVC and exome and RNA sequencing on a subset of these and compared the data to the same OSCC parameters. CN 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.

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

Key words: oral verrucous carcinoma, genomics, molecular signature

Abbreviations: CN: copy number; DEGs: differentially expressed genes; FFPE: formalin-fixed paraffin-embedded; FPKM: fragments per kilobase per million mapped; HNSCC: head and neck squamous cell carcinoma; HPV: human papillomavirus; OSCC: oral squamous cell carcinoma; OVC: oral verrucous carcinoma; PCA: principal component analysis; WHO: World Health Organisation

Additional Supporting Information may be found in the online version of this article.

Conflict of interest: Nothing to report

M.S. and H.M.W. are joint first authors

Caroline Conway's current address is School of Biomedical Sciences, University of Ulster, Coleraine, Northern Ireland, BT52 1SA, United Kingdom

Stefano Berri's current address is Illumina UK Ltd., Chesterford Research Park, Saffron Walden, CB10 1XL, United Kingdom

Burcu Senguven's current address is Department of Oral Pathology, Faculty of Dentistry, Gazi University, Ankara, Turkey

Grant sponsors: Saudi Arabian Ministry of Higher Education, King Fahad Medical City, Saudi Arabia, Betty Wolsey Endowment, Wellcome Trust; **Grant sponsor:** Yorkshire Cancer Research; **Grant number:** L341PG; **Grant sponsor:** University of Leeds; **Grant number:** RGCALA101195

DOI: 10.1002/ijc.29615

History: Received 5 Jan 2015; Accepted 15 May 2015; Online 25 May 2015

Correspondence to: Henry Wood, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds LS9 7TF, United Kingdom, Tel.: +44-0-113-2064070, E-mail: h.m.wood@leeds.ac.uk

What's new?

Oral verrucous carcinoma (OVC) is considered to be a variant of oral squamous cell carcinoma (OSCC), due mainly to similarities in appearance. However, observed differences in growth patterns and metastatic behavior have raised questions about their classification. Here, comparisons of genomic data derived from copy number sequencing, exome sequencing, and RNA sequencing indicate that OVC is distinct from its squamous cell counterpart. In particular, OVCs were characterized by fewer genomic changes than OSCCs, and the two lacked common driver events. The unique genomic features of OVC warrant its reclassification within the oral cancer taxonomy.

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 formalin-fixed paraffin-embedded (FFPE) blocks as a source of tumour DNA for both next-generation sequencing at low coverage to generate copy number (CN) profiles,[2] 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 (OSCC). 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,[9] leading us to conclude that verrucous carcinoma should be considered a separate entity from classical squamous carcinoma.

Material and Methods

Sample selection

Four different sources of 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 was 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[6] (Figs. 1*a* and 1*b*). Squamous cell carcinoma lesions with verrucous appearance were classified as squamous cell carcinoma with papillary architecture[6] (Figs. 1*b* and 1*c*). Five of these samples were analysed by CN sequencing (Fig. 2*d*), 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 of 77 cases were suitable for NGS CN analysis, all of which have previously been shown not to contain human papillomavirus (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 CN analysis, 16 for RNA-seq and 20 for exome sequencing. Supporting Information Table 1S summaries the clinical data from these patients.

DNA isolation for CN analysis

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 macrodissected FFPE tissue using Qiagen DNA extraction kits (Qiagen, Sussex, UK) according to the manufacturer's instructions. DNA concentration and purity were quantified using a Nanodrop UV spectrophotometer: Nanodrop-8000 (Thermo Scientific, Loughborough, UK). Additionally, double-stranded DNA

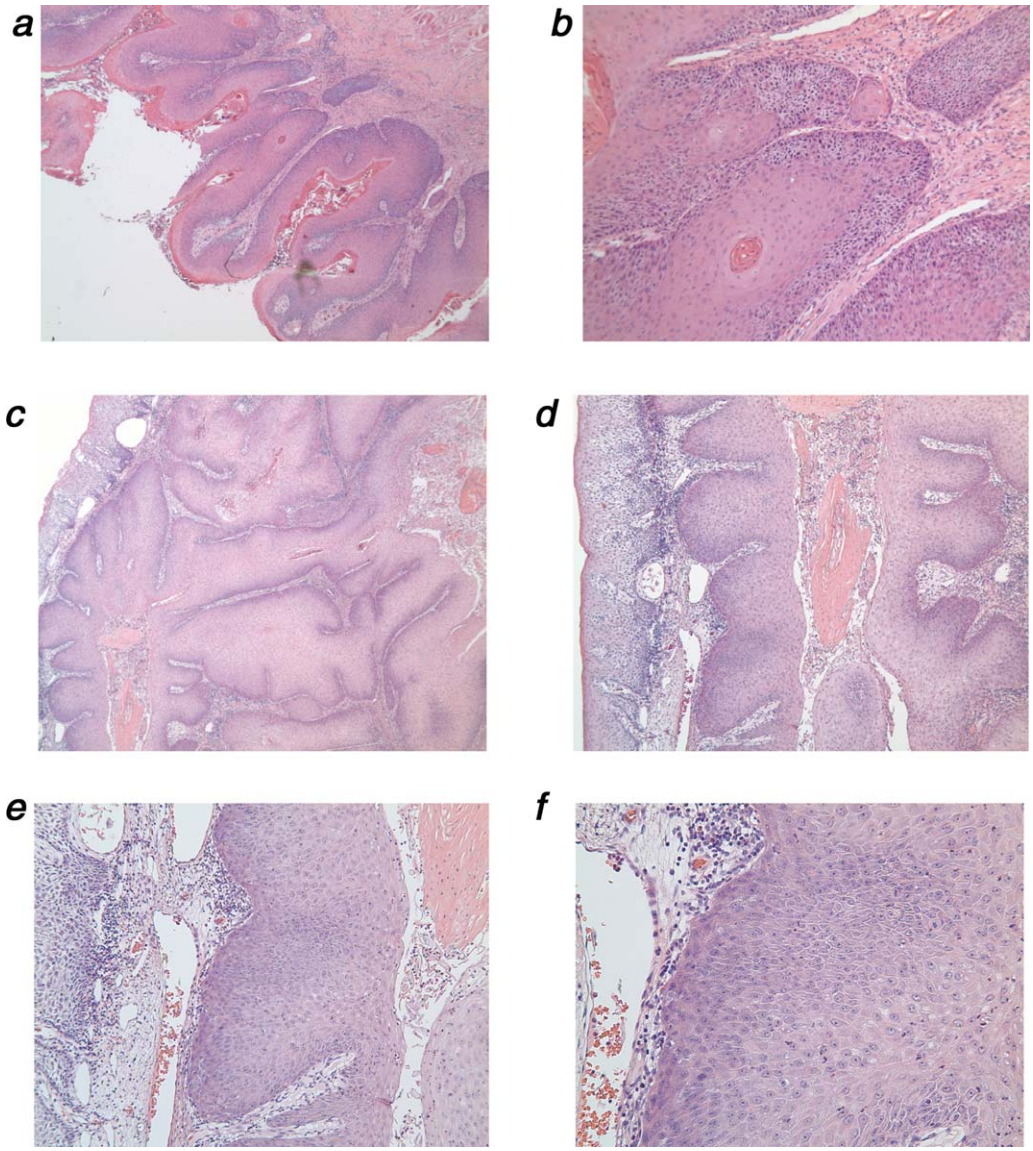

Figure 1. Photomicrographs of oral SCC with papillary architecture (*a* and *b*) and OVC (*c–f*). (*a*) H&E stain of oral SCC with papillary architecture at low power (magnification: ~ ×25), but with cytological atypia and invasion present (*b*) seen at magnification: ~ ×100. (*c*) H&E stain showing OVC classical features at low power (magnification: ~ ×40) and without cytological atypia or invasion (*d*) seen at magnification: ~ ×100, (*e*) magnification: ~ ×200 and (*f*) magnification: ~ ×250.

concentration was precisely measured using the Quant-iT PicoGreen dsDNA BR assay (Invitrogen, Paisley, 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 All-Prep DNA/RNA FFPE kit (Qiagen) according to the manufacturer's instructions. Nucleic acid concentration and purity were quantified as before. RNA concentration was measured using The Qubit RNA Assay kit (Invitrogen).

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, Hitchin, 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 (2X100 bp) 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 CN was analysed using CNAnorm.[14] The genome was split into 400 kb

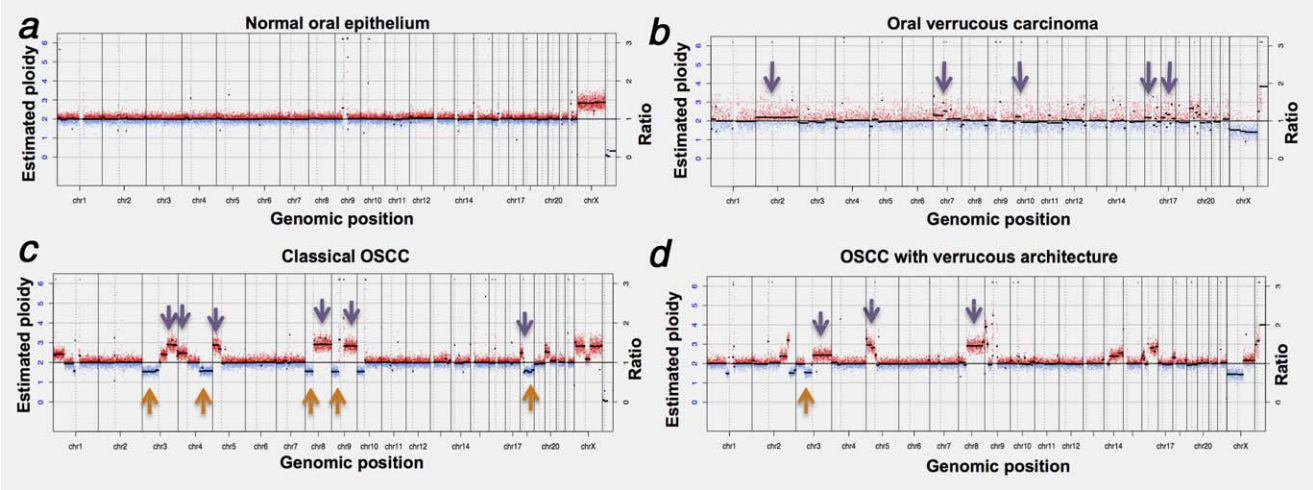

Figure 2. Examples of the genomic profiles (karyograms) of (*a*) histologically normal oral epithelium (a patient with an ulcerative lesion), (*b*) OVC, (*c*) OSCC, (*d*) OSCC with papillary architecture. Each data point represents one window of ~400 reads. Genomic position is on the *x*-axis and tumour:normal 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.

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 CN profiles of OVC and OSCC were also categorised using GISTIC2,[15] identifying genes within CN-altered regions, and generating segmented CN heat maps.

The ability to distinguish OVC and OSCC samples using their CN 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) and then exome DNA was captured with Agilent exome capture kit (Agilent Technologies, Stockport, UK). Briefly, 750 ng of each library was hybridized for target enrichment, followed by washing, indexing and ten-cycle PCR amplification. Indexed samples were pooled up to four samples per lane and paired-end sequenced (2X100 bp) on an Illumina HiSeq 2500. Sequencing was performed to an average of 90× coverage. Reads were trimmed using cutadapt[12] to aligned to the human genome (hg19) using BWA.[13] PCR duplicates were removed using Picard (http://picard.sourceforge.net), and indel realignment and quality score calibration were performed using GATK.[17] Variant calling was performed by Varscan2 in somatic mode,[18] and variant consequences were then predicted using the Variant Effect Predictor.[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 values <0.05); mutations had to be present in more than 50% of tumour cells, calculated taking into account tumour cell percentages and local CN, 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.[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, Madison, WI, USA). All libraries were constructed according to the manufacturer's instructions and Ribo-Zero reagents were used to remove ~99% of rRNA (Supporting Information Table 4S).

Libraries were sequenced to an equivalence of six 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 five times, with two mismatches.

Alignment statistics (Supporting Information 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 (DEGs) between OVC and OSCC.[20]

Results

DNA CN analysis of OVC *versus* OSCC

Between 1 and 10 million reads per sample were generated for CN analysis, equating to one read to every 300 bp–3 kb, or 0.033×–0.33× coverage. CN 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 CN 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 CN of two groups of "known" samples. Of the 45 OSCC samples, four were misclassified, whereas only one of 57 OVC samples was incorrectly predicted.

Mutation detection in OVC compared to OSCC

Exome sequencing data were produced from the tumours and matched adjacent normal epithelium for 12 OVC and 20 OSCC patients. Two OVC patients were excluded owing 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 90× coverage.

From the 12 OVC patients, a total of 224 somatically mutated genes passed filters, ranging from between 9 (V-123-01) and 53 (V-116-01) genes per patient. OVC patients' individual mutation profiles are listed in Supporting Information 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 Table 1. The top five genes (*TP53*, *CDKN2A*, *NOTCH1*, *NOTCH2* and *FAT1)* in the OSCC cohort have previously been identified as significant HNSCC cancer genes.[23,24] None of these genes was found to be mutated in the OVC cohort. However, four genes (*DSPP*, *MUC4*, *NEFH* and *ANP32E*) were mutated in the verrucous cohort with a frequency >30% (Table 1 and Supporting Information 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.[23,24]

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

Gene expression in OVC compared to OSCC

Gene transcription profiles were generated using RNA-Seq for all 12 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 and 0.1% of the total reads (Supporting Information 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 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 (Supporting Information Table 5S), while the significant differentially upregulated genes in OSCC *versus* OVC were more involved in cell growth and migration and angiogenesis (Supporting Information Table 5S). PCA of all expressed protein-coding genes in OVC and OSCC revealed an evident separation of the two cohorts (Fig. 4).

Comparing genomic changes and expression levels

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

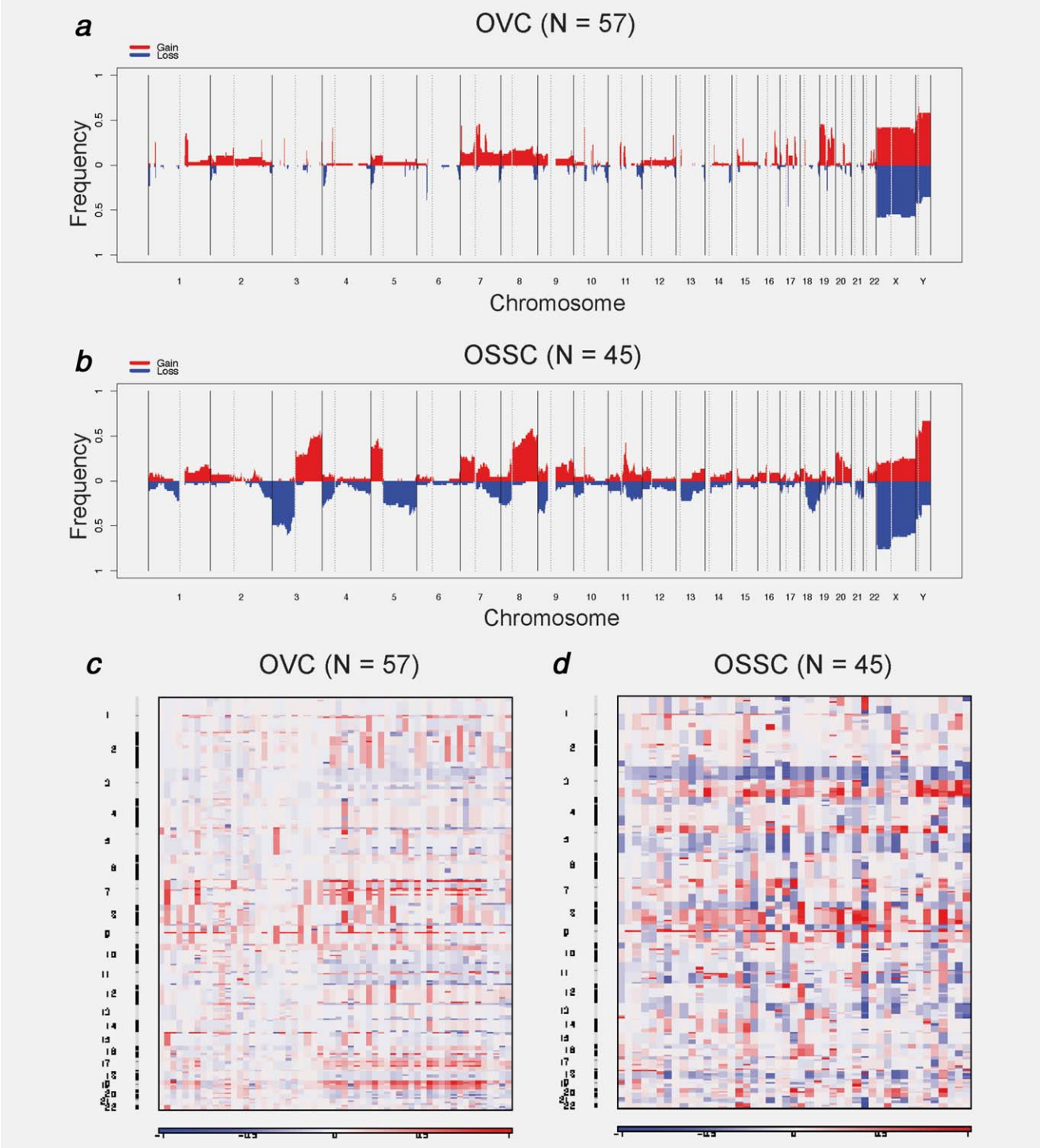

Figure 3. Frequency of genomic gain and loss for OVC (*a*) and OSCC (*b*). Genomic position is on the *x*-axis, frequency (%) of gains (red) and losses (blue) are shown on the *y*-axis. Heat map images of OVC (*c*) and OSCC (*d*) based on total segmented DNA copy number variation profiles. Heat map images 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.

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, they may be relatively smooth. Furthermore, OSCC can have a superficial verrucous appearance,

Table 1. Frequently occurring mutations in OVC and OSCC cohorts

Genes	Mutated in OSCC	Percentage	Mutated in OVC	Percentage
TP53	Yes	70	No	
CDKN2A	Yes	35	No	
NOTCH1	Yes	10	No	
NOTCH2	Yes	25	No	
FAT1	Yes	20	No	
DSPP	Yes	5	Yes	40
MUC4	Yes	15	Yes	30
NEFH	Yes	20	Yes	30
ANP32E	Yes	5	Yes	30

thus morphological features alone can lead to misclassification with consequences for patient management that may 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 CN patterns, somatic gene mutations and transcriptional changes that occur in OVC and compare them with the CN, 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.[27,28] 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,[26] 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.[29] We used logistic regression analysis to obtain statistical quantitation of the CN differences, correctly identifying 56 of 57 OVC samples and 41 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[21,22] 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 subgroups 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 upregulation of *DSPP* along with other genes in histologically aggressive and poorly differentiated OSCC and in some oral epithelial dysplasia.[30] Other studies have demonstrated a positive correlation between *MUC4* expression and tumour growth and malignant progression in pancreatic lesions.[31] *NEFH* was downregulated along with other genes in metastatic lung squamous cell carcinoma samples when compared with non-metastatic samples.[32] *ANP32E* expression was upregulated in gastric cancer cells when compared with non-neoplastic epithelial cells.[33] 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.[34] 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.[23,24] 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.[35] 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.[36] Of the 167 DEGs reported, 39 were shared between OSCC and OVC, and eight (*HLF*, *TGFBI*, *SERPINE1*, *MMP1*, *INHBA*, *COL4A2*, *COL4A1* and *ADAMTS12*) were

Table 2. Significant differential expression list for protein-coding genes in oral verrucous samples *versus* OSCCs

Gene name	Gene function	Log FC	P.adj
UNC45B	Myoblast fusion	5.73176541	1.02E−09
ANKRD30BL	–	5.068125637	3.00E−09
KRT2	Keratin gene, involved in differentiation	3.28887873	4.25E−07
DLG2	Guanylate kinase	3.142861272	3.25E−06
RP11-181C3.1	–	2.589914427	4.46E−06
KRT76	Keratin gene, structural integrity of epithelial cells	4.393647354	4.37E−05
LOR	Keratinocyte cell envelope protein	3.696791035	0.000182642
ELOVL4	Elongates fatty acids	1.913915467	0.000341653
HTRA3	Probable serine protease	−1.624779995	0.000341653
PDK4	Glucose metabolism	−2.442699908	0.000497839
FLG2	Filaggrin family	4.155257725	0.000682528
MT2A	Binds heavy metals	−2.067002668	0.000737808
HPGD	Prostoglandin inactivation. Inhibits proliferation in colon cancer	1.75292207	0.00078544
TNFRSF12A	Induces apoptosis, promotes angiogenesis, may modulate cell adhesion to matrix proteins	−1.923891955	0.000790342
IGFBP6	Affects growth-promoting effects of insulin-like growth factors	−2.386076713	0.000966215
ATP10B	–	1.718836376	0.001069055
HBB	Affects blood pressure	3.554687453	0.001114319
SLC11A1	Iron metabolism, pathogen resistance	−1.677099745	0.001188788
FDCSP	Binds to surface of B-lymphoma cells	3.705508687	0.001313018
FSTL3	Bone formation, differentiation of haemopoietic progenitor cells	−1.641022383	0.001395803
PTGDR2	Prostaglandin receptor	2.991133602	0.001395803
DSC1	Keratinisation of epithelial tissue	2.515586707	0.001634268
TCAP	Muscle assembly	−4.437644385	0.001634268
DDIT4	Promotes neuronal cell death	−2.274816537	0.002274577
INPP5F	Decreases AKT and GSK3B phosphorylation	2.093068765	0.002500484
PID1	–	1.742685358	0.002934709
ENTPD3	ATD/ADP hydrolysis	1.037851739	0.002934709
PLA2G4D	Inflammation	2.117301287	0.003105351
CTC-236F12.4	–	1.109388981	0.003105351
PDLIM3	Actin organisation	−2.03957167	0.003434804
C10orf53	–	2.290929173	0.004033844
THBS1	Cell–cell and cell–matrix interactions	−2.189567433	0.005454379
CXCL5	Neutrophil activation	−2.570674984	0.006168571
SERPINH1	Collagen binding	−1.43209521	0.007136242
MT1X	Binds heavy metals	−1.54913047	0.007136242
WNT9A	Probable developmental protein	−1.802040084	0.007747887
LCE1A	Precursor of stratum corneum	2.964477071	0.008599762
LAMC2	Migration, attachment and organisation of cells in embryonic development	−2.28993045	0.008730485
CYC1	Mitochondrial activity	−1.369128527	0.008958146
SMR3B	Androgen-regulated protein	−3.661364151	0.009270752
HLF	Member of PAR family	1.349932377	0.009352683
CTGF	Cell proliferation, differentiation and adhesion	−1.972202278	0.009409324

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.

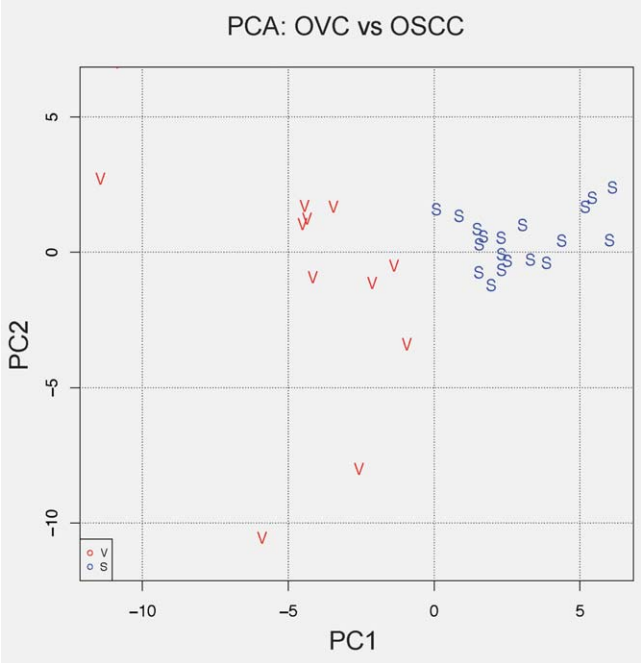

Figure 4. PCA biplot of PCs 1 and 2 using all expressed genes. This biplot best separates the two oral tumour groups: V: verrucous carcinoma; S: squamous cell carcinoma.

significantly differentially expressed between the two groups. Seven of the eight were upregulated in OSCC compared to OVC, with only *HLF* being comparatively upregulated in OVC. For all eight significantly DEGs 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 upregulated in OVC compared to OSCC. *HLF* is thought to be tumour suppressor gene that plays a role in the detoxification processes.[37]

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.[38–43] These were overexpressed in the OSCC cohort compared to the OVC samples. In contrast, downregulation of the *DLG2*, *HBB* and *DSC1* genes has been associated with progression.[44–46] These all were expressed more highly in the OVC samples. Similarly, *KRT76* and *KRT2* have been reported to be downregulated in OSCC[47] and were again more highly expressed in our OVC samples.

Amongst genes associated with metastasis, downregulation of *HPGD* has been shown to be a marker for metastasis.[48] This gene is more highly expressed in the OVC cohort. In contrast, overexpression of *CXCL5*, *THBS1*, *MT2A* and *MT1X* has been linked to metastasis.[49–51] 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 DEGs and mutated genes are located in the plasma membrane, participated in cell adhesions and keratinocyte differentiation and implemented in calcium ion (Ca^{2+}) binding. On the other hand, the OSCC data revealed that the majority of the upregulated 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 OSCCs. We used the three independent, complementary techniques of CN 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.[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.

Acknowledgements

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

1. Brodeur GM, Seeger RC, Schwab M, et al. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* 1984;224:1121–4.
2. Wood HM, Belvedere O, Conway C, et al. Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens. *Nucleic Acids Res* 2010;38:e151.
3. Dawson SJ, Rueda OM, Aparicio S, et al. A new genome-driven integrated classification of breast

cancer and its implications. *EMBO J* 2013;32:617–28.
4. Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 2014;158:929–44.
5. Qaisi M, Vorrasi J, Lubek J, et al. Multiple primary squamous cell carcinomas of the oral cavity. *J Oral Maxillofac Surg* 2014;72:1511–16.
6. Barnes L, Eveson JW, Reichart P, et al. WHO classification of tumours, pathology and genetics of head and neck tumours. Lyon: IARC Press, 2005. 174–5.
7. Ackerman LV. Verrucous carcinoma of the oral cavity. *Surgery* 1948;23:670–8.
8. Rekha KP, Angadi PV. Verrucous carcinoma of the oral cavity: a clinico-pathologic appraisal of 133 cases in Indians. *Oral Maxillofac Surg* 2010;14:211–18.
9. Samman M, Wood H, Conway C, et al. Next-generation sequencing analysis for detecting human papillomavirus in oral verrucous carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2014;118:117–25.e1.
10. Conway C, Chalkley R, High A, et al. Next-generation sequencing for simultaneous determination of human papillomavirus load, subtype, and associated genomic copy number changes in tumors. *J Mol Diagn* 2012;14:104–11.
11. Genomes Project Consortium; Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–73.
12. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJ* 2011;17:10–12.
13. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:60.
14. Gusnanto A, Wood HM, Pawitan Y, et al. Correcting for cancer genome size and tumour cell content enables better estimation of copy number alterations from next-generation sequence data. *Bioinformatics* 2012;28:40–7.
15. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011;12.R41:1–14.
16. Gusnanto A, Tcherveniakov P, Shuweihdi F, et al. Stratifying tumour subtypes based on copy number alteration profiles using next-generation sequence data. *Bioinformatics* 2015 doi: 10.1093/bioinformatics/btv191.
17. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
18. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;22:568–76.
19. McLaren W, Pritchard B, Rios D, et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 2010;26:2069–70.
20. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2008;4:44–57.
21. Liu CJ, Lin SC, Chen YJ, et al. Array-comparative genomic hybridization to detect genomewide changes in microdissected primary and metastatic oral squamous cell carcinomas. *Mol Carcinogen* 2006;45:721–31.
22. Jarvinen AK, Autio R, Kilpinen S, et al. High-resolution copy number and gene expression microarray analyses of head and neck squamous cell carcinoma cell lines of tongue and larynx. *Gene Chromosome Cancer* 2008;47:500–9.
23. Agrawal N, Frederick MJ, Pickering CR, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* 2011;333:1154–7.
24. Stransky N, Egloff AM, Tward AD, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science* 2011;333:1157–60.
25. Devaney KO, Ferlito A, Rinaldo A, et al. Verrucous carcinoma (carcinoma cuniculatum) of the head and neck: what do we know now that we did not know a decade ago? *Eur Arch Otorhinolaryngol* 2011;268:477–80.
26. Pentenero M, Donadini A, Di Nallo E, et al. Distinctive chromosomal instability patterns in oral verrucous and squamous cell carcinomas detected by high-resolution DNA flow cytometry. *Cancer* 2011;117:5052–7.
27. Lin HP, Wang YP, Chiang CP. Expression of p53, MDM2, p21, heat shock protein 70, and HPV 16/18 E6 proteins in oral verrucous carcinoma and oral verrucous hyperplasia. *Head Neck* 2011;33:334–40.
28. Mohtasham N, Babakoohi S, Shiva A, et al. Immunohistochemical study of p53, Ki-67, MMP-2 and MMP-9 expression at invasive front of squamous cell and verrucous carcinoma in oral cavity. *Pathol Res Pract* 2013;209:110–14.
29. Zargaran M, Eshghyar N, Baghaei F, et al. Assessment of cellular proliferation in oral verrucous carcinoma and well-differentiated oral squamous cell carcinoma using Ki67: a non-reliable factor for differential diagnosis? *Asian Pac J Cancer* 2012;13:5811–5.
30. Ogbureke KUE, Abdelsayed RA, Kushner H, et al. Two members of the SIBLING family of proteins, DSPP and BSP, may predict the transition of oral epithelial dysplasia to oral squamous cell carcinoma. *Cancer* 2010;116:1709–17.
31. Chaturvedi P, Singh AP, Moniaux N, et al. MUC4 mucin potentiates pancreatic tumor cell proliferation, survival, and invasive properties and interferes with its interaction to extracellular matrix proteins. *Mol Cancer Res* 2007;5:309–20.
32. Wang N, Zhou F, Xiong H, et al. Screening and identification of distant metastasis-related differentially expressed genes in human squamous cell lung carcinoma. *Anat Rec* 2012;295:748–57.
33. Tsukamoto Y, Uchida T, Karnan S, et al. Genome-wide analysis of DNA copy number alterations and gene expression in gastric cancer. *J Pathol* 2008;216:471–82.
34. Hua S, Kallen CB, Dhar R, et al. Genomic analysis of estrogen cascade reveals histone variant H2A.Z associated with breast cancer progression. *Mol Syst Biol* 2008;4:1–14.
35. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. *Nat Rev Cancer* 2011;11:9–22.
36. Wang YH, Tian X, Liu OS, et al. Gene profiling analysis for patients with oral verrucous carcinoma and oral squamous cell carcinoma. *Int J Clin Exp Med* 2014;7:1845–52.
37. Woenckhaus M, Klein-Hitpass L, Grepmeier U, et al. Smoking and cancer-related gene expression in bronchial epithelium and non-small-cell lung cancers. *J Pathol* 2006;210:192–204.
38. Lindberg P, Larsson A, Nielsen BS. Expression of plasminogen activator inhibitor-1, urokinase receptor and laminin gamma-2 chain is an early coordinated event in incipient oral squamous cell carcinoma. *Int J Cancer* 2006;118:2948–56.
39. Pyeon D, Newton NA, Lambert PF, et al. Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. *Cancer Res* 2007;67:4605–19.
40. Cacalano N, Le D, Paranjpe A, et al. Regulation of IGFBP6 gene and protein is mediated by the inverse expression and function of c-jun N-terminal kinase (JNK) and NF kappa B in a model of oral tumor cells. *Apoptosis* 2008;13:1439–49.
41. Freier K, Knoepfle K, Flechtenmacher C, et al. Recurrent copy number gain of transcription factor SOX2 and corresponding high protein expression in oral squamous cell carcinoma. *Gene Chromosome Cancer* 2010;49:9–16.
42. Hassan NMM, Hamada J, Kameyama T, et al. Increased expression of the PRL-3 gene in human oral squamous cell carcinoma and dysplasia tissues. *Asian Pac J Cancer* 2011;12:947–51.
43. Ambatipudi S, Gerstung M, Pandey M, et al. Genome-wide expression and copy number analysis identifies driver genes in gingivobuccal cancers. *Gene Chromosome Cancer* 2012;51:161–73.
44. Onda M, Akaishi J, Asaka S, et al. Decreased expression of haemoglobin beta (HBB) gene in anaplastic thyroid cancer and recovery of its expression inhibits cell growth. *Br J Cancer* 2005;92:2216–24.
45. Reshmi SC, Huang X, Schoppy DW, et al. Relationship between FRA11F and 11q13 gene amplification in oral cancer. *Gene Chromosome Cancer* 2007;46:143–54.
46. Teh MT, Parkinson EK, Thurlow JK, et al. A molecular study of desmosomes identifies a desmoglein isoform switch in head and neck squamous cell carcinoma. *J Oral Pathol Med* 2011;40:67–76.
47. Zain RB, Anwar A, Karen-Ng LP, et al. Selected transcriptome profiles of oral cancer suggestive of field cancerisation using second generation sequencing. *Oral Dis* 2010;16:521.
48. Kawamata H, Furihata T, Omotehara F, et al. Identification of genes differentially expressed in a newly isolated human metastasizing esophageal cancer cell line, T.Tn-AT1, by cDNA microarray. *Cancer Sci* 2003;94:699–706.
49. Miyazaki H, Patel V, Wang HX, et al. Growth factor-sensitive molecular targets identified in primary and metastatic head and neck squamous cell carcinoma using microarray analysis. *Oral Oncol* 2006;42:240–56.
50. Mendez E, Houck JR, Doody DR, et al. A genetic expression profile associated with oral cancer identifies a group of patients at high risk of poor survival. *Clin Cancer Res* 2009;15:1353–61.
51. Skubitz KM, Francis P, Skubitz APN, et al. Gene expression identifies heterogeneity of metastatic propensity in high-grade soft tissue sarcomas. *Cancer* 2012;118:4235–43.