Clinical relevance of breast cancer-related genes as potential biomarkers for oral squamous cell carcinoma

Toshima Z. Parris1,†,*
* Corresponding author
Email: toshima.parris@oncology.gu.se

Luaay Aziz2,†,*
* Corresponding author
Email: luaay.aziz@vgregion.se

Anikó Kovács3
Email: aniko.kovacs@vgregion.se

Shahin Hajizadeh3
Email: shahin.hajizadeh@vgregion.se

Szilárd Nemes4,5
Email: Szilard.Nemes@registercentrum.se

May Semaan1
Email: may.semaan@oncology.gu.se

Chang Yan Chen1,6
Email: chang.yan.chen@gu.se, cchen6@bidmc.harvard.edu

Per Karlsson1
Email: per.karlsson@oncology.gu.se

Khalil Helou1
Email: khalil.helou@oncology.gu.se

†These two authors contributed equally to the work.

1Department of Oncology, Institute of Clinical Sciences, Sahlgrenska Cancer Center, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

2Department of Otolaryngology, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
Running title: Prognostic markers for OSCC

Keywords: oral squamous cell carcinoma, outcome prediction, molecular biomarker, immunohistochemistry, model validation

Word count: 3,997 words (from Background to Conclusions)

Corresponding authors: Toshima Z. Parris, University of Gothenburg, Department of Oncology, Sahlgrenska Cancer Center, Medicinaregatan 1G, SE-41345 Gothenburg, Sweden. Phone: 46-31-7866751; Fax: 46-31-820114; E-mail: toshima.parris@oncology.gu.se. Luaay Aziz, University of Gothenburg, Department of Otolaryngology, Gröna stråket 9, SE-41345 Gothenburg, Sweden. Phone: 46-31-3429128; Fax: 46-31-413258; E-mail: luaay.aziz@vgregion.se.

Grant support: This work was supported by grants from the Swedish Cancer Society (KH), the King Gustav V Jubilee Clinic Cancer Research Foundation (KH), Assar Gabrielsson Research Foundation (LA), and the Wilhelm and Martina Lundgren Research Foundation (TZP and LA).
Abstract

Background: Squamous cell carcinoma of the oral cavity (OSCC) is a common cancer form with relatively low 5-year survival rates, due partially to late detection and lack of complementary molecular markers as targets for treatment. Molecular profiling of head and neck cancer has revealed biological similarities with basal-like breast and lung carcinoma. Recently, we showed that 16 genes were consistently altered in invasive breast tumors displaying varying degrees of aggressiveness.

Methods: To extend our findings from breast cancer to another cancer type with similar characteristics, we performed an integrative analysis of transcriptomic and proteomic data to evaluate the prognostic significance of the 16 putative breast cancer-related biomarkers in OSCC using independent microarray datasets and immunohistochemistry. Predictive models for disease-specific (DSS) and/or overall survival (OS) were calculated for each marker using Cox proportional hazards models.

Results: We found that CBX2, SCUBE2, and STK32B protein expression were associated with important clinicopathological features for OSCC (peritumoral inflammatory infiltration, metastatic spread to the cervical lymph nodes, and tumor size). Consequently, SCUBE2 and STK32B are involved in the hedgehog signaling pathway which plays a pivotal role in metastasis and angiogenesis in cancer. In addition, CNTNAP2 and S100A8 protein expression were correlated with DSS and OS, respectively.

Conclusions: Taken together, these candidates and the hedgehog signaling pathway may be putative targets for drug development and clinical management of OSCC patients.
Background

Oral squamous cell carcinoma (OSCC) is the most common malignancy form in the head and neck region, accounting for about 260,000 new cases and 124,000 OSCC-related deaths worldwide annually [1, 2]. In western countries, the etiology of some newly diagnosed primary SCCs of the head and neck has shifted from tobacco and alcohol abuse to human papillomavirus (HPV) infections, possibly as a result of changes in sexual practices [3-8]. Despite aggressive treatment modalities, 5-year survival rates for advanced head and neck cancers have remained low and relatively unchanged (about 50-60%) for several decades, partially due to early locoregional recurrences within 2 years of initial treatment [9]. There is therefore a pressing need for molecular predictors that enable earlier detection of the disease, describe tumor behavior, and improve risk assessment to identify patients at risk for recurrence and OSCC-related death.

Molecular profiling has become a common and effective method for cancer gene discovery and classification of cancer. Almost a decade ago, Chung et al. identified four intrinsic subtypes for head and neck squamous cell carcinoma with clinical and biological implications [10]. Consequently, one of the subtypes with the most unfavorable prognosis also displayed strikingly similar transcriptional patterns with the breast carcinoma basal-like phenotype and lung squamous cell carcinoma. Recently, additional evidence of shared cellular processes was found between breast carcinoma and oral squamous cell carcinoma, i.e. mechanisms for tumor lymphangiogenesis and metastasis to the regional lymph nodes as well as HER2/neu polymorphisms [11, 12]. These findings suggest that cancers derived from different sites of origin may perturb common signaling pathways and thereby display similar tumor characteristics [13].

To test this hypothesis, we evaluated the prognostic potential of 16 putative prognostic biomarkers (AZGP1, BTG2, CBX2, CNTNAP2, DNALI1, LOC389033, NME5, PIP, S100A8, SCUBE2,
SERPINA11, STC2, STK32B, SUSD3, UBE2C, and WHSC1L1) previously identified in breast carcinoma using oral squamous cell carcinoma [14, 15]. Interestingly, several of the putative biomarkers have been implicated in the carcinogenesis of more than one cancer form. We, and others, have also been able to show that AZGP1, S100A8, and STK32B, as well as CNTNAP2 are associated with the basal-like phenotype and lymph node metastasis, respectively [15-17].

Here, we investigated the prognostic potential of the gene expression signature in relation to clinical outcome, disease-specific survival (DSS) and/or overall survival (OS), in two steps. First, transcriptional levels for each gene were evaluated with respect to the clinical endpoints using publicly available Affymetrix one-channel microarray (n = 168) and Illumina RNASeq datasets (n = 198) for OSCC from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) repositories, respectively. Second, because correlation between mRNA/protein levels is frequently low, Cox proportional hazards models for DSS and OS were also calculated using immunohistochemical protein expression patterns from 43 OSCC cases together with established clinicopathological features (tumor size and node status or tumor size and age, respectively).

Methods

Patient cohorts

To evaluate the prognostic potential of the AZGP1, BTG2, CBX2, CNTNAP2, DNALI1, LOC389033, NME5, PIP, S100A8, SCUBE2, SERPINA11, STC2, SUSD3, STK32B, UBE2C, and WHSC1L1 genes in OSCC specimens, three patient cohorts were compiled primarily from squamous cell carcinomas of the oral cavity. External gene expression datasets and corresponding clinical information for Cohorts I-II were compiled from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) repositories, respectively. Cohort I included two Affymetrix U133 Plus 2.0 GeneChip
datasets (GEO accession numbers GSE41613 and GSE42743) containing 168 OSCC samples (oropharynx samples were excluded from the analysis) [18]. Cohort II consisted of normalized RNAseq by Expectation-Maximization (RSEM) gene datasets from 198 OSCC patients (oral cavity: buccal mucosa, floor of mouth, tongue), which were downloaded from The Broad Institute TCGA GDAC (http://gdac.broadinstitute.org/runs/stddata__2014_01_15/). Cohort III consisted of 43 OSCC cases originating from the oral cavity (buccal gingiva, floor of mouth, tongue), which had been diagnosed between 1997-2004 at Sahlgrenska University Hospital in Gothenburg, Sweden. All patients underwent diagnostic battery inclusive biopsy of the primary tumor, palpation of the neck, radiological examination with MRT and/or CT, and TNM classified according to the American Joint Committee on Cancer (AJCC) staging system. Surgical excision of the primary tumor and supraomohyoid neck dissection (SOHND) were performed. In total, 16 patients had cervical lymph node metastases (pN1) of which 5/16 patients had micrometastases (pNmic) as assessed using anti-human monoclonal cytokeratin AE1/AE3, and 27 patients were lymph node-negative (pN0). Lymph node-positive patients received post-operative radiotherapy to the neck, whereas pN0 and pNmic patients were followed up clinically. All patients were followed up for at least five years during which seven patients (16%) developed local and/or regional recurrence, including two patients with pN1, three patients with pNmic, and two patients with pN0 disease. Three of the five patients with micrometastases developed recurrence (60%), of which two (40%) died within three years due to OSCC-related causes. The clinicopathological features for Cohorts I-III are summarized in Table 1.

**Immunohistochemistry**

For Cohort III, 45 FFPE samples corresponding to the 43 patients were obtained from the Department of Pathology at Sahlgrenska University Hospital and used in immunohistochemistry.
experiments in accordance with the Declaration of Helsinki and approved by the Medical Faculty Research Ethics Committee (Gothenburg, Sweden). The ethics committee approved a waiver of written consent to use the tumor specimens in the study. Histological classification and TNM staging of the tumor specimens were performed according to the WHO classification and International Union Against Cancer (UICC), respectively [19, 20]. Optimal antibody dilutions and assay conditions were achieved for immunohistochemistry using OSCC as positive controls. Four micrometer full-face FFPE sections were pretreated using the Dako PTLink system (Dako, Carpinteria, CA, USA) and processed using the Dako Envision™ FLEX High pH Link Kit (pH 9) for p16, AZGP1, BTG2, CBX2, CNTNAP2, NME5, S100A8, SCUBE2, SERPINA11, STC2, SUSD3, STK32B, SUSD3, UBE2C, and WHSC1L1 as listed in Additional file 1: Table S1. Peroxidase-catalyzed diaminobenzidine was used as the chromogen, followed by hematoxylin counterstain. The slides were then rinsed with deionized water, dehydrated in absolute alcohol, followed by 95% alcohol, cleared in xylene, and mounted. H&E staining was performed on one FFPE section to facilitate histological assessment. The degree of lymphoplasmacytic infiltration (inflammatory infiltration) was classified as minimal (few inflammatory cells), moderate (1-2 mm margin), and strong (>2 mm margin) according to the density of inflammatory cells. Immunostaining was evaluated by a head and neck pathologist, blinded to patient clinical outcome, and scored as previously described using the semi-quantitative H-score method to calculate the sum of the percentage and intensity of positively stained tumor cells within the invasive tissue component (negative staining = 0; weak staining = 1+; moderate staining = 2+; strong staining = 3+). The H-score ranged from 0 to 300, where H-score = (1 x %1+) + (2 x %2+) + (3 x %3+) [21]. The X-tile software (version 3.6.1) was used to determine an H-score cut-off for positive staining by dichotomizing patients according to H-score value and clinical outcome, as listed in Additional file 1: Table S1 [22]. FFPE specimens lacking an invasive tissue component were removed from the analysis. Each tumor specimen was scored once, where multiple FFPE sections representing the same tumor were
Statistical Analysis

Statistical analyses were performed using a 0.05 P-value cutoff (two-sided) in R/Bioconductor (version 2.15.0). Putative prognostic biomarkers for OSCC were identified in two steps. First, the prognostic potential of aberrant biomarker gene expression was evaluated in external microarray and RNASeq datasets (Cohorts I-II). Then, predictive models for DSS and OS were developed using biomarker protein expression (Cohort III).

A) Evaluation of gene expression patterns for the 16-marker signature in external microarray and RNASeq datasets. Univariate Cox proportional hazard models were calculated for each gene using the endpoints disease specific-survival (DSS) and/or overall survival (OS). OSCC survival rates were defined as a) the period from initial diagnosis to OSCC-related death for DSS and b) period from initial diagnosis to death from any cause for OS. Data processing and Cox regression analysis of the Affymetrix one-channel microarray datasets (Cohort I) and normalized RNASeq RSEM values (Cohort II) were performed using Nexus Expression 3.0 (BioDiscovery).

B) Development of a predictive model for DSS and OS using protein expression. Survival rates (DSS and OS) at different protein expression levels were depicted with Kaplan-Meier curves and tested with log-rank test. The relationship between clinicopathological features and protein expression was evaluated using two-tailed Fisher’s exact test. Multivariate analysis was conducted using the Cox proportional hazard model for DSS or OS with stepwise selection to assess the predictive strength and additive accuracy of protein expression after
adjusting for established clinicopathological features (tumor size and node status or tumor size and age, respectively). A concordance index (C-index) for the time-dependent area under the ROC curve (AUC(t)) was calculated to assess model predictive performance, varying from C-index = 0.5 (no predictive power) to C-index = 1 (perfect prediction).

Results

Prognostic potential of the molecular biomarkers in external gene expression microarray and RNASeq datasets

In previous work, we showed the clinical significance of 16 candidate molecular biomarkers (AZGP1, BTG2, CBX2, CNTNAP2, DNALI1, LOC389033, NME5, PIP, S100A8, SCUBE2, SERPINA11, STC2, STK32B, SUSD3, UBE2C, and WHSC1L1) in invasive breast carcinoma [14, 15, 23]. To investigate whether these putative prognostic biomarkers may also play a pivotal role in the aggressive nature of OSCCs, the effect of altered gene expression patterns on clinical outcome was evaluated using two external OSCC patient cohorts (Cohorts I-II; Table 1). Cox proportional hazard models were calculated for each gene with relation to clinical endpoints (DSS and/or OS; Table 2).

In Cohort I, two genes (LOC389033 and SERPINA11) were not found on the Affymetrix platform and therefore excluded from the analysis. Univariate Cox regression analysis showed that low levels of AZGP1 \( (P = 0.001) \), BTG2 \( (P = 0.020) \), PIP \( (P = 0.010) \), and SCUBE2 \( (P = 0.026) \) were indicative of a more unfavorable prognosis, whereas elevated levels of STC2 \( (P = 0.001) \) and UBE2C \( (P = 0.042) \) had an adverse effect on DSS. For OS, low levels of BTG2 \( (P = 0.005) \) and
elevated levels of \textit{CBX2} ($P = 0.005$), \textit{STC2} ($P = 0.001$), and \textit{STK32B} ($P = 0.015$) were predictive of outcome. In addition, low \textit{S100A8} mRNA levels ($P = 0.053$) were borderline significant for OS.

For Cohort II, the 16-gene signature was evaluated in RNASeq expression profiling data for 198 OSCC patients. Univariate Cox regression analysis showed that elevated levels of \textit{UBE2C} mRNA levels ($P = 0.005$) were indicative of OS. On the other hand, elevated levels of \textit{NMES} ($P = 0.050$), low \textit{S100A8} levels ($P = 0.072$), and low \textit{SCUBE2} levels ($P = 0.067$) were borderline significant for OS. Furthermore, low \textit{S100A8} levels ($P < 0.001$; log$_2$ ratio $= -2.61$) and elevated \textit{UBE2C} levels ($P = 0.007$; log$_2$ ratio $= 0.713$) were significantly associated with high histological grade.

\textbf{Protein expression levels of the molecular biomarkers in OSCC specimens}

Protein expression levels for the candidate biomarkers were evaluated using immunohistochemistry with 45 full-face FFPE specimens representing 43 OSCC patients (Cohort III; Table 1). PIP and DNALI1 were excluded from further analysis due to low expression levels in OSCC samples, whereas LOC389033 was excluded because the gene is not expressed at the protein level. Immunopositivity was shown for all of the examined proteins in peritumoral normal mucous membrane, the salivary glands, and dysplasia, with the exception of AZGP1, SUSD3, UBE2C, and WHSC1L1. AZGP1 and UBE2C were strongly positive in the basal cell layer; SUSD3 was positive in the mucous membrane but negative in the salivary glands, whereas WHSC1L1-positivity was shown in the layers of muscle tissue. In addition, no \textit{S100A8} staining was observed in the basal cell layer. In invasive tissue, immunopositivity for the 14 analyzed proteins ranged from 14-86\% with CNTNAP2 and WHSC1L1 having the lowest and highest incidence rates, respectively (Table 3). Interestingly, there was only one reported case (6\%) of SCUBE2-positivity in
lymph node-positive tumors, compared with SCUBE2-positivity in 41% of lymph node-negative tumors. In addition, p16 immunopositivity was observed in 8/43 tumor specimens (19%).

**Correlation of the molecular biomarkers with clinicopathological features**

To investigate whether heterogeneous protein expression of the analyzed antigens is clinically relevant, a correlation analysis was performed with established clinicopathological features (Table 4). S100A8 was strongly associated with tumor differentiation ($P = 0.009$), e.g. tumors with enhanced S100A8 expression levels were frequently well differentiated (64%) compared with 17% in S100A8-negative tumors. In addition, SCUBE2 was significantly associated with lymph node status ($P = 0.01$), CBX2 with tumor inflammatory infiltration ($P = 0.03$), and STK32B with tumor size ($P = 0.04$). Interestingly, a high proportion of SCUBE2-positive tumors (11/12) were lymph node-negative and all STK32B-positive tumors (15/15) were smaller in size (T1-T2 tumors); minimal peritumoral inflammatory infiltration was found in tumors with reduced CBX2 levels. Additionally, we also found a slight indication that UBE2C and SCUBE2, SERPINA11, and NME5 were associated with tumor differentiation ($P = 0.06$ and $P = 0.09$, respectively), inflammatory infiltration ($P = 0.09$), and p16 expression ($P = 0.08$), respectively.

**Prognostic significance of the molecular biomarkers**

Next, we examined the prognostic significance of the proposed biomarkers using disease-specific survival and overall survival. OSCC patients with tumors displaying enhanced CNTNAP2 levels had significantly shorter DSS ($P = 0.010$; HR (95% CI) = 5.70 (1.27-25.57)), whereas patients with S100A8-negative tumors had significantly shorter OS ($P = 0.0063$; HR (95% CI) = 0.10 (0.014-0.76); Figure 1). Our data suggest a slight association between SCUBE2 expression and DSS ($P = 0.090$),
as well as UBE2C expression and OS ($P = 0.074$). CNTNAP2 expression had no significant effect on DSS after adjusting for tumor size and lymph node status ($P = 0.10$ - HR (95% CI) = 3.56 (0.78-16.17)). Furthermore, outcome prediction was not improved using a predictive model for DSS including CNTNAP2 expression, lymph node status, and tumor size (C-index = 0.949) compared with a model containing lymph node status and tumor size (C-index = 0.941; Figure 2). Following multivariate analysis adjusting for tumor size, lymph node status, differentiation and age, S100A8 was still statistically significant ($P = 0.013$ - HR (95% CI) = 0.11 (0.013-0.92; Table 5)). Combining S100A8 in a predictive model for OS with tumor size, lymph node status, differentiation and age improved outcome prediction significantly from 0.605 to 0.833 (Figure 2).

**Discussion**

Oral squamous cell carcinoma is a heterogeneous disease with diverse clinical, pathological, and biological behavior [10]. Nevertheless, the strongest determinants of prognosis still include tumor stage and the presence of cervical metastases at the time of diagnosis, as well as the time to locoregional recurrences [9, 24-26]. Unfortunately, up to 50% of OSCCs are diagnosed at an advanced stage with 5-year survival rates at approximately 60%, e.g. delayed diagnosis [27-30]. Therefore, many patients could benefit greatly from complementary molecular markers, which may help guide treatment decisions and be of value in the development of new therapeutic agents. Extensive efforts are currently being made to identify and validate biomarkers based on the biology of oral cancers that can complement established clinicopathological features and improve clinical management of the disease. Recent work to characterize OSCC using transcriptomic profiling has mainly focused on the identification of biomarkers for disease progression and lymph node metastasis prediction [31-38]. Surprisingly, few gene expression
signatures have been developed to improve patient risk assessment [18, 34, 39, 40]. Although transcriptome analyses can give an indication of biological activity within a tissue, mRNA and protein levels may not correlate because gene expression is controlled by a multistage system. We have therefore performed immunohistochemistry using readily available FFPE samples to evaluate the clinical significance of the 16 putative biomarkers in OSCC. Here, we show the prognostic potential of the S100A8 and CNTNAP2 proteins, as well as the relationship between SCUBE2, CBX2, and STK32B protein levels and important clinicopathological features for OSCC, i.e. regional metastasis to the cervical lymph nodes, tumor inflammatory infiltration, and tumor size, respectively.

In the present work, we evaluated the applicability of breast cancer prognostic biomarkers for OSCC, given the biological similarities between the two cancer types [10]. In breast carcinoma, we showed the recurrent upregulation of the CBX2, CNTNAP2, S100A8, UBE2C, and WHSC1L1 genes as well as downregulation of the AZGP1, BTG2, DNALI1, LOC389033, NME5, PIP, SCUBE2, SERPINA11, STC2, STK32B, and SUSD3 genes in more aggressive tumors [14, 15, 23]. The prognostic potential of individual biomarkers in the 16-gene signature were evaluated in three OSCC patient cohorts containing clinical information, two of which (Cohorts I-II) were compiled from publicly available Affymetrix and Illumina RNASeq datasets. The Affymetrix one-channel system and Illumina RNAseq platform have recently been shown to correlate well, in particular for high abundance genes [41]. Although the majority of the markers in the signature were in agreement for both breast carcinoma and OSCC, five proteins (S100A8, STC2, STK32B, SUSD3, and WHSC1L1) were inversely regulated in aggressive OSCC samples. In addition, several of the biomarkers which showed promise at the mRNA level (CBX2, S100A8, SCUBE2, and STK32B) were also either predictive of clinical outcome or associated with clinicopathological features at the protein level. The differences in the prognostic potential of specific biomarkers at the mRNA or
protein levels may possibly be the result of small sample sizes, discordant mRNA-protein
eexpression patterns or differences in sample preparation for microarray analyses and
immunohistochemistry, i.e. microarray is frequently performed using a tumor mass, which
consists of both malignant and nonmalignant cells, resulting in an over- or underestimation of
expression levels, whereas cell type-specific protein expression patterns can be easily interpreted
using immunohistochemistry. In breast carcinoma, elevated levels of the S100A8 protein were
associated with moderate/strong tumor inflammatory infiltration and significantly shorter DSS
rates. In contrast, lower S100A8 protein levels were associated with significantly shorter OS rates
in OSCC, but there was no association with tumor infiltration. These findings suggest differences
in cell type-specific gene expression and thereby activation and/or inhibition of diverse cellular
mechanisms. Furthermore, CBX2, CNTNAP2, S100A8, SCUBE2, and STK32B protein levels were
significantly associated with established clinicopathological features for OSCC and/or patient
clinical outcome. We propose here two potential prognostic biomarkers for OSCC (the CNTNAP2
protein for DSS and the S100A8 protein for OS). There may be several reasons why these
candidate biomarkers were associated with different clinical endpoints: differences in gene
function and whether the biomarkers correlated with strong determinants of OSCC-related
prognosis, e.g. tumor stage, lymph node metastasis or time to locoregional recurrences.

Chen et al. identified upregulation of CNTNAP2 mRNA levels as a distinctive characteristic of OSCC
in comparison with normal mucosa [32]. In a genome-wide association study for oral cancers,
CNTNAP2 was found to be associated with cell migration [42]. The CNTNAP2 gene also functions
as a cell adhesion molecule and has been found to be either methylated or deleted in several
different cancer types, e.g. glioma, myeloid leukemia, and pancreatic adenocarcinoma [43-45]. In
the present study, elevated CNTNAP2 protein levels were prevalent in tumors from pN1 patients,
but this association was not statistically significant likely due to the low number of events. These
findings and the known function of the CNTNAP2 protein in tumorigenesis suggest that it may be a determinant of metastatic spread to the cervical lymph nodes and thereby a more aggressive phenotype with an adverse effect on DSS. Although elevated CNTNAP2 levels were predictive of a more unfavorable prognosis in univariate Cox regression models, disease-specific survival rates were not significantly improved by including the protein in a model containing lymph node status and tumor size. These results show the profound effect node metastases have on clinical outcome. In future studies, it will be necessary to alter the study design to test the prognostic potential of CNTNAP2 further, e.g. using a larger cohort containing node-negative patients.

Furthermore, the \textit{S100A8} gene, also known as MRP-8 or calgranulin A, has been studied extensively in head and neck cancer. Inclusion of S100A8 protein expression patterns in multivariate models together with tumor size, node status, and age significantly improved OS prediction. In addition, we found an association between low S100A8 levels (mRNA levels in Cohort II and protein levels in Cohort III) and tumor differentiation. S100A8 belongs to the S100 gene family, several of which form a gene cluster on chromosome 1q21 and are commonly induced by chronic and acute inflammation [46-48]. In the present study, positive S100A8 immunostaining was also observed in a peritumoral inflammatory cell population, albeit similar in both S100A8-positive and negative tumors. Overexpression of the S100A8 protein has been observed in breast, colorectal, gastric, lung, pancreatic, and prostate cancer, whereas underexpression has been shown in various squamous cell carcinomas of the head and neck [49]. Recently, the S100A8/A9 heterocomplex was identified as a regulator of cell cycle progression and cell proliferation in cancer cell lines originating from the head and neck region [50].

OSCC outcome prediction may also be improved by identifying biomarkers that can predict which tumors will inevitably spread to the cervical nodes. When we initiated this study, we believed that the \textit{CNTNAP2} gene would be a strong candidate for lymph node status because of its
association with metastatic spread to the axillary lymph nodes in breast carcinoma [23]. However in the present study, loss of SCUBE2 protein expression in the tumor epithelial component was the only marker significantly associated with lymph node status and borderline significant for DSS. Interestingly, all 12 patients with SCUBE2-positive tumors were long-term disease-specific survivors and only 1/12 tumors had metastasized to the cervical lymph nodes. These findings are consistent with reports in breast carcinoma where elevated levels of the SCUBE2 protein not only inhibited proliferation of breast cancer cell lines, but was indicative of a more favorable prognosis [23, 51]. SCUBE2 is a multidomain, secreted glycoprotein that was first implicated in functioning upstream of hedgehog signaling in zebrafish, but has since been shown to also play a pivotal role in human tumorigenesis [52, 53]. Furthermore, SCUBE2 serves two different functions by using either the CUB domain at the C-terminal end or the EGF-like repeats at the N-terminus, i.e. control of cell proliferation by antagonizing bone morphogenetic protein (BMP) activity and regulation of cell-cell contacts, and possibly tumor metastasis, by forming complexes with E-cadherin in adherens junctions, respectively [54].

The role CBX2 and STK32B play in OSCC tumorigenesis is largely unknown. In a correlation analysis between protein expression patterns and established clinicopathological features, a clear link was found between CBX2-positivity and tumor inflammatory infiltration as well as STK32B-positivity and small tumor size (T1-T2). CBX2 plays a role in epigenetic regulation and hematopoietic stem cell differentiation, whereas the STK32B gene is one of several serine/threonine kinases (STK32B, STK36, and STK39) with similar gene expression patterns in basal-like breast cancers [16, 55, 56]. Consequently, STK36 is a component of the hedgehog signaling pathway. The function of the STK32B gene is not known, but single nucleotide polymorphisms in the STK32B gene have been associated with oral clefts [57].
Hedgehog signaling has also been implicated in OSCC and basal cell carcinoma, a skin cancer frequently found in the head and neck region [58, 59]. Leovic and colleagues illustrated the clinical relevance and aberrant expression of key hedgehog signaling components in SCCs of the oral cavity and oropharynx, i.e. \textit{PTCH1}, \textit{GLI1}, \textit{SMO}, and \textit{SHH} [58]. In Cohort II, several genes which play a role in hedgehog signaling were also found to be associated with high histological grade (the \textit{KIF7}, \textit{TCTN1}, and \textit{TCTN2} genes) and metastatic spread to the cervical lymph nodes (the \textit{FGF4} and \textit{PTCHD1} genes). These findings further emphasize the importance of hedgehog signaling in OSCC.

**Conclusions**

In summary, our results illustrate that cancer is a complex, heterogeneous disease. However, certain clinically relevant biomarkers may be useful therapeutic targets for several different cancer types. Despite the limited number of patients included in the study, the integrative analysis identified at least five of the putative biomarkers originally developed in breast carcinoma (\textit{CBX2}, \textit{CNTNAP2}, \textit{S100A8}, \textit{SCUBE2}, and \textit{STK32B}) that also improve risk assessment for OSCC patients and may play a crucial role in cancer-related processes, e.g. hedgehog signaling, regulation of cell cycle progression, cell proliferation, metastasis. These targets should be studied in a larger cohort to further evaluate their clinical significance in OSCC.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

TZP, LA, AK, CYC, PK, KH contributed to the study design. TZP, SH, MS, AK performed the immunohistochemical analyses. TZP performed the data and statistical analyses, and drafted the manuscript. LA collected clinical data, performed the evaluation of lymph nodes, data analyses, and drafted the manuscript. AK performed the pathologic review of the FFPE samples. SN performed the statistical analyses. AK, SH, SN, MS, CYC, PK, KH critically reviewed the manuscript. All authors have read and approved the final manuscript.
Acknowledgments
The authors would like to thank Ulric Pedersen for his technical assistance.

Abbreviations
OSCC: squamous cell carcinoma of the oral cavity; DSS: disease-specific; OS: overall survival; pN1: cervical lymph node metastases; pN0: neck negative; pNmic: micrometastasis; H&E: hematoxylin and eosin; FFPE: formalin-fixed, paraffin-embedded tumor tissues; C-index: concordance index; AUC(t): time-dependent area under the ROC curve.

References


increased overall survival in breast cancer, as detected by immunohistochemistry using tissue microarray. BMC Cancer 2010, 10:296.


496 D, Pimentel N, Hoebers FJ, Rasch CR et al: Gene expression profiling to predict outcome after
498 41. Guo Y, Sheng Q, Li J, Ye F, Samuels DC, Shyr Y: Large scale comparison of gene expression
499 levels by microarrays and RNAseq using TCGA data. PLoS One 2013, 8(8):e71462.
500 42. Bhatnagar R, Dabholkar J, Saranath D: Genome-wide disease association study in chewing
502 43. Bralten LB, Gravendeel AM, Kloosterhof NK, Sacchetti A, Vrijenhoek T, Veltman JA, van den
503 Bent MJ, Kros JM, Hoogenraad CC, Sillevi Smitt PA et al: The CASPR2 cell adhesion molecule
506 Paepe A, Van Roy N et al: Identification of two critically deleted regions within chromosome
509 profiling of methylated promoters in pancreatic adenocarcinoma. Cancer Biol Ther 2008,
510 7(7):1146-1156.
511 46. Cross SS, Hamdy FC, Deloulme JC, Rehman I: Expression of S100 proteins in normal human
512 tissues and common cancers using tissue microarrays: S100A6, S100A8, S100A9 and S100A11
513 are all overexpressed in common cancers. Histopathology 2005, 46(3):256-269.
514 47. Odink K, Cerletti N, Bruggen J, Clerc RG, Tarcsay L, Zwadlo G, Gerhards G, Schlegel R, Sorg C:
515 Two calcium-binding proteins in infiltrate macrophages of rheumatoid arthritis. Nature 1987,
516 330(6143):80-82.
518 Durka S et al: The Ca2+-binding proteins S100A8 and S100A9 are encoded by novel injury-


Figures

Figure 1. Prognostic potential of CNTNAP2 and S100A8 protein expression in OSCC. (a-b) Kaplan-Meier estimates of the probability of disease-specific survival and overall survival according to dichotomized protein expression for CNTNAP2 and S100A8, respectively. Patients with CNTNAP2-positive and S100A8-negative tumors had significantly shorter survival times. P-values, hazard ratios (HR), and 95% confidence intervals (95% CI) were calculated using the log-rank test and Cox proportional hazards regression, respectively. The x-axes depict Months after initial diagnosis and the y-axes depict Disease-specific survival or Overall survival. (c) Representative immunohistochemical staining showing protein expression patterns in the invasive tissue component.

Figure 2. Predictive performance of prognostic models including CNTNAP2 or S100A8. (a) The lines represent the time-dependent area under the ROC curve (AUC(t)) for disease-specific survival using established clinical variables (lymph node status and tumor size) assessed individually and in conjunction with CNTNAP2 protein expression. The estimated performance of the model including CNTNAP2 was marginally better than
the model containing the established clinical variables alone, increasing the C-index from 0.941 to 0.949. (b) The lines represent the time-dependent area under the ROC curve (AUC(t)) for overall survival using established clinical variables (lymph node status, tumor size, differentiation, age) assessed individually and in conjunction with S100A8 protein expression. Combining the established clinical variables with S100A8 protein expression increased the C-index significantly from 0.605 to 0.833. The x-axes depict Survival time in months and the y-axes depict AUC(t).

Tables

Table 1. Clinicopathological features for OSCC patients in Cohorts I-III
Table 2. Univariate Cox proportional hazard regression models for OSCC Cohorts I-II
Table 3. Incidence of molecular marker immunopositivity in OSCC (Cohort III)
Table 4. Relationship between protein expression and clinicopathological features in OSCC (Cohort III)
Table 5. Univariate and multivariate survival analysis for disease-specific and overall survival in OSCC (Cohort III)

Additional files

Additional file 1: Table S1. Optimal antibody conditions and H-score cutoffs
Figure 1. Parris Aziz et al. (2014)
Additional files provided with this submission:

Additional file 1: Table 1.xlsx, 13K
http://www.biomedcentral.com/imedia/1660707665124049/supp1.xlsx
Additional file 2: Table 2.xlsx, 14K
http://www.biomedcentral.com/imedia/4928252001240491/supp2.xlsx
Additional file 3: Table 3.xlsx, 12K
http://www.biomedcentral.com/imedia/7092371491240491/supp3.xlsx
Additional file 4: Table 4.xlsx, 27K
http://www.biomedcentral.com/imedia/8033858211240491/supp4.xlsx
Additional file 5: Table 5.xlsx, 13K
http://www.biomedcentral.com/imedia/1898719479126955/supp5.xlsx
Additional file 6: Additional file 1 Table S1.xlsx, 11K
http://www.biomedcentral.com/imedia/1004380851240491/supp6.xlsx