



HAL
open science

A three-gene expression signature associated with positive surgical margins in tongue squamous cell carcinomas: Predicting surgical resectability from tumour biology?

Zuzana Saidak, Caroline Pascual, Jebrane Bouaoud, Louise Galmiche, Florian Clatot, Stéphanie Dakpé, Cyril Page, Antoine Galmiche

► To cite this version:

Zuzana Saidak, Caroline Pascual, Jebrane Bouaoud, Louise Galmiche, Florian Clatot, et al.. A three-gene expression signature associated with positive surgical margins in tongue squamous cell carcinomas: Predicting surgical resectability from tumour biology?. *Oral Oncology*, 2019, 94, pp.115-120. 10.1016/j.oraloncology.2019.05.020 . hal-02352881

HAL Id: hal-02352881

<https://normandie-univ.hal.science/hal-02352881>

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

A three-gene expression signature associated with positive surgical margins in tongue squamous cell carcinomas: predicting surgical resectability from tumour biology?

Zuzana Saidak ^{1,2,10}, Caroline Pascual ^{1,3}, Jebrane Bouaoud ^{4,5}, Louise Galmiche ⁶, Florian Clatot ^{7,8},
Stéphanie Dakpé ^{1,4}, Cyril Page ³, Antoine Galmiche ^{1,9,10}

1: CHIMERE team, EA7516, Université de Picardie Jules Verne, Amiens, France

2: Department of Molecular Oncobiology, Centre de Biologie Humaine, CHU Amiens, France

3: Department of Otorhinolaryngology, CHU Amiens, France

4: Department of Maxillofacial surgery, CHU Amiens, France

5: Department of Maxillofacial surgery and stomatology, APHP, Hôpital Pitié Salpêtrière, Paris, France

6: Department of Pathology, APHP, Hôpital Necker-Enfants Malades, Paris, France.

7: INSERM U1245 / IRON team, Rouen, France

8: Centre Henri Becquerel, Rouen, France

9: Department of Biochemistry, Centre de Biologie Humaine, CHU Amiens, France

10: **Corresponding authors:** Zuzana Saidak, PhD, Laboratoire d'Oncobiologie Moléculaire, Centre de Biologie Humaine (CBH), CHU Amiens Sud, Avenue Laennec, 80054 Amiens Cedex, France.

Tel:033-(0)322087074; Fax: 033-(0)322087026; E-mail: Saidak.Zuzana@chu-amiens.fr

Antoine Galmiche, MD PhD, Service de Biochimie, Centre de Biologie Humaine (CBH), CHU Amiens Sud, Avenue Laennec, 80054 Amiens Cedex, France. Tel:033-(0)322087017; Fax: 033-

(0)322087026; E-mail: Galmiche.antoine@chu-amiens.fr

Running title : Surgical margins and tumour biology in HNSCC

Word count for the text: 2758

List of abbreviations: ALI: angiolymphatic invasion; AUC: Area Under the Curve; DEG: differentially expressed genes; CNV: Copy Number Variations; ECS: extracapsular spread; FDR: false discovery rate; HNSCC: head and neck squamous cell carcinoma; PNI: perineural invasion; ROC: Receiver Operating Characteristic; SM: Surgical Margins; TCGA: The Cancer Genome Atlas.

Abstract

Objectives: Achieving complete tumour resection is one of the main goals of surgery for head and neck squamous cell carcinoma (HNSCC) tumours. Whether biological characteristics of tumours contribute to the surgical resectability and the presence of positive surgical margins (SM) after resection of HNSCC is unclear. We aimed to address this issue.

Materials and Methods: We used data from The Cancer Genome Atlas (TCGA) to relate the SM status of 356 HNSCC tumours covering five major primary locations (tongue, larynx, tonsils, floor of mouth and buccal mucosa) with data from multiple omics approaches (transcriptomic, genomic and proteomic analyses).

Results: We identified three differentially expressed genes whose expression was significantly associated with the presence of positive SM in tongue tumours (n=144). The three genes (*CCDC66*, *ZRANB2* and *VCPKMT*) displayed significantly higher mRNA levels in tongue tumours with positive SM compared to tumours with negative SM. The corresponding gene expression signature identified tongue tumours with a positive SM with high sensitivity and specificity (85% and 76%, respectively, Area Under the Curve (AUC)=0.84). Tongue tumours with this signature were characterised by a high grade, elevated proliferation levels and a tumour stroma with fewer fibroblasts and endothelial cells.

Conclusion: Positive SM were found to be strikingly associated with tumour biology in tongue tumours. These findings offer interesting perspectives for biomarker identification and precision surgery in these tumours.

(223 words)

Key words: Head and neck squamous cell carcinoma; The Cancer Genome Atlas (TCGA); gene expression; Surgical Margins; Tongue tumours.

Introduction

Head and neck squamous cell carcinomas (HNSCC) are a group of tumours with heterogeneous clinical presentation that arise from the mucosal epithelia of the head and neck [1,2]. Surgical resection of the primary tumour is often the first line of treatment for HNSCC, and complete tumour resection (R0, negative margins) is the major aim of surgery [1-3]. However, the extent of resection is usually restrained because of the proximity of vital organs and anatomical structures. The surgical procedure is therefore not only guided by oncological, but also anatomical considerations and the sequels that will impact the patient's quality of life [1-3]. Among the parameters that define tumour aggressiveness, surgical margins (SM) pose a unique problem. Positive SM are one of the strongest determinants of prognosis for HNSCC patients, making the final SM status an essential element in the decision of whether to apply adjuvant therapy [3,4]. In addition, compared to angiolymphatic invasion (ALI), lymph node extracapsular spread (ECS) or perineural invasion (PNI), SM are intimately related to the surgical procedure. Because of this, SM have been the focus of recent investigations, aiming for example to define an optimal margin that is considered safe for resection [5-9]. New studies have also investigated optimised tools and strategies to rapidly and unambiguously explore the SM status, using for example spectral properties for optimal detection of tumour cells [10]. Finally, molecular analyses hold considerable promise for a better understanding of SM in HNSCC [11-14] and could ultimately lead to the identification of clinically-applicable biomarkers.

In this context, the emergence of omics technologies and the rise of systems biology have provided powerful tools for the identification of the molecular events associated with HNSCC [15]. Exploring the molecular alterations present in each tumour is made easier thanks to these powerful techniques. The increasing availability of the data also facilitates the advancement of studies exploring novel strategies for personalised medicine. The Cancer Genome Atlas (TCGA), an integrated effort coordinated by the National Cancer Institute (NCI), offers for example free access to clinical and biological characteristics for more than 500 HNSCC tumours [16,17]. The accessibility of this large collection of surgically-resected HNSCC tumours, combined with extensive characterisation of tumour

biology, provides an unprecedented opportunity to study the determinants of the outcome of surgery at the molecular level. In the present study, we addressed the possibility that SM might arise as a consequence of tumour biology, rather than solely due to anatomical constraints or the surgeon's ability. Using data available from TCGA, we compared genomic, transcriptomic and proteomic data of HNSCC tumours with different SM status in different primary locations. Our findings point to the existence of a link between SM and tumour biology in tongue tumours.

Materials and Methods

Patient data and tumour stratification. Basic clinical, pathological and RNA expression data (RNA SeqV2 data normalised using RNA-Seq by Expectation Maximization: RSEM) were retrieved for 528 HNSCC patients using cBioportal at: <http://cbioportal.org> [18,19] in December 2018. Information regarding the SM status was available for 469 patients. « Negative » SM status indicates that the tumour is more than 2mm from specimen surface, « Close » SM: tumour is less than or equal to 2mm from specimen surface, « Positive » SM: tumour on specimen surface. For this study, we defined tumours with positive SM as those noted as “positive” or “close”. Importantly, SM status reflects the status of the main resection sample, and not the additional resections eventually performed after extemporaneous pathological examination. Tumours were stratified according to the primary tumour location, keeping only those anatomical locations with data for more than 20 patients (i.e. floor of mouth, larynx, tonsil, buccal mucosa and tongue) (Table 1).

Identification of Differentially Expressed Genes (DEG) associated with SM. A non-parametric unpaired Wilcoxon-Mann-Whitney test was used to compare the groups with positive / negative SM for each anatomical location. For each gene, the p value was adjusted using the False Discovery Rate (FDR) Benjamini-Hochberg procedure to control for Type I error. Genes with a $p < 0.05$ FDR were retained for further analysis. Gene methylation, copy number and proteomic Reverse Phase Protein Array (RPPA) data were retrieved from cBioportal in December 2018. Aneuploidy scores and data regarding mutational burden were obtained for all HNSCC tumours from the study by Taylor *et al.* [20].

Analysis of tumour grade, proliferation and pathological examination. Tumour grade was classified from G1 to G4 (G1 corresponding to well-differentiated tumours, and G4 to undifferentiated tumours). The assessment of tumour proliferation was based on the study by Wolf *et al.* [21], which used gene expression data to identify a proliferation module consisting of 120 genes, co-regulated across multiple datasets and closely correlated to cancer cell proliferation [21]. For pathological examination of tumour samples, digitised whole-slide images of HNSCC frozen sections, taken from

regions adjacent to the tissue used for molecular analysis and scanned at 20X resolution, were retrieved from Cbioportal and analysed by a local pathologist (L.G.).

Tumour microenvironment analysis. The STROMAL and IMMUNE scores were used to evaluate tumour purity, *i.e.* the infiltration levels of stromal and immune cells based on gene expression data (<http://bioinformatics.mdanderson.org/estimate/>) [22]. The Microenvironment Cell Population-counter (MCP counter) method was also used to quantify the absolute abundance of immune and stromal cell populations [23]. The MCP counter analyses the presence of gene signatures specific for 8 types of immune cells, fibroblasts and endothelial cells [23].

Statistics: Comparisons of two groups of numeric data were done using the unpaired Wilcoxon-Mann-Whitney test. Chi-squared test was used for comparisons of categorical data. $p < 0.05$ after FDR was the threshold of significance. All analyses were done with R version 3.4.2 (<https://www.r-project.org>).

Results

In order to relate SM with tumour biology, we used data for 469 HNSCC tumours from TCGA for which the SM status was available. Because the location of any primary tumour is a key determinant of the surgical procedure used, we organised all tumours according to five major locations for which data on the SM status were available for more than 20 tumours (Table 1). The tumour locations retained were: tongue (n=144, 18.7% SM+), tonsil (n=30, 43.3% SM+), floor of mouth (n=61, 24.6% SM+), larynx (n=98, 13.3% SM+) and buccal mucosa (n=23, 43.5% SM+). The Wilcoxon-Mann-Whitney statistical test with FDR correction was applied in order to identify all DEG linked to the SM status. Using this strategy, we did not find any DEG associated with the SM status in tumours of the tonsils, floor of mouth, larynx or buccal mucosa. On the other hand, for tongue tumours, we have identified three genes with significantly different mRNA expression levels according to the SM status (*CCDC66*, p=0.017 FDR ; *ZRANB2*, p=0.048 FDR and *VCPKMT*, p=0.048 FDR) (Fig. 1A,B). Other aspects of tumour biology were explored in parallel, using the various Omics data available. An analysis centered on the frequency of mutations and copy number variations (CNV) of the genes reported to be key drivers in HNSCC progression [24] showed no statistically significant association between SM and specific genomic alterations in tongue tumours (Suppl. Table 1). Similarly, we found no differences in total DNA methylation, mutational burden or aneuploidy between tongue tumours depending on their SM status (Suppl. Fig. 1). The three differentially expressed genes, *CCDC66*, *ZRANB2* and *VCPKMT* are located on distinct chromosomes (3p14, 1p31 and 14q21, respectively), with infrequent mutations or CNV in HNSCC tumours (frequency <3%, data not shown). Further analyses showed no correlation between the expression levels of *CCDC66*, *ZRANB2* and *VCPKMT* mRNA levels and DNA methylation at specific loci, global mutational burden or aneuploidy (data not shown).

In order to examine the association between the three genes *CCDC66*, *ZRANB2* and *VCPKMT* and the SM status, we calculated a SM signature score by averaging the z scores for the three genes for each individual tongue tumour (Fig 1C). A ROC analysis was used to investigate the potential ability of this score to predict positive SM, giving an AUC of 0.84 [0.76-0.93, 95% CI,

$p < 0.0001$] (Fig. 1D). At the optimal cutoff point of 0.085, the SM signature score stratified patients with tongue tumours into two groups: 50 patients with a positive SM signature (high SM score) (35.5%) and 91 patients with a negative SM signature (low SM score) (64.5%). With this optimal set up, the SM signature identified the SM+ tumours with a surprisingly high specificity (76%) and sensitivity (85%). We further examined the SM signature according to the TCGA annotation of « positive » or « close » SM. No statistical difference in the SM signature score was found between the subgroups « positive » and « close » (Suppl Fig. 2). The performance of the SM signature for margins classified as « close » was AUC = 0.82 [0.72-0.93, 95% CI, $p < 0.0001$]. Overall, these findings highlight a strong and specific association between the expression of the three DEGs and the SM status in tongue tumours.

We pursued our investigation by exploring the pathological significance of the SM signature in tongue tumours. We found no significant correlation between the SM signature score and Tumour (T) or Node (N) status (Fig. 2A). The SM signature score was not associated with the presence of Angiolymphatic Invasion (ALI), Lymph node Extracapsular Spread (ECS) or Perineural Invasion (PNI) (Suppl. Fig. 3). The three genes were not part of the transcriptional signatures reported for ALI, ECS or PNI in HNSCC tumours from TCGA [25]. We found no significant difference in tumour weight or longest dimension between tongue tumours with a high value of the SM score (high SM score) vs the rest (low SM score) (Suppl. Fig. 4). Conversely, tumours with high histological grade had a significantly higher SM signature score compared to low grade tumours ($p < 0.01$, Wilcoxon-Mann-Whitney test) (Fig. 2A). When tongue tumours were sorted according to their SM signature (using the previously defined optimal cut-off), we also noticed a significant positive association between the SM signature and tumour proliferation, determined according to Wolf *et al.* [21] (Fig. 2B). Considering the potential importance of the tumour microenvironment in the aggressiveness of tongue tumours, we examined the possibility that the SM signature might be linked to the tumour content of stromal and immune cells (Fig. 2C). To address this possibility, we used the STROMAL and IMMUNE scores, reported by Yoshihara *et al.* [22]. This analysis revealed that tumours with high vs low SM signature score did not differ in terms of inflammatory / immune cell infiltration (Fig. 2C). However, tumours with high SM signature score had a significantly lower Stromal Score (median=-178 for low SM score vs -526 for high SM score, $p = 0.006$) (Fig. 2C). These findings were confirmed with a separate algorithm, the MCP counter, which showed that tongue tumours with a high SM signature score had lower levels

of fibroblasts and endothelial cells. (Suppl. Fig. 5). Overall, this analysis suggested that a high SM signature score identifies tongue tumours with higher grade, higher proliferation levels and lower stromal cell content. These findings were supported by the histological examination of the pathological sections of tongue tumours with a high SM signature score (Fig. 2D).

Discussion

The primary aim of the present study was to examine whether tumour biology might in some cases contribute to positive SM and failure to perform a successful HNSCC resection. Because of the different clinical situations generated by the complex anatomy of the head and neck [1-3], we stratified patients according to the primary tumour location for further data analysis. Only with tongue tumours did we observe a significant association between SM and tumour biology, with the expression of three genes (*CCDC66*, *ZRANB2* and *VCPKMT*) linked to the SM status. Taking these three genes into consideration, a gene expression signature was constructed, and found to be closely associated with SM in tongue tumours (sensitivity: 85%, specificity: 76%). Our findings are important because the tongue is the most frequent primary location for HNSCC tumours [1-3]. Tongue tumours are often treated by surgery with the main goal of achieving a complete tumour resection, but little is known regarding the biological parameters that dictate the behaviour and response to treatment of this subtype of tumours [26,27]. Interestingly, our findings suggest that surgical resectability might be linked to biological tumour characteristics in a large fraction of tongue tumours.

Why we found a link between tumour biology and SM in the tongue and not in tumours from other primary locations is still speculative at this stage. The homogeneous nature of the tongue tissue, consisting essentially of muscle cells, and the lack of anatomical constraints might account for the apparent contribution of tumour biology to SM in tongue tumours. This does not necessarily exclude the contribution of tumour biology to surgical resectability in other primary tumour locations. For example, HNSCC tumours of the larynx are highly heterogeneous in their biological and clinical presentation [28]. A further level of complexity comes from the different surgical procedures that can be used, ranging from total to partial laryngectomy, as well as the use of laser or transoral robotic surgery [29]. This heterogeneity constitutes an obstacle to studies aiming to relate tumour biology to surgical resectability. The number of patients in TCGA is likely too small to address the contribution of tumour biology for specific tumour sublocations, such as for example glottic tumours.

To the best of our knowledge, the present study is the first to suggest that tumour biology is linked to surgical resectability for a subtype of HNSCC. Our study nevertheless suffers from some limitations, such as the lack of an independent validation cohort. Another important point is the fact that the SM status was defined on the initial tumour sample. The SM status that we analysed in the present study more closely reflects tumour resectability than the final outcome of surgery, considering that additional resections were performed in almost half of the patients with positive SM in tongue tumours in our study (data not shown). How tumour biology could contribute to positive SM remains unclear at this stage. None of the three genes that we identified here is an established oncogene or a gene reported to play a role in HNSCC aggressiveness [15,24]. Little has been reported about *ZRANB2* and *VCPKMT* in cancer. However, a couple of recent studies have addressed the contribution of *CCDC66* to tumour growth and the acquisition of locally invasive properties in lung and colon tumours [30,31]. These properties were linked to the induction of an Epithelial-Mesenchymal Transition (EMT) in cancer cells, an evolutionarily conserved process that converts an epithelial cell into a mobile, presumably more aggressive mesenchymal state [31]. Irrespective of their cellular function, the expression of the three genes appeared to be related to some important biological characteristics of these tumours : i) a higher grade, *i.e.* the presence of less differentiated cancer cells ; ii) higher proliferation levels ; iii) a tumour stroma with fewer fibroblasts and endothelial cells. Considering these elements, we propose that the SM score might identify aggressive tongue tumours, with an invasive growth pattern [32,33]. Another possibility, not mutually exclusive to the previous one, is that the biological properties reported here might interfere with the diagnosis and staging of tongue tumours. For example, a less-dense stroma might reduce the chances of detecting tumours by palpation or compromise tumour detection by magnetic resonance imaging [34,35]. Taken together, our study provides a proof of principle that tumour biology may be an important determinant of tumour resectability for HNSCC. An important goal of future studies should be to extend the conclusions of the present study in a separate cohort, ideally providing more detail regarding clinical parameters and tumour imaging.

Provided that the conclusions of our study can be independently validated, we propose that molecular analysis of tongue tumours could help to predict the resectability, and perhaps even the success of surgical resection, as has been suggested for example for colorectal and ovarian tumours [36-38]. The concept of precision surgery, *i.e.* the possibility of choosing a surgical approach based on molecular analysis of the tumour is promising for HNSCC, especially considering the recent progress made in functional genomics for this tumour type [39]. RNA sequencing of tumour material, obtained either through tumour biopsy or body fluids (blood or saliva), may be helpful in the prospective assessment of tumour resectability [40]. Interestingly, the remarkable property of *CCDC66*, which was reported to form circular RNA with high stability as a consequence of its splicing [30], might be an advantage for the clinical transposition of this type of analysis. Whether this analysis will in the end help physicians to improve clinical staging, risk assessment and decision making in surgery of tongue tumours is an interesting possibility that deserves further investigation.

Acknowledgements

The authors gratefully acknowledge the contribution of the patients and researchers involved in TCGA. We thank the Amiens University Hospital and la Ligue contre le Cancer for financial support.

Role of the Funding source

Work in our laboratory is supported by the CHU Amiens Picardie and Ligue contre le Cancer. The funders took no part in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- [1] Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. *Lancet* 2008;371:1695-709.
- [2] Bose P, Brockton NT, Dort JC. Head and neck cancer: from anatomy to biology. *Int J Cancer* 2013;133:2013-23.
- [3] Adelstein D, Gillison ML, Pfister DG, Spencer S, Adkins D, Brizel DM, et al. NCCN Guidelines Insights: Head and Neck Cancers, Version 2.2017. *J Natl Compr Canc Netw* 2017;15:761-70.
- [4] Backes C, Bier H, Knopf A. Therapeutic implications of tumor free margins in head and neck squamous cell carcinoma. *Oncotarget* 2017;8:84320-8.
- [5] Tasche KK, Buchakjian MR, Pagedar NA, Sperry SM. Definition of "Close Margin" in oral cancer surgery and association of margin distance with local recurrence rate. *JAMA Otolaryngol Head Neck Surg* 2017;143:1166-72.
- [6] Williams MD. Determining adequate margins in head and neck cancers: Practice and continued challenges. *Curr Oncol Rep* 2016;18:54.
- [7] Smits RW, Koljenović S, Hardillo JA, Ten Hove I, Meeuwis CA, Sewnaik A, et al. Resection margins in oral cancer surgery: room for improvement. *Head Neck* 2016;38:E2197-203.
- [8] Shapiro M, Salama A. Margin Analysis: Squamous Cell Carcinoma of the Oral Cavity. *Oral Maxillofac Surg Clin North Am* 2017;29:259-67.
- [9] Sim FW, Xiao HD, Bell RB. Margin Analysis: Squamous Cell Carcinoma of the Oropharynx. *Oral Maxillofac Surg Clin North Am* 2017;29:269-80.

- [10] Clark DJ, Mao L. Understanding the Surgical Margin: A Molecular Assessment. *Oral Maxillofac Surg Clin North Am* 2017;29:245-58.
- [11] Brennan JA, Mao L, Hruban RH, Boyle JO, Eby YJ, Koch WM, et al. Molecular assessment of histopathological staging in squamous-cell carcinoma of the head and neck. *N Engl J Med* 1995;332:429-35.
- [12] Sinha P, Bahadur S, Thakar A, Matta A, Macha M, Ralhan R, et al. Significance of promoter hypermethylation of p16 gene for margin assessment in carcinoma tongue. *Head Neck* 2009;31:1423-30.
- [13] Ogbureke KU, Weinberger PM, Looney SW, Li L, Fisher LW. Expressions of matrix metalloproteinase-9 (MMP-9), dentin sialophosphoprotein (DSPP), and osteopontin (OPN) at histologically negative surgical margins may predict recurrence of oral squamous cell carcinoma. *Oncotarget* 2012;3:286-98.
- [14] Liu SA, Wang CC, Jiang RS, Wang WY, Lin JC. Genetic analysis of surgical margins in oral cavity cancer. *Br J Surg* 2018;105:e142-e149.
- [15] Leemans CR, Snijders PJF, Brakenhoff RH. The molecular landscape of head and neck cancer. *Nat Rev Cancer* 2018;18:269-82.
- [16] The Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 2015;517:576-82.
- [17] Cooper LAD, Demicco EG, Saltz JH, Powell RT, Rao A, Lazar AJ. Pan-cancer insights from The Cancer Genome Atlas: the pathologist's perspective. *J Pathol* 2018;244:512-24.

- [18] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2:401-4.
- [19] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;6:pl1.
- [20] Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* 2018;33:676-89.
- [21] Wolf DM, Lenburg ME, Yau C, Boudreau A, van 't Veer LJ. Gene co-expression modules as clinically relevant hallmarks of breast cancer diversity. *PLoS One* 2014;9:e88309.
- [22] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Commun* 2013;4:2612.
- [23] Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* 2016;17:218.
- [24] Dotto GP, Rustgi AK. Squamous Cell Cancers: A unified perspective on biology and genetics. *Cancer Cell* 2016;29:622-37.
- [25] Saidak Z, Clatot F, Chatelain D, Galmiche A. A gene expression profile associated with perineural invasion identifies a subset of HNSCC at risk of post-surgical recurrence. *Oral Oncol* 2018;86:53-60.
- [26] Almangush A, Heikkinen I, Mäkitie AA, Coletta RD, Läärä E, Leivo I, et al. Prognostic biomarkers for oral tongue squamous cell carcinoma: a systematic review and meta-analysis. *Br J Cancer* 2017;117:856-66.

- [27] Hussein AA, Forouzanfar T, Bloemena E, de Visscher J, Brakenhoff RH, Leemans CR, et al. A review of the most promising biomarkers for early diagnosis and prognosis prediction of tongue squamous cell carcinoma. *Br J Cancer* 2018;119:724-36.
- [28] Mor N, Blitzer A. Functional Anatomy and Oncologic Barriers of the Larynx. *Otolaryngol Clin North Am.* 2015;48:533-45.
- [29] Smith RV. Transoral robotic surgery for larynx cancer. *Otolaryngol Clin North Am.* 2014;47:379-95.
- [30] Hsiao KY, Lin YC, Gupta SK, Chang N, Yen L, Sun HS, Tsai SJ. Noncoding Effects of Circular RNA CCDC66 Promote Colon Cancer Growth and Metastasis. *Cancer Res.* 2017;77:2339-50.
- [31] Joseph NA, Chiou SH, Lung Z, Yang CL, Lin TY, Chang HW, et al. The role of HGF-MET pathway and CCDC66 cirRNA expression in EGFR resistance and epithelial-to-mesenchymal transition of lung adenocarcinoma cells. *J Hematol Oncol.* 2018;11:74.
- [32] Yu P, Wang W, Zhuang Z, Xie N, Xu J, Wang C, et al. A novel prognostic model for tongue squamous cell carcinoma based on the characteristics of tumour and its microenvironment: iBD score. *Histopathology.* 2018 [Epub ahead of print].
- [33] Almangush A, Heikkinen I, Bakhti N, Mäkinen LK, Kauppila JH, Pukkila M, et al. Prognostic impact of tumour-stroma ratio in early-stage oral tongue cancers. *Histopathology* 2018;72:1128-35.
- [34] Choi N, Noh Y, Lee EK, Chung M, Baek CH, Baek KH, et al. Discrepancy between cTNM and pTNM staging of oral cavity cancers and its prognostic significance. *J Surg Oncol* 2017;115:1011-8.

[35] Baba A, Okuyama Y, Ikeda K, Kozakai A, Suzuki T, Saito H, et al. Undetectability of oral tongue cancer on magnetic resonance imaging; clinical significance as a predictor to avoid unnecessary elective neck dissection in node negative patients. *Dentomaxillofac Radiol.* 2019;20180272.

[36] Borley J, Wilhelm-Benartzi C, Brown R, Ghaem-Maghamsi S. Does tumour biology determine surgical success in the treatment of epithelial ovarian cancer? A systematic literature review. *Br J Cancer* 2012;107:1069-74.

[37] Margonis GA, Buettner S, Andreatos N, Sasaki K, Ijzermans JNM, van Vugt JLA, et al. Anatomical resections improve disease-free survival in patients with KRAS-mutated colorectal liver metastases. *Ann Surg* 2017;266:641-9.

[38] Evrard S, Torzilli G, Caballero C, Bonhomme B. Parenchymal sparing surgery brings treatment of colorectal liver metastases into the precision medicine era. *Eur J Cancer* 2018;104:195-200.

[39] Foy JP, Saintigny P, Goudot P, Schouman T, Bertolus C. The promising impact of molecular profiling on treatment strategies in oral cancers. *J Stomatol Oral Maxillofac Surg* 2017;118:242-7.

[40] Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol.* 2017;14:531-48.

Figure legends

Figure 1: Identification of Differentially expressed genes (DEG) associated with the surgical margin (SM) status in surgically-resected HNSCC tumours from TCGA.

A. A Wilcoxon-Mann-Whitney test was carried out to identify the genes that are differentially expressed in tumours with positive vs negative SM in five primary locations. Note that the analysis was performed after tumour stratification according to primary tumour location, retaining only those locations for which data were available for more than 20 tumours (Tongue n=144; Tonsils n=30, Floor of mouth n=61; Larynx n=98; Buccal mucosa n=23). The illustration indicates the number of DEG with $p < 0.05$ (FDR corrected) in tumours +/- SM. For tongue tumours, three genes were identified with different mRNA levels according to SM (*CCDC66*, *ZRANB2* and *VCPKMT*). **B.** Violin plots showing the mRNA expression (RSEM) of *CCDC66*, *ZRANB2* and *VCPKMT*, the three genes significantly associated with positive SM in tongue tumours with positive or negative SM. *** $p < 0.001$. **C.** Heatmap showing the expression of the three genes, *CCDC66*, *ZRANB2* and *VCPKMT* (high expression=red, low expression=blue) and their co-occurrence with positive SM (black) and negative SM (green). **D.** ROC analysis for the prediction of positive SM, using the SM signature score *i.e.* the average z score for the three genes *CCDC66*, *ZRANB2* and *VCPKMT*. AUC (Area under the curve) = 0.84 [0.76-0.93, 95% CI] $p < 0.0001$.

Figure 2: Pathological tumour characteristics associated with the SM signature.

A. Boxplots showing the SM signature score in tongue tumours according to tumour size (T), lymph node involvement (N) and grade (G). **B.** Tumour proliferation was estimated using gene expression analysis centered on the tumour proliferation module (n=120 genes) in tumours stratified according to the SM signature score (high SM score: n=50, low SM score: n=91). **C.** Stromal and immune cell

tumour content was assessed using the corresponding algorithms in tumours stratified according to the SM signature score (high SM score: n=50, low SM score: n=91). **p<0.01, ***p<0.001. **D.** A representative histological slide from a tongue tumour with a high SM score. Note the poor differentiation of the HNSCC cells, assembled into trabeculae of immature cells with minimal keratinisation (arrow heads).

Figure 1

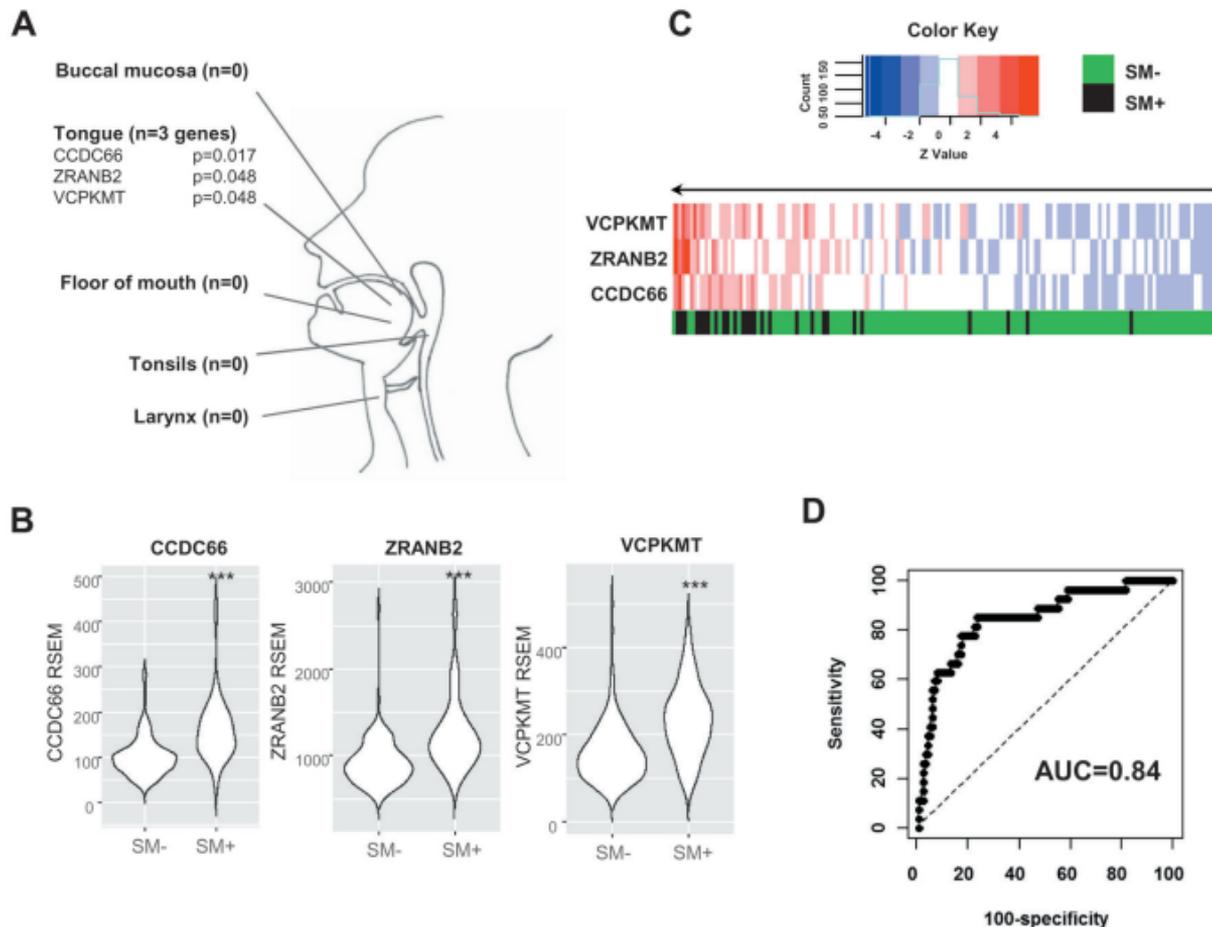


Figure 2

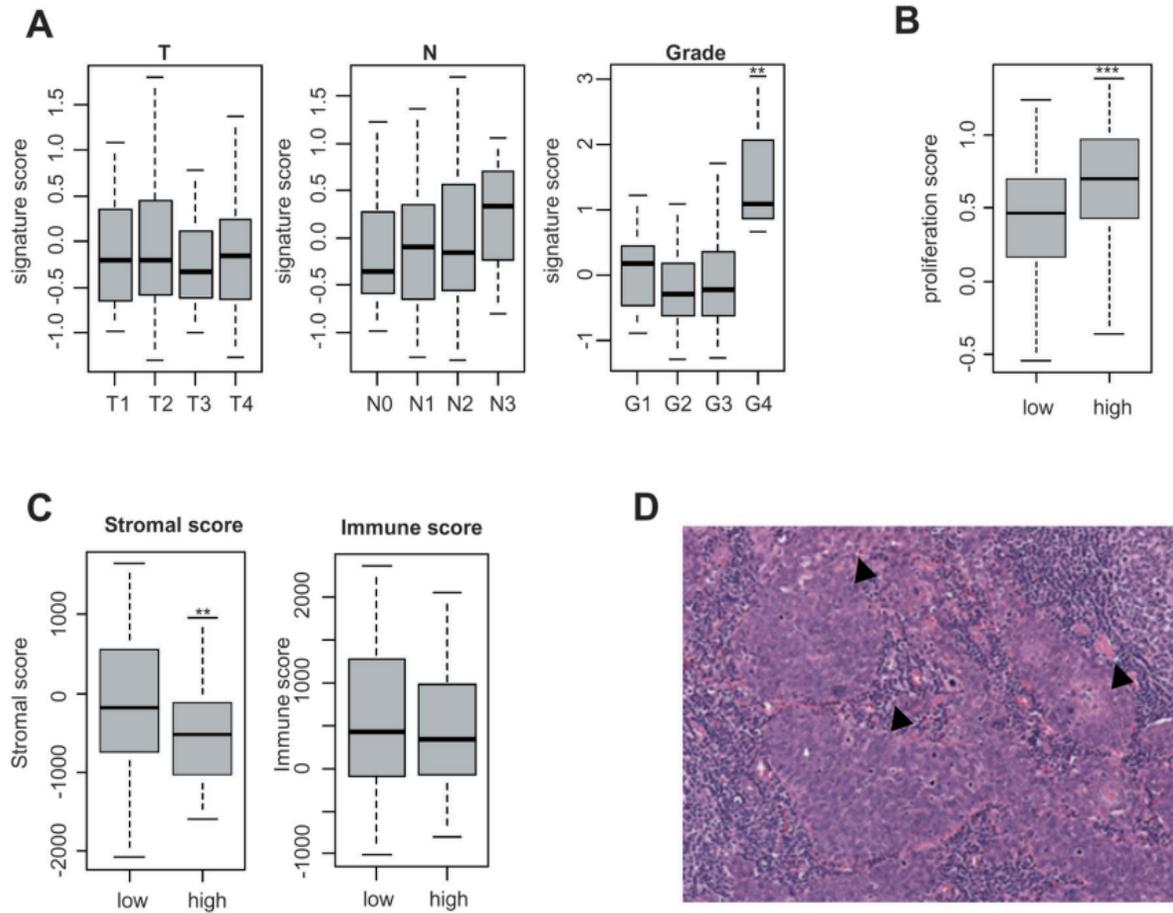


Table 1: Clinical characteristics of HNSCC patients, stratified according to primary tumor location and surgical margin (SM) status

		Tongue (n=144)		Larynx (n=98)		Tonsils (n=30)		Floor of mouth (n=61)		Buccal Mucosa (n=23)	
		SM- (n=117)	SM+ (n=27)	SM- (n=85)	SM+ (n=13)	SM- (n=17)	SM+ (n=13)	SM- (n=46)	SM+ (n=15)	SM- (n=13)	SM+ (n=10)
Patients	Male	79 (68%)	16 (59%)	72 (85%)	11 (85%)	13 (76%)	10 (77%)	32 (70%)	14 (93%)	8 (62%)	8 (80%)
	Female	38 (32%)	11 (41%)	13 (15%)	2 (15%)	4 (24%)	3 (23%)	14 (30%)	1 (7%)	5 (38%)	2 (20%)
Age (median years)		59	63	61	62	51	57	60	56	66	70
HPV status *	Negative	21 (18%)	2 (7%)	13 (15%)	2 (15%)	1 (6%)	0 (0%)	5 (11%)	3 (20%)	4 (31%)	2 (20%)
	Positive	3 (3%)	4 (15%)	1 (1%)	1 (8%)	9 (53%)	5 (38%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
	N/A	93 (80%)	21 (78%)	71 (84%)	10 (77%)	7 (41%)	8 (62%)	41 (89%)	11 (73%)	9 (69%)	8 (80%)
T Stage	T1	19 (16%)	6 (22%)	5 (6%)	1 (8%)	5 (29%)	1 (8%)	3 (7%)	0 (0%)	0 (0%)	0 (0%)
	T2	41 (35%)	9 (33%)	13 (15%)	0 (0%)	11 (65%)	8 (62%)	11 (24%)	4 (27%)	7 (54%)	1 (10%)
	T3	34 (29%)	8 (30%)	20 (24%)	3 (23%)	0 (0%)	2 (15%)	10 (22%)	2 (13%)	1 (8%)	6 (60%)
	T4	16 (14%)	4 (15%)	44 (52%)	9 (69%)	0 (0%)	1 (8%)	19 (41%)	9 (60%)	4 (31%)	3 (30%)
	N/A	7 (6%)	0 (0%)	3 (4%)	0 (0%)	1 (6%)	1 (8%)	3 (7%)	0 (0%)	1 (8%)	0 (0%)
Angiolymphatic Invasion	Yes	21 (18%)	10 (37%)	26 (31%)	7 (54%)	4 (24%)	4 (31%)	7 (15%)	10 (67%)	6 (46%)	3 (30%)
	No	69 (59%)	9 (33%)	39 (46%)	2 (15%)	5 (29%)	3 (23%)	31 (67%)	5 (33%)	6 (46%)	5 (50%)
	N/A	27 (23%)	8 (30%)	20 (24%)	4 (31%)	8 (47%)	6 (46%)	8 (17%)	0 (0%)	1 (8%)	2 (20%)
Extracapsular spread	Gross	4 (3%)	6 (22%)	3 (4%)	3 (23%)	1 (6%)	3 (23%)	3 (7%)	3 (20%)	0 (0%)	2 (20%)
	Micro	20 (17%)	2 (7%)	15 (18%)	4 (31%)	1 (6%)	4 (31%)	7 (15%)	4 (27%)	3 (23%)	1 (10%)
	None	73 (62%)	12 (44%)	47 (55%)	3 (23%)	6 (35%)	3 (23%)	27 (59%)	5 (33%)	7 (54%)	7 (70%)
	N/A	20 (17%)	7 (26%)	20 (24%)	3 (23%)	9 (53%)	3 (23%)	9 (20%)	3 (20%)	3 (23%)	0 (0%)
Perineural Invasion	Yes	48 (41%)	14 (52%)	17 (20%)	6 (46%)	0 (0%)	2 (15%)	18 (39%)	7 (47%)	6 (46%)	7 (70%)
	No	42 (36%)	7 (26%)	44 (52%)	3 (23%)	10 (59%)	8 (62%)	20 (43%)	8 (53%)	6 (46%)	3 (30%)
	N/A	27 (23%)	6 (22%)	24 (28%)	4 (31%)	7 (41%)	3 (23%)	8 (17%)	0 (0%)	1 (8%)	0 (0%)

* Positive HPV status based on p16 or ISH; N/A: data not available. Data show are n (%).