Oncogenic drivers in 11q13 associated with prognosis and response to therapy in advanced oropharyngeal carcinomas


Published in:
Oral Oncology

DOI:
10.1016/j.oraloncology.2018.06.010

Publication date:
2018

Document version
Accepted manuscript

Document license
CC BY-NC-ND

Citation for published version (APA):

Terms of use
This work is brought to you by the University of Southern Denmark through the SDU Research Portal. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 29. Apr. 2021
Oncogenic drivers in 11q13 associated with prognosis and response to therapy in advanced oropharyngeal carcinomas.

MC Barros-Filho1*, LA Reis-Rosa2*, M Hatakeyama1*, FA Marchi4, T Chulam3, C Scapulatempo-Neto4, UR Nicolau5, AL Carvalho6, CAL Pinto7, SA Drigo8, LP Kowalski3,9, SR Rogatto9,10§.

*These authors contributed equally.

1International Research Center (CIPE), A. C. Camargo Cancer Center, São Paulo, Brazil;
2Instituto de Medicina Tropical de SP Universidade de São Paulo-USP, São Paulo, Brazil;
3Department of Head and Neck Surgery and Otorhinolaryngology, A. C. Camargo Cancer Center, São Paulo, Brazil;
4Molecular Oncology Research Center, Barretos, and Diagnósticos da América (DASA), São Paulo, Brazil;
5Department of Oncology, A. C. Camargo Cancer Center, São Paulo, Brazil;
6Department of Head and Neck Surgery, Barretos Cancer Hospital, Barretos, Brazil;
7Department of Pathology, A. C. Camargo Cancer Center, São Paulo, Brazil;
8Department of Surgery and Orthopedics, School of Medicine and Department of Veterinary Clinic, School of Veterinary Medicine and Animal Science, UNESP, Botucatu, Brazil;
9National Institute of Science and Technology in Oncogenomics (INCiTO), São Paulo, Brazil;
10Department of Clinical Genetics, Vejle Hospital, Institute of Regional Health Research, University of Southern Denmark, Denmark.

§Corresponding author: Professor Silvia Regina Rogatto
Department of Clinical Genetics, Vejle Hospital
Institute of Regional Health Research, University of Southern Denmark
Beriderbakken 4, 7100 Vejle – DK
silvia.regina.rogatto@rsyd.dk

Word Count: 3,495
Abstract

Objectives: To identify potential molecular drivers associated with prognosis and response to treatment in advanced oropharyngeal squamous cell carcinomas (OPSCC).

Materials and Methods: Thirty-three OPSCC biopsies from untreated Brazilian patients were evaluated for human papilloma virus genotyping, genome wide copy number alterations and gene expression profiling. Data were integrated using CONEXIC algorithm. Validation with TCGA dataset and confirmation by RT-qPCR of candidate genes were performed.

Results: High-risk HPV positive cases, detected in 55% of advanced OPSCC, were associated with better outcome. Losses of 8p11.23-p11.22, 14q11.1-q11.2 and 15q11.2, and gains of 11q13.2 and 11q13.2-q13.3 were detected as recurrent alterations. Gains of 3q26.31 and 11q13.2 and losses of 9p21.3 were exclusively detected in HPV-negative tumors. Two clusters of expression profiles were observed, being one composed mostly by HPV positive cases (83%). HPV-positive enriched cluster showed predominantly immune response-related pathways. Integrative analysis identified 10 modulators mapped in 11q13, which were frequently cancer-related. These 10 genes showed copy number gains, overexpression and an association with worse survival, further validated by TCGA database analyses. Overexpression of four genes (ORAOV1, CPT1A, SHANK2 and PPFIA1) evaluated by RT-qPCR confirmed their association with poor survival. Multivariate analysis showed that PPFIA1 overexpression and HPV status are independent prognostic markers. Moreover, SHANK2 overexpression was significantly associated with incomplete response to treatment.

Conclusion: The integrative genomic and transcriptomic data revealed potential driver genes mapped in 11q13 associated with worse prognosis and response to treatment, giving fundamentals for the identification of novel therapeutic targets in OPSCC.
Introduction

In the last decades, several epidemiological studies have revealed decreased incidence of head and neck squamous cell carcinomas (HNSCC) in oral cavity and larynx as a consequence of lower exposure to the tobacco products. Nevertheless, an increasing incidence of oropharyngeal squamous cell carcinomas (OPSCC) mainly associated with oncogenic human papillomavirus (HPV) has been reported [1,2,3,4,5,6].

In general, HPV-positive OPSCC is associated with good prognosis, presenting better survival in comparison with HPV-negative cases [7,8,9,10,11]. De-escalation of radiation and chemotherapy for HPV-positive cases has been proposed and tested in different clinical trials aiming to avoid overtreatment and long-term toxicities [8,12,13]. However, accurate identification of cases with good prognosis and treatment-responsive tumors are critical findings, since distant metastasis may occur in a set of HPV-positive OPSCC [8,14]. In contrast, few advances have been made for treatment of HPV-negative OPSCC patients and a large number of them will present loco-regional recurrence [9].

The molecular mechanisms underlying oropharyngeal carcinogenesis have been investigated and potential biomarkers were reported, however the data are still unclear and controversial [15,16,17,18,19]. The integration of genomic and transcriptomic analysis can be used to identify cancer-driver genes and disrupted pathways, which can be drug targetable [20]. This strategy has revealed functionally relevant drivers involved in the carcinogenic process in different tumor types, including oral carcinoma [21], ovarian cancer [22], penile carcinoma [23,24], uterine leiomyoma [25] and leiomyosarcoma [26].

Integration of genomic, transcriptomic and epigenomic data of 279 HNSCC, including oral (n=172, 62%), oropharyngeal (n=33, 12%) and laryngeal (n=72, 26%)
carcinomas was reported by The Cancer Genome Atlas (TCGA) [27]. Distinct genetic alterations were observed between HPV-positive HNSCCs (68% in oropharynx) and HPV-negative cases. Recurrent deletions and truncating mutations of TRAF3 found in HPV-positive tumors were associated with anti-viral immune response. Conversely, HPV-negative HNSCCs presented loss of 9p21.3 (including CDKN2A gene) and co-amplifications of 11q13 and 11q22, which contain genes implicated in cell death/NF-kB and Hippo pathways [27]. A distinct genetic subgroup of HPV-negative tumors is also being reported, characterized by low frequency of copy number alterations (CNA), wild-type TP53, mutation in HRAS and CASP8 and more favorable prognosis [13].

Nevertheless, prognostic and predictive biomarkers in advanced OPSCC are still limited and need to be further investigated. In this study, we integrated DNA CNA and gene expression analyses to identify drivers in advanced OPSCC according to HPV status. In silico functional analysis was performed to identify genes and pathways associated with oropharyngeal carcinogenesis, which can reveal potential drug targets.
Patients and methods

Patients and samples

Fresh-frozen tumor biopsy samples from 40 OPSCC patients naive of treatment were obtained from A.C. Camargo Cancer Center and Barretos Cancer Hospital, Brazil. Eligibility criteria included patients harboring locally advanced clinical stages III, IVA and IVB according to AJCC (7th Edition). The follow-up time ranged from 0.5 to 190 months (mean of 53 months). The study was approved by the Human Research Ethics Committee from both Institutions (A.C. Camargo Cancer Center #1249/09 and Barretos Cancer Hospital #139/2008). All patients provided written informed consent. Patients underwent curative therapy according to standard clinical protocol taking into account the medical decisions for each patient, which included induction chemotherapy (IC) followed by radiotherapy with concurrent chemotherapy (CT-RT), and upfront surgical resection followed by radiotherapy with or without chemotherapy. Pretreatment risk stratification was defined as low, intermediate and high, based on HPV status, smoking history and tumor/node stage [7].

Treatment and response assessment

The majority of patients included in this study were treated by IC followed by CT-RT (n=24). Patients were treated with a combination of docetaxel, cisplatin and 5-fluorouracil (TPF) as IC followed by radiotherapy and weekly carboplatin or cisplatin concurrent to radiotherapy. Cetuximab was employed in eight cases as concurrent therapy to radiation (one case received IC before). Six patients were treated by surgery followed by adjuvant therapy (RT or RT+CT); one patient deceased surgery and one patient lost the follow-up before completing the treatment (supplementary Table S1). Six patients were treated only by surgery followed by adjuvant therapy (RT and or CT),
one patient deceased soon after surgery and one patient lost the follow-up before treatment begins.

**Nucleic acids extraction and HPV genotyping**

OPSCC samples (80% of tumor cells) and surrounding normal tissues were macrodissected for DNA (Qiagen DNeasy Blood & Tissue Kit; Qiagen, Valencia, CA) and total RNA extraction (RNasy MiniKit; Qiagen, Valencia, CA). The Linear Array HPV Genotyping Test Kit (Roche Molecular Systems, Inc., Branchburg, NJ, USA) was used for HPV detection.

**CNA analysis by array-based Comparative Genomic Hybridization**

OPSCC (n=33) and normal commercial DNA (Promega) samples were differentially labeled (Genomic DNA Enzymatic Labeling Kit; Agilent Technologies) and the hybridized on Agilent Human CGH 180K Oligo Microarrays. Genomic data were extracted by Feature Extraction 10.1.1.1 software (Agilent Technologies) and analyzed using the Nexus Copy Number software (v.6.0, Biodiscovery, El Segundo, CA, USA). CNA was defined as exceeding the significance threshold of $1\times10^{-6}$ and containing at least five consecutive altered probes per segment. The thresholds were defined as the average log$_2$ CGH fluorescence ratio for copy gains $\geq$0.6, high copy number gains $\geq$1.4, losses $\leq$-0.6 and homozygous losses $\leq$-1.25. Genomic variants detected in control individuals from worldwide populations and classified as common (>1%) according to DGV database (http://dgv.tcag.ca/dgv/app/home) were excluded. Alterations detected in at least 20% of the cases were selected for further analysis. The unsupervised hierarchical clustering analysis was performed using complete linkage and Euclidian distance.
Gene expression microarray

Total RNA from OPSCC (n=33) and surrounding non-neoplastic oropharyngeal tissues (n=3) were labeled and hybridized using the Two-Color Human GE 4x44K microarray platform (Agilent Technologies), following the manufacturer instructions. Data processing, quality control filtering and normalization (Lowess) were performed using the Feature Extraction v.10.1.1.1 software (Agilent Technologies) and an in-house pipeline. Gene expression analysis was performed using R version 2.15 (http://www.bioconductor.org/) and BRB ArrayTools software (v.4.4.0). An unsupervised hierarchical clustering analysis was employed with the most variable probes (interquartile range >0.1) using complete linkage and Euclidian distance. Transcriptomic variations among clusters were identified by significance analysis of microarray (SAM) (false discovery ratio <1%).

The CNA and expression microarray data are available at the Gene Expression Omnibus (GEO) (GSE111395).

Integrative analysis

Paired CNA and gene expression data of 33 OPSCC was integrated using COpy Number and EXpression In Cancer (CONEXIC) algorithm to identify drivers, which results in a ranked list with high scores modulators [28]. In this analysis, unbalanced expressed genes are correlated with the expression of group of genes (modules), and genomic regions with significant alterations indicate a greater possibility of the gene has some adaptive advantage on the tumor phenotype. Parameters used in this analysis were described elsewhere [25]. Mann–Whitney test (Prismv.5.0, GraphPad Software,
La Jolla, CA) was applied to analyze the relationship between gene expression and CNA.

**External data validation and in silico functional analysis of driver candidates**

Driver candidates were validated in 78 OPSCC available in the TCGA public database (SNP arrays and RNAseq), using cBio Cancer Genomics Portal (http://www.cbioportal.org/public-portal/index.do) assessed in February 2018). Ingenuity® Pathway Analysis (IPA v2.3; http://www.ingenuity.com) and KOBAS (v.3.0; http://kobas.cbi.pku.edu.cn/) software was additionally used to highlight enriched canonical pathways modulations.

**Data validation by reverse transcription quantitative polymerase chain reaction (RT-qPCR)**

Thirty OPSCC (23 array-dependent and 7 independent samples) and four non-neoplastic oropharyngeal tissues were assessed to evaluate the *ORAOV1, CPT1A, PPFIA1* and *SHANK2* transcript levels by RT-qPCR (primer sequences are showed in supplementary Table S2). *HMBS and HPRT1* were selected as reference genes using geNorm algorithm (tested in combination with *ACTB, GAPDH* and *GUSB*) [29]. Robotic pipetting was carried out using QIAgility (QIAGEN) in a total volume of 12.5uL containing Power SYBR Green PCR Master Mix (Applied Biosystems), 20 ng of cDNA and 200nM of each primer. All samples were analyzed in duplicate. Normalized relative gene expression was obtained according to Pfaffl method [30].

**Statistical analysis**
Fisher exact test was used to associate the clustering analysis and copy number alterations with clinicopathological features. Transcript levels were evaluated according to CNA and response to therapy using Mann–Whitney test. High expression levels were defined as at least two-fold in tumor tissues compared to normal tissues. Overall and disease-free survival probabilities were calculated using Kaplan-Meyer method and statistically compared with log rank. Multivariate analysis was applied using Cox proportional hazards including all variables with \( p \) value < 0.1 in the univariate analysis. Receiver Operating Characteristic (ROC) curve was applied to verify the predictive value of the variables in relation to response to treatment, defined by the Area Under the ROC Curve (AUC). Statistical analysis was performed with GraphPad Prism v5.0 (GraphPad Software Inc., La Jolla, CA) and SPSS v.21.0 (SPSS; Chicago, IL).
Results

HPV infection status and prognosis

Clinical and pathological features of the Brazilian patients included in this study are presented on Table 1. Fifty-eight percent (23/40 cases) of the advanced OPSCC were HPV-positive, being HPV16 the most prevalent subtype (19 cases; 1 HPV16/18; 1 HPV16/33; 1 HPV18 and 1 HPV33) (supplementary Table S1). Tonsil was the most common affected site by HPV (57%) (Table 1). Longer overall survival was observed for HPV-positive compared to HPV-negative cases (p=0.008, supplementary Figure S1A). Risk stratification based on HPV status, smoking history and tumor/node stage [7] was performed. High-risk patients (n=15) showed shorter survival compared with low (n=15) and intermediate-risk patients (n=10) (p=0.013, supplementary Figure S1B).

Genomic and transcriptomic analyses in OPSCCs

Figure 1 summarizes the genomic and transcriptomic analysis approaches and the main results. Recurrent CNAs included losses of 8p11.23-p11.22, 14q11.1-q11.2, 15q11.2 and gains of 11q13.2 and 11q13.2-q13.3. After the comparison of all significant CNA (p<0.05) with DGV database, only gains of 11q13.2 and 11q13.2-q13.3 were classified as new alterations. Genes and ORFs mapped in these regions are shown in supplementary Table S3.

CNA profiles revealed two clusters of samples, being the cluster 1 enriched by HPV-positive cases (cluster 1: 67% and cluster 2: 28%, p=0.038), low/intermediate risk stratification (cluster 1: 77% and cluster 2: 25%, p=0.005), complete chemoradiation response (cluster 1: 89% and cluster 2: 26%, p=0.004) and lower mortality (cluster 1: 27% and cluster 2: 82%, p=0.008) (Figure 2A). CNA analysis according to HPV status
revealed gains of 3q26.31 and 11q13.2 and losses of 9p21 exclusively detected in HPV-negative cases ($p<0.05$, supplementary Table S4).

The gene expression analysis also demonstrated a cluster of samples enriched by HPV-positive cases (cluster 1: 87% and cluster 2: 44%, $p=0.027$). A trend of low/intermediate risk enrichment (cluster 1: 85% and cluster 2: 50%, $p=0.067$) was also observed in the cluster 1, which presented a partial overlapping with CNA cluster 1 (Figure 2B). The comparison between these two gene expression clusters revealed 423 differentially expressed transcripts (supplementary Table S5). The pathways identified in HPV-enriched cluster were related to immune response, including interferon-gamma (IFNG) and programmed cell death-1 (PD-1) signaling (supplementary Table S6) (IPA and KOBAS v3.0). The main molecule predicted as activated by IPA upstream regulator analysis was IFNG (supplementary Table S7), which was detected as overexpressed in HPV-enriched cluster.

Integrated genomic and gene expression data

The integrative analysis using CONEXIC revealed 16 putative modulators mapped at 11q13 (ANO1, CCND1, CPT1A, FADD, FGF3, FGF4, FGF19, IGHMBP2, MRGPRD, MRGPRF, MRPL21, ORAOV1, PPFIA1, PPP6R3, SHANK2 and TPCN2). Ten of 16 modulators showed statistically significant association between copy number gains and higher expression levels (Figure 3A). Interestingly, copy number gains of these genes were associated with shorter overall survival (Figure 3B). In silico analysis showed interactions between these 10 driver candidates and their association with cell cycle, cancer, organismal injury and abnormalities (supplementary Figure S2). In addition, interactions with recognized cancer-related molecules (calmodulin and p21) were observed, being CCND1 and CPT1A susceptible for drug inhibition (arsenic
trioxide and perhexiline, respectively) (drugbank: https://www.drugbank.ca/. Accession in February 2018).

Validation of candidate driver genes

The putative drivers were further tested using the TCGA database of OPSCC, confirming the increased gene expression levels and genomic copy number gains of 10 genes (supplementary Figure S3A). Similarly to our data, these genomic alterations were also associated with shorter overall survival (supplementary Figure S3B).

In 30 OPSCC samples with available tissues for RNA extraction RT-qPCR confirmed a significant association of genomic gains in 11q13 and overexpression of ORAOV1, CPT1A, PPFIA1 and SHANK2 genes (Figure 4A). High expression levels of these genes were associated with shorter survival (Figure 4B). The multivariate analysis revealed PPFIA1 expression and HPV status as independent prognostic markers in advanced OPSCC (Table 2).

Among the four potential markers identified, significant decreased expression level of SHANK2 was detected as predictive marker of complete response (Figure 4C) and complete response without recurrence in at least five years (Figure 4D) after concurrent radiotherapy and chemotherapy (p=0.010 and p=0.012, respectively). Compared to risk stratification (low, intermediate or high) or HPV status (negative or positive), SHANK2 relative expression presented a higher performance in predicting concurrent CT-RT complete response (AUC=0.733, AUC=0.646, AUC=0.833; respectively) (Figure 4C) and complete response without recurrence (AUC=0.659, AUC=0.606, AUC=0.811; respectively) (Figure 4D).
Discussion

This study was designed to reveal potential drivers in locally advanced OPSCC, aiming to identify new therapeutic targets and biomarkers. About half of OPSCC cases were positive for high-risk HPV, which was associated with improved survival, corroborating previous reports [7,8,10,11,31,32]. Among the HPV positive cases, 95% (22/23) were HPV16/18, similar to recent findings described in HNSCC [33]. Moreover, worse prognosis was observed for high-risk patients (according to HPV, TNM and smoking), supporting the hypothesis that the clinical course of HPV-positive tumors is modified by tobacco usage [7]. Curiously, the distribution of HPV infection was uneven between the patient recruitment centers. Tumors collected from Sao Paulo (the largest metropolis in Brazil) presented 77% (23/30) of positivity compared to 0% (0/10) in the Barretos cohort (an inland city in Brazil southeast). The Brazilian OPSCC patients were recently characterized as having 4% of HPV positivity compared to 59% and 31% in USA and Europe, respectively [34]. However, the predominance of HPV in Brazilian patients can drastically differ due to geographic heterogeneity of our population, which can vary socioeconomically, in alcohol-tobacco consumption and in sexual behavior.

Significant recurrent gains of 11q and two clusters based on CNA similarities were observed, being the cluster 1 enriched by HPV-positive tumors. Particularly, gains of 11q13.2 (LRP5, PPP6R3 and SAPS3 genes) were exclusively detected in HPV-negative samples. Amplification of 11q13 is commonly reported in different tumors, including HNSCC [20,27,35,36,37,38], particularly in HPV-negative tumors [39,40] and OPSCC [15]. Mutually exclusive amplifications of 11q13 (CCND1 and FADD) and 11q22 (BIRC2 and YAP1) were detected in HPV-negative HNSCC [27]. These data
emphasizes that distinct molecular alterations drive the oropharynx carcinogenesis according to the HPV status.

Large-scale gene expression analysis also showed distinct clusters according to HPV status. HPV-enriched cluster presented predominantly immune response-related genes, mainly involving negative regulators of T-cell immune function, as cytotoxic T-lymphocyte–associated antigen 4 (CTLA-4) and programmed death 1 (PD-1) and its ligand PD-L1 immune checkpoints pathway members. The PD-1/PD-L1 and CTLA-4, have shown promising clinical success as a cancer immunotherapy target [41]. Nonetheless, many factors influence in the response to immunotherapies, as the degree of tumor lymphocyte infiltration and the checkpoint proteins levels [13]. A previous study showed that 70% of HPV-positive HNSCCs presents PD-L1 positive immunostaining [42]. The authors reported increased transcript levels of IFNG in oropharyngeal tumors HPV+/PD-L1+, suggesting that IFN-γ might induce PD-L1 in tumor cells. In our study, IFN-γ was highly overexpressed in the HPV-positive enriched cluster and was predicted by IPA as the one of the main upstream regulator. Tumor infiltrating lymphocytes overexpressing PD-1 presented better clinical outcome, mostly in head and neck tumors HPV-positives [43]. Furthermore, CD8+ tumor infiltrating lymphocytes showing high expression levels of PD-1 were found more frequently in HPV-negative tumors with compromised IFNG secretion. These patients presented worse prognosis. Conversely, HPV-positive tumors presenting low PD-1 expression in T cells were associated with better response to immunotherapy (anti-PD-1) and outcome [43]. In melanoma, the immunostaining positivity for MHC (Major Histocompatibility Complex) class II molecules was associated with therapeutic response and better clinical outcome in anti-PD-1-treated patients [44]. In our study, overexpression of MHC class II genes was detected in HPV-related cluster (supplementary Table S4).
Overall, our findings give additional evidence that PD-1/PD-L1 is a target for new therapeutic strategies in OPSCC, particularly in HPV-positive cases.

Correlation between CNAs mapped in 11q13.2-11q13.3 and corresponding genes expression levels was previously described in HNSCC [21]. In our study, 10 cancer-related genes (ANO1, CCND1, CPT1A, FADD, IGHMBP2, MRPL21, ORAOV1, PPFIA1, SHANK2 and TPCN2) mapped in 11q13 showed copy number gains and overexpression. These findings were further validated using the TCGA, suggesting that other genes besides CCND1 may confer an advantage to OPSCC cells [45]. Moreover, gains involving these genes were associated with worse survival in our study and in TCGA, thus supporting their relevance in OPSCC outcomes.

Further analysis using RT-qPCR confirmed an association among ORAOV1, CPT1A, PPFIA1 and SHANK2 overexpression and shorter survival. A meta-analysis comprising 15 publically available microarray gene expression data assessing 140 normal and 277 HNSCC samples revealed a list of 181 differentially expressed genes. ANO1 and FADD (both mapped in 11q13) were overexpressed in at least 30% of the TCGA samples (528 HNSCC), and associated with higher risk of recurrence and death [46]. Gains of 11q13.2, including CPT1A, were also associated with worse survival in patients with esophageal carcinoma [47]. CPT1A (Carnitine palmitoyl transferase 1A) gene encodes a rate-limiting enzyme for fatty acid transport inside the mitochondria for β-oxidation. Fatty acid oxidation is essential for ATP production in metabolic stress conditions and particularly in cancer [48]. Enzymes involved in fatty acid oxidation, such as CPT1, are promising targets for cancer therapy [48] and might be considered as a potential target in OPSCC. Interestingly, CPT1 enzyme can be pharmacologically inhibited by perhexiline [49].
Although increased expression levels of ORAOV1 (oral cancer overexpressed 1) have been reported in several tumors [50,51,52], its biological function is not completely characterized. In oral carcinomas, reduced ORAOV1 expression was associated with delayed cell growth \textit{in vitro} and inhibition of tumor growth and angiogenesis \textit{in vivo} [51]. A protective response to reactive oxygen species was related to ORAOV1 overexpression, a common feature of many cancers [53], including HNSCC [54]. Despite the small sample size and number of events (10 patients died during a 5-year follow-up), PPFIA1 overexpression was identified as an independent prognostic marker (independent of the HPV status and/or tobacco consumption). PPFIA1 was reported as the highest up-regulated gene associated with copy number gains of 11q13 in HNSCC cell lines [55]. In breast cancer cells, the silencing of PPFIA1 promotes inhibition of migration and invasion [56]. Overall, our data suggest that PPFIA1 acts as an oncogene in OPSCC, although its function in these tumors requires further studies.

\textit{SHANK2} overexpression was associated with incomplete response to therapy, being able to distinguish treatment-unresponsive from responsive tumors more efficiently than HPV status. \textit{SHANK2} is a member of scaffold protein family involved in excitatory synapses in the central nervous system. In oral carcinomas, co-amplification and overexpression of \textit{SHANK2} and \textit{CTTN} was reported, suggesting a cooperative role of these proteins on tumor cell motility and invasiveness [57]. Amplification and overexpression of \textit{SHANK2} was also associated with poor prognosis in esophageal squamous cell carcinoma [58]. In oral carcinomas, \textit{SHANK2} was described as one of seven prognostic markers evaluated by DNA methylation [59]. These finding suggest that CNA and/or epigenetic mechanisms may result in \textit{SHANK2} increased expression and consequent HNSCC aggressiveness. To our knowledge, the association between genes mapped in 11q13 and chemoradiation response was not
previously reported. The role of these genes in oropharyngeal carcinogenesis, particularly in tumor aggressiveness and their ability to predict therapy response, should be further investigated.

Considering that only locally advanced OPSCC biopsies obtained from patients subsequently treated with specific therapeutic strategies were included in our study, the major limitation was the small sample size (n=40). Nonetheless, with the addition of an independent sample set (n=78) from TCGA (not treatment-standardized), we were able to confirm the potential clinical relevance of genes mapped in 11q13.

In conclusion, we described a higher frequency of HPV infection than previous Brazilian OPSCC reports, which varied greatly according to the geographic region and tobacco-alcohol consumption. HPV-positive tumors showed an immune response-related gene signature, which can have an impact on immunotherapy response prediction leading to a more precise treatment indication. In addition to CCND1 previously described, novel driver genes including ORAOV1, CPT1A, PPFIA1 and SHANK2 genes mapped in 11q13 presented a role in the prognosis and chemoradiation response prediction, regardless of the HPV status.

Conflict of interest statement
None declared

Acknowledgments
The authors would like to thank the Macromolecules Laboratory at the A.C. Camargo Cancer Center and Barretos Cancer Hospital. This study was supported by grants from the National Institute of Science and Technology in Oncogenomics (FAPESP #2008/57887–9 and CNPq #573589/08–9).
Authors’ contributions

SRR and LPK conceived and designed the experiments; MCBF, MH and LARR conducted the experiments; MCBF, FAM, LARR, SAD analyzed the data; SRR and LPK contributed with reagents/materials; SRR, LPK and SAD supervised the study; CALP performed the histopathological evaluation; URN, ALC and LPK selected the cases and obtained the clinical data; MCBF, SAD and SRR wrote and edited the manuscript. All authors read and approved the final version of the manuscript.
References


FIGURES CAPTIONS

**Figure 1.** Flowchart illustrating the methodologies and major findings of the study.

**Figure 2.** Unsupervised hierarchical clustering analysis of (A) array-CGH and (B) global gene expression data according to the clinical features. CT: chemotherapy; RT: radiotherapy; *Risk stratification (Ang et al. 2010), #Patients submitted to induction chemotherapy followed by concurrent platinum/cetuximab and radiotherapy or only concurrent platinum/cetuximab and radiotherapy.

**Figure 3.** **A.** Association of DNA copy number gains with expression levels of ten genes identified as driver candidates in OPSCC. **B.** Overall survival analysis according to DNA copy number gains in 10 candidate driver genes. *p<0.05; **p<0.01; ***p<0.001 (Mann-Whitney test); Overall survival analysis performed by Kaplan Meyer method with log rank test.

**Figure 4.** **A.** Relative gene expression analysis using RT-qPCR in non-neoplastic samples and OPSCC with two copies (normal) and 11q13 copy number gains. **B.** Disease-free survival curves of tumors in normal range of expression and overexpression of the genes assessed by RT-qPCR (p value obtained from log rank test). **C and D.** Expression level of SHANK2 (RT-qPCR) as a predictive marker of complete response (C) and durable complete response (D) to concurrent radiotherapy and chemotherapy, comparing HPV status and risk stratification (Ang et al., 2010). *p<0.05; **p<0.01 (Mann Whitney non-parametric test); AUC: area under the ROC (receiver operating characteristic) curve.
### Table 1. Patient distribution according to the demographic and clinical variables and 11q13 amplification status.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Number of patients (%)</th>
<th>11q13 amplification (%)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2 (5)</td>
<td>0/2 (0)</td>
<td>0.542</td>
</tr>
<tr>
<td>Male</td>
<td>38 (95)</td>
<td>11/31 (35.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (interquartile range)</td>
<td>55.5 (47.8-62.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;60 years</td>
<td>26 (65)</td>
<td>6/20 (30)</td>
<td>0.714</td>
</tr>
<tr>
<td>≥60 years</td>
<td>14 (35)</td>
<td>5/13 (38.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Anatomic Site</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base of the tongue</td>
<td>16 (40)</td>
<td>4/14 (28.6)</td>
<td>1.000</td>
</tr>
<tr>
<td>Soft palate</td>
<td>1 (2.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tonsil</td>
<td>23 (57.5)</td>
<td>6/19 (31.8)</td>
<td></td>
</tr>
<tr>
<td><strong>HPV infection status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>17 (42.5)</td>
<td>7/15 (46.7)</td>
<td></td>
</tr>
<tr>
<td>HPV16</td>
<td>19 (47.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other HPV (18/33)</td>
<td>2 (5)</td>
<td>4/18 (22.2)</td>
<td>0.163</td>
</tr>
<tr>
<td>Co-infection (HPV16 and HPV18/33)</td>
<td>2 (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tobacco Consumption</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>9 (22.5)</td>
<td>1/8 (12.5)</td>
<td>0.218</td>
</tr>
<tr>
<td>Yes</td>
<td>31 (77.5)</td>
<td>10/25 (40)</td>
<td></td>
</tr>
<tr>
<td><strong>Alcohol Consumption</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>14 (35)</td>
<td>1/10 (10)</td>
<td>0.109</td>
</tr>
<tr>
<td>Yes</td>
<td>26 (65)</td>
<td>10/23 (43.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Clinical T stage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>1 (2.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>7 (17.5)</td>
<td>7/23 (30.4)</td>
<td>0.696</td>
</tr>
<tr>
<td>T3</td>
<td>21 (52.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>11 (27.5)</td>
<td>4/10 (40)</td>
<td></td>
</tr>
<tr>
<td><strong>Clinical N stage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>11 (27.5)</td>
<td>2/9 (22.2)</td>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>5 (12.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N2</td>
<td>18 (45)</td>
<td>9/24 (37.5)</td>
<td>0.681</td>
</tr>
<tr>
<td>N3</td>
<td>6 (15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stratification Risk</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>15 (37.5)</td>
<td>4/20 (20)</td>
<td>0.065</td>
</tr>
<tr>
<td>Intermediate</td>
<td>10 (25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>15 (37.5)</td>
<td>7/13 (53.8)</td>
<td></td>
</tr>
</tbody>
</table>
Treatment response (RXT+CT)

<table>
<thead>
<tr>
<th></th>
<th>HR (CI&lt;sub&gt;95%&lt;/sub&gt;)</th>
<th>p value</th>
<th></th>
<th>HR (CI&lt;sub&gt;95%&lt;/sub&gt;)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stable disease/progression</td>
<td>3 (9.7)</td>
<td></td>
<td>5/9 (55.6)</td>
<td>0.097</td>
<td></td>
</tr>
<tr>
<td>Partial</td>
<td>7 (22.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete</td>
<td>21 (67.7)</td>
<td></td>
<td>4/19 (21.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not available*</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Clinical outcome

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Free of disease</td>
<td>25 (62.5)</td>
<td></td>
<td>8/19 (42.1)</td>
<td>0.278</td>
</tr>
<tr>
<td>Local recurrence</td>
<td>11 (27.5)</td>
<td></td>
<td>3/14 (21.4)</td>
<td>0.278</td>
</tr>
<tr>
<td>Distant Metastasis</td>
<td>4 (10)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Death

<table>
<thead>
<tr>
<th></th>
<th>HR (CI&lt;sub&gt;95%&lt;/sub&gt;)</th>
<th>p value</th>
<th></th>
<th>HR (CI&lt;sub&gt;95%&lt;/sub&gt;)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>26 (65)</td>
<td></td>
<td>5/20 (25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>14 (35)</td>
<td></td>
<td>6/13 (46.2)</td>
<td>0.270</td>
<td></td>
</tr>
</tbody>
</table>

*risk stratification according to Ang et al., 2010; *nine patients not submitted to concurrent radiotherapy and chemotherapy (CT-RT); °amplification in at least one gene located in 11q13; p value: Fisher exact test.

Table 2. Univariate and multivariate analysis of OPSCC overall survival.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariate analysis</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (CI&lt;sub&gt;95%&lt;/sub&gt;)</td>
<td>p value</td>
<td>Multivariate analysis</td>
<td>HR (CI&lt;sub&gt;95%&lt;/sub&gt;)</td>
</tr>
<tr>
<td>HPV infection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Yes</td>
<td>0.3 (0.1-1)</td>
<td>0.056</td>
<td>0.1 (0.02-0.56)</td>
<td>0.001</td>
</tr>
<tr>
<td>Ang's risk stratification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low/Intermediate</td>
<td>1</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>High</td>
<td>3.7 (1-14.5)</td>
<td>0.056</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CPT1A expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal range</td>
<td>1</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Overexpression</td>
<td>6 (1.7-21.9)</td>
<td>0.006</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ORAOV1 expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal range</td>
<td>1</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Overexpression</td>
<td>18.3 (4.1-80.9)</td>
<td>&lt;0.001</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PPFIA1 expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal range</td>
<td>1</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Overexpression</td>
<td>11.9 (3.2-44.1)</td>
<td>&lt;0.001</td>
<td>30.9 (5.3-179.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SHANK2 expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal range</td>
<td>1</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Overexpression</td>
<td>7.6 (2-28.1)</td>
<td>0.003</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

HR: hazard ratio; CI<sub>95%</sub>: confidence interval of 95%. p value: Cox proportional-hazards regression.
FIGURES

Figure 1.
Figure 2.
Figure 3.
Figure 4.