



OPEN

Impact of genetic variants in clinical outcome of a cohort of patients with oropharyngeal squamous cell carcinoma

Ana Carolina de Carvalho¹, Sandra Perdomo^{2,3}, Wellington dos Santos¹, Gabriela Carvalho Fernandes⁴, Lais Machado de Jesus¹, Raiany Santos Carvalho⁵, Cristovam Scapulatempo-Neto^{1,6}, Gisele Caravina de Almeida⁵, Bruna Pereira Sorroche¹, Lidia Maria Rebolho Batista Arantes¹, Matias Eliseo Melendez^{1,7,8}, Pedro De Marchi^{9,10}, Neil Hayes¹¹, Rui Manuel Reis^{1,12,13} & André Lopes Carvalho^{1,3}✉

Tobacco- or human papillomavirus- driven oropharyngeal squamous cell carcinomas (OpSCC) represent distinct clinical, biological and epidemiological entities. The aim of this study was to identify genetic variants based on somatic alterations in OpSCC samples from an admixed population, and to test for association with clinical features. The entire coding region of 15 OpSCC driver genes was sequenced by next-generation sequencing in 51 OpSCC FFPE samples. Thirty-five percent of the patients (18/51) were HPV-positive and current or past tobacco consumption was reported in 86.3% (44/51). The mutation profile identified an average of 2.67 variants per sample. Sixty-three percent of patients (32/51; 62.7%) were mutated for at least one of the genes tested and *TP53* was the most frequently mutated gene. The presence of mutation in *NOTCH1* and *PTEN*, significantly decreased patient's recurrence-free survival, but only *NOTCH1* mutation remained significant after stepwise selection, with a risk of recurrence of 4.5 (HR 95% CI = 1.11–14.57; Cox Regression $p = 0.034$). These results show that Brazilian OpSCC patients exhibit a similar clinical and genetic profile in comparison to other populations. Molecular characterization is a promising tool for the definition of clinical subgroups, aiding in a more precise tailoring of treatment and prognostication.

Head and neck cancers (HNC) are ranked as the 8th most common tumors worldwide, with a total of 834,860 new cases and 431,131 deaths estimated for 2018¹.

The most prevalent site is the oral cavity, followed by the oropharynx and larynx, and more than 95% are squamous cell carcinomas (SCC)². Excessive tobacco and alcohol consumption and high-risk human papillomavirus (HPV) infection (mainly HPV-16 and HPV-18) in the oropharynx are the main risk factors^{3–6}. Besides etiology, HPV-positive and HPV-negative oropharyngeal squamous cell carcinoma (OpSCC) cases configure diseases with remarkably different clinical presentation, epidemiological and molecular profiles^{3–6}. Regarding prognosis, HPV-positive OpSCC have a far more favorable outcome when compared with HPV-negative cases⁶. For this reason, the tumor-node-metastasis (TNM) staging for HNC was adapted in the eighth edition to include p16^{INK4A} immunostaining as a surrogate for HPV status in OpSCC⁷.

¹Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, SP, Brazil. ²Institute of Nutrition, Genetics and Metabolism Research, Faculty of Medicine, Universidad El Bosque, Bogotá, Colombia. ³International Agency of Research on Cancer, Lyon, France. ⁴Center of Molecular Diagnosis, Barretos Cancer Hospital, Barretos, SP, Brazil. ⁵Research Support Center, Barretos Cancer Hospital, Barretos, SP, Brazil. ⁶Pathology and Molecular Diagnostics Service, Diagnósticos da América-DASA, São Paulo, SP, Brazil. ⁷Pelé Little Prince Research Institute, Curitiba, PR, Brazil. ⁸Little Prince College, Curitiba, PR, Brazil. ⁹Department of Medical Oncology, Barretos Cancer Hospital, Barretos, SP, Brazil. ¹⁰Oncoclinicas, Rio de Janeiro, RJ, Brazil. ¹¹Department of Medicine, Division of Oncology, UTHSC Center for Cancer Research, University of Tennessee Health Science Center, Memphis, USA. ¹²Life and Health Sciences Research Institute (ICVS), Medical School, University of Minho, Braga, Portugal. ¹³ICVS/3B's-PT Government Associate Laboratory, Braga/Guimarães, Portugal. ✉e-mail: carvalhoal@gmail.com

Treatment for OpSCC has evolved in the past decades and often includes approaches for organ preservation based on concomitant chemotherapy and radiotherapy followed by salvage surgery in non-responders⁸. This treatment frequently has long-term side effects, impacting patients' quality of life⁹. Several treatment de-escalation trials for HPV-positive OpSCC have been initiated, which may lead to personalized treatment based on HPV status¹⁰.

Around 2/3 of HNC patients are diagnosed with advanced disease in different parts of the world^{9,11–13} and at least 50% of them will relapse locally, regionally or at distant sites, within the first two years after treatment^{9,14,15}. While for HPV-positive OpSCC cases, 5-years overall survival rates (OS) are close to 80%, regardless of stage^{16–18}, TNM stage has a significant impact in OS for HPV-negative OpSCC: 70%, 58%, 50%, and 30% for stage I, II, III, and IV¹⁸. Therefore, there is an increasing need for predictors of treatment response and disease progression/relapse.

Studies characterizing the spectrum of genetic alterations in cancers have been enabling a better understanding of the molecular alterations playing a role in head and neck carcinogenesis^{19–22}. A high level of intertumoral heterogeneity has been observed, confirming the biological complexity of these tumors^{22–24}. A common finding is that HPV-negative tumors have a higher mutation burden while HPV-positive tumors have far fewer genes mutated and variants per tumor, regardless of smoking status^{22,23}. Results from The Cancer Genome Atlas (TCGA) consortium on the molecular profile of 279 HNSCC tumors confirmed the differences at the molecular level between HPV-positive and negative cases, with frequent activating helical domain mutations in *PIK3CA* in HPV-positive cases and *TP53* mutations and *CDKN2A* loss-of-function mutations in HPV negative cases²⁴. These findings suggest that the absence of the oncogenic effect from HPV oncoproteins in HPV-unrelated tumors requires the accumulation of multiple genetic aberrations to allow malignant transformation²³.

Despite the different strategies for diagnosis and treatment of OpSCC patients, and the advent of HPV as an important prognostic marker, there is no molecular biomarker to guide selection among treatment and follow-up options that directly have an impact on patient survival. Therefore, the lack of risk categories based on clinical features and/or biomarkers that can be used for personalized treatment approaches for OpSCC has propelled research into the molecular landscape of these tumors. Moreover, there is an underrepresentation of genomic data on the mutation profile of HNC patients from admixed populations. Therefore, this study aims at identifying molecular alterations in a set of HNC patients from Brazil and to better characterize, together with clinical variables, OpSCC cases according to their outcome.

Methods

Sample population and DNA isolation. This study included pre-treatment formalin-fixed paraffin embedded (FFPE) tissue biopsies from 51 patients with primary oropharyngeal squamous cell carcinoma (OpSCC) diagnosed at the Department of Head and Neck Surgery of Barretos Cancer Hospital in Brazil between 2009 and 2017. These cases were selected based on very stringent criteria to ensure reliability of the clinical and molecular data as follows: pre-treatment FFPE tumor biopsies available from patients treated with curative intent by chemo-radiation protocols and with available follow-up data; samples with at least 60% of tumor cells at the histopathological examination and, with enough DNA for library preparation at an acceptable level of integrity as evaluated by multiplex PCR²⁵. Finally, after sequencing, only samples with acceptable scores of DNA sequencing quality were included (as described below).

Hematoxylin & eosin stained sections of paraffin blocks containing the tumor tissue from the patients included were reviewed by two expert pathologists (CSN and GCA) for diagnosis confirmation and characterization of cellular components. Scrapings from regions containing at least 60% of tumor cells were processed using the QIAamp DNA FFPE Tissue Kit (Qiagen, Germany). Isolated DNA samples were eluted in 35 μ L of water and quantified in the Qubit fluorometer (Invitrogen, Carlsbad, CA) prior to storage at -20°C until use.

Information regarding tobacco and alcohol consumption, clinical and pathological features were retrieved from patient's charts. p16 immunohistochemistry was conducted as a surrogate marker for high-risk HPV (pre-diluted, monoclonal mouse antihuman p16INK4A protein, Clone E6H4TM, ready for use, Ventana, Tucson, AZ, USA). Samples with strong and diffuse nuclear and cytoplasmic staining in more than 75% of the tumor cells were considered positive for p16^{26–28}. Moreover, HPV-DNA detection of types HPV-16 and HPV-18 was performed in a subset of cases (n = 12) using droplet digital PCR as previously described and detailed in the Supplementary Methods²⁹.

The present study was approved by the Barretos Cancer Hospital Institutional Review Board (approval number 425/2013) and all methods were performed in accordance with the relevant guidelines and regulations.

Next-generation sequencing (NGS) and genetic ancestry determination. To identify somatic mutations, we performed NGS of a panel of 15 genes, including some of the most frequently mutated genes in HNC, *TP53*, *NOTCH1*, *CDKN2A*, *PTEN*, *PIK3CA*, *FBXW7*, *HRAS*, *TP63*, *CASP8*, *FAT1*, *KMT2D*, *RB1*, *IRF6*, *EZH2* and *NSD1*, based on a previous study³⁰. Bearing in mind the consensus regarding DNA fragmentation when recovered from FFPE, primers were designed to amplify fragments with sizes ranging from 125–175 base pairs, to minimize the effect of DNA fragmentation in the efficiency of library preparation. Therefore, three pools of primers were used for the simultaneous amplification of 923 amplicons from the 15 genes.

Ten nanograms of genomic DNA from each sample were subjected to library preparation using the Ion AmpliSeq Library Kit 2.0 (Life Technologies, Carlsbad, CA) and an AmpliSeq Custom Panel (Life Technologies, Carlsbad, CA) designed to specifically amplify the entire coding region of the 15 genes, with a coverage of 99.26%. Samples were barcoded using IonXpress Barcode Adapters (Life Technologies, Carlsbad, CA). Automated template preparation and enrichment using an input of 30pM of purified library were performed in an Ion Chef System (Life Technologies, Carlsbad, CA) and sequencing was conducted in an Ion Torrent PGM (Life Technologies, Carlsbad, CA) using the Ion PGM Hi-Q View Sequencing Kit (Life Technologies, Carlsbad, CA).

The genetic ancestry of a subset of patients ($n = 11$) was determined using AIMs (Ancestry Informative Markers) as previously reported^{31–35}. A more detailed description of the methodology used is in the Supplementary Methods.

Data analysis. Sequencing results obtained in the Ion Torrent PGM run were processed in the Torrent Server. Reads generated were aligned to hg19 human reference genome using the Torrent Mapping Alignment Program (TMAP) and the Torrent Variant Caller (TVC) plugin version v5.8.0.19 (464) was used to call for variants. Variants were annotated using wANNOVAR (<http://wannovar.wglab.org/>) and Cancer Genome Interpreter³⁶. After quality filters were applied per sample (average depth of coverage of at least 600x, at least 80% of reads aligned to the target region and at least 100,000 mapped reads)³⁷, variants with the following characteristics were selected: non-synonymous and frameshift variants occurring within exons; variants with a depth of coverage of at least 200x; variants with an allele frequency of at least 10%; frequency observed in population databases of germline variants (such as 1000 genomes, ExAC and ESP6500) lesser than 1%; classified as drivers or predicted drivers using the CGI tool (<https://www.cancergenomeinterpreter.org/home>); and previously reported at the COSMIC (<https://cancer.sanger.ac.uk/cosmic>) and/or TCGA databases (<http://www.cbioportal.org/>).

All remaining variants were manually confirmed on the Integrative Genomics Viewer (IGV) version 2.7. We further explored *TP53* status by using the Evolutionary Action score of *TP53*-coding variants (EAp53) to stratify patients with tumors harboring *TP53* mutations as high or low risk for unfavorable outcome, based on a model of the genotype-phenotype relationship described in^{38,39}.

Tobacco, HPV status and mutational data from 39 OpSCC patients evaluated in The Cancer Genome Atlas (TCGA) were accessed with the online Xena Browser (<https://xenabrowser.net>). Mutation frequencies of the 15 genes tested in this study were compared between both cohorts.

All NGS data generated from the 51 OpSCC samples evaluated here was deposited in the European Genome-phenome Archive (EGA) under study/dataset identifiers EGAS00001004430/EGAD00001006151.

Statistical analysis. Statistical analysis was performed using the software IBM SPSS Statistics 23 for Windows. Categorical variables were compared using Fisher's exact or Chi-square test. Survival curves were calculated by Kaplan-Meier method and differences between groups were compared using the log-rank test. Recurrence-free survival was defined as the interval between the date of initial treatment and the diagnosis of recurrence. For all analysis, we considered statistical significance when p value < 0.05 .

Results

Clinical description of the study population. Clinical and histopathological data of the 51 oropharyngeal cancer patients enrolled in this study are presented in Table 1. Most of the patients profiled in this cohort were male (98.0%) with age ranging from 35 to 76 years (mean/median = 55.3/55.0 years). Tobacco and alcohol consumption were self-reported by 86.3% and 96.0% of the cases, respectively. The majority of patients had advanced disease (according to the AJCC TNM 8th edition⁴⁰) at diagnosis (84.3%), T3/T4 tumors (37/51, 72.5%) and with clinically positive lymph nodes at diagnosis (N+; 45/51, 88.2%). All patients were treated with a platinum-based chemotherapy concomitant to radiotherapy between 2009 and 2017.

The median follow-up was 26 months (range: 1.00 to 86.00 months) and 44.3% of the cases were alive in 5-years. Recurrence occurred in 14 cases (27.5%). Tobacco smoking was significantly associated with reduced recurrence-free survival (50.2% versus 100% for never smokers; log-rank $p = 0.045$), while HPV-status association with increased recurrence-free survival was marginally significant (75.5% for HPV-positive versus 48.8% for HPV-negative; log-rank $p = 0.082$) (Supplementary Fig. 1A,B).

The expression of p16 protein was detected in 18/51 samples (35.3%). For a small subset of cases with available DNA ($n = 12$), droplet digital PCR (ddPCR) was used for the detection of HPV-DNA for types HPV-16 and HPV-18 as part of another study (unpublished data). The agreement between p16-IHC and HPV-DNA results was of 91.7% (Cohen's Kappa Value = 0.833; $p = 0.003$): 6 were p16+ /HPV-DNA+, 5 were p16-/HPV-DNA-, but 1 had discordant a result, being p16+ /HPV-DNA- (Supplementary Table 1). Given the high agreement between p16-IHC and HPV-DNA results in the subset of samples tested, and the recommendation of this test as a surrogate marker for HPV-positivity on a clinical setting⁴¹, we considered patients with tumor p16-IHC expression as positive for HPV infection in this study.

We also included a set of 39 OpSCC with available mutational, tobacco consumption and HPV status (by p16-IHC and/or *in-situ* hybridization) from the TCGA study recovered from the Xena database (Supplementary Table 2). Of the 39 OpSCC cases, 79.5% (31/39) were HPV-positive and 69.2% were current or former smokers (26/39). Therefore, our cohort had a substantially lower rate of HPV-positive cases and higher prevalence of current or former tobacco smoker patients in comparison to this set of patients from the TCGA study.

Detection of somatic variants and genetic ancestry determination in OpSCC cases. After filtering-out common germline variants, only nonsynonymous variants, with VAF $> 10\%$, and previously reported by TCGA in head and neck tumors and/or in COSMIC were considered. The mutation profile identified a total of 136 variants, with an average of 2.67 variants per sample (ranging from 1–31). The majority of variants were classified as missense (117/136; 86.0%), followed by nonsense variants (16/136; 11.8%) and frameshift (3/136; 2.2%). We found an average of 1.65 (ranging from 1–12) genes mutated per sample and the majority of the patients (32/51; 62.7%) were mutated for at least one of the genes tested: 17/32 (53.1%) had only 1 gene mutated, 9/32 (28.1%) had between 2 and 5 genes mutated, and 6/32 (18.8%) had more than 5 genes mutated.

The frequency of mutated samples per gene was as follows: 45.1% (23/51) for *TP53*, 21.6% (11/51) for *NOTCH1*, 17.6% (9/51) for *FAT1*, 13.7% (7/51) for *NSD1* and *PIK3CA*, 11.8% (6/51) for *CDKN2A*, 9.9% (5/51) for *RBI*, 7.8% (4/51) for *KMT2D* and *PTEN*, 5.9% (3/51) for *HRAS*, 3.9% (2/51) for *IRF6* and *FBXW7*, and 2.0%

Characteristic	Number of cases	%
Gender		
Male	50	98.0
Female	1	2.0
Age		
<=55 years	26	51.0
>55 years	25	49.0
Smoking status		
Never	7	13.7
Current or Former	44	86.3
Alcohol consumption status		
Never	2	4.0
Current or Former	48	96.0
HPV status		
Negative	33	64.7
Positive	18	35.3
T stage		
T1/T2	14	27.5
T3/T4	37	72.5
N stage		
N0	6	11.8
N+	45	88.2
Stage*		
I/II - early	8	15.7
III/IV - advanced	43	84.3
Response to treatment		
Non-responder	19	38.8
Responder	30	61.2
Recurrence		
No	37	72.5
Yes	14	27.5
Live Status		
Alive without disease	25	49.0
Alive with disease	3	5.9
Death by cancer	23	45.1

Table 1. Clinical and pathological features of oropharyngeal patients (n = 51). *According to the 8th edition of the AJCC TNM staging system of oropharyngeal tumor 40.

(1/51) for *EZH2* (Fig. 1 and Supplementary Fig. 2). None of the patients presented mutations for *CASP8* and *TP63*.

For *TP53*, the most frequently mutated gene in our cohort, most of the mutations were missense (28/51), 5 were truncating (3 frameshift and 2 nonsense) and the majority occurred within the p53 DNA-binding domain (between codons 95 and 288) (Supplementary Fig. 3A). *NOTCH1* was the second most frequently mutated gene in our cohort (11/51) and all variants observed were missense with a high fraction happening in the N-terminal epidermal growth factor (EGF)-like ligand binding domains and another fraction in the C-terminal end, mainly around and within the Ankyrin repeat domain (Supplementary Fig. 3B).

In the set of OpSCC patients from TCGA evaluated, 64.1% of the samples had mutation in at least 1 of the genes tested (25/39), a very similar rate to the one we observed in our cohort (32/51; 62.7%). *TP53* was the most frequently mutated gene (9/39; 23.1%), followed by *PIK3CA* (8/39; 20.5%), *RB1* (5/39; 12.8%), *KMT2D* (4/39; 10.3%), *NOTCH1* and *NSD1* (3/39; 7.7%), *FAT1*, *FBXW7* and *PTEN* (2/39; 5.1%), and *CDKN2A* (1/39; 2.6%) (Supplementary Fig. 2 and Supplementary Table 2). No mutations in *CASP8*, *EZH2*, *HRAS*, *IRF6* and *TP63* were observed.

Besides somatic variants characterization, for a subset of patients (n = 9) the genetic ancestry was determined (Supplementary Table 1). A great admixture was observed in the genetic composition of the individuals tested based on the four ancestry groups evaluated. As expected, a high proportion of European ancestry was observed (median: 63.8%, range: 42.0–88.5%), followed by African (median: 22.2%, range: 5.5–38.6%), Native American (median: 6.2%, range: 2.0–28.7%) and East Asian (median: 5.0%, range 3.8–23.9%) (Supplementary Table 1 and Supplementary Fig. 4).

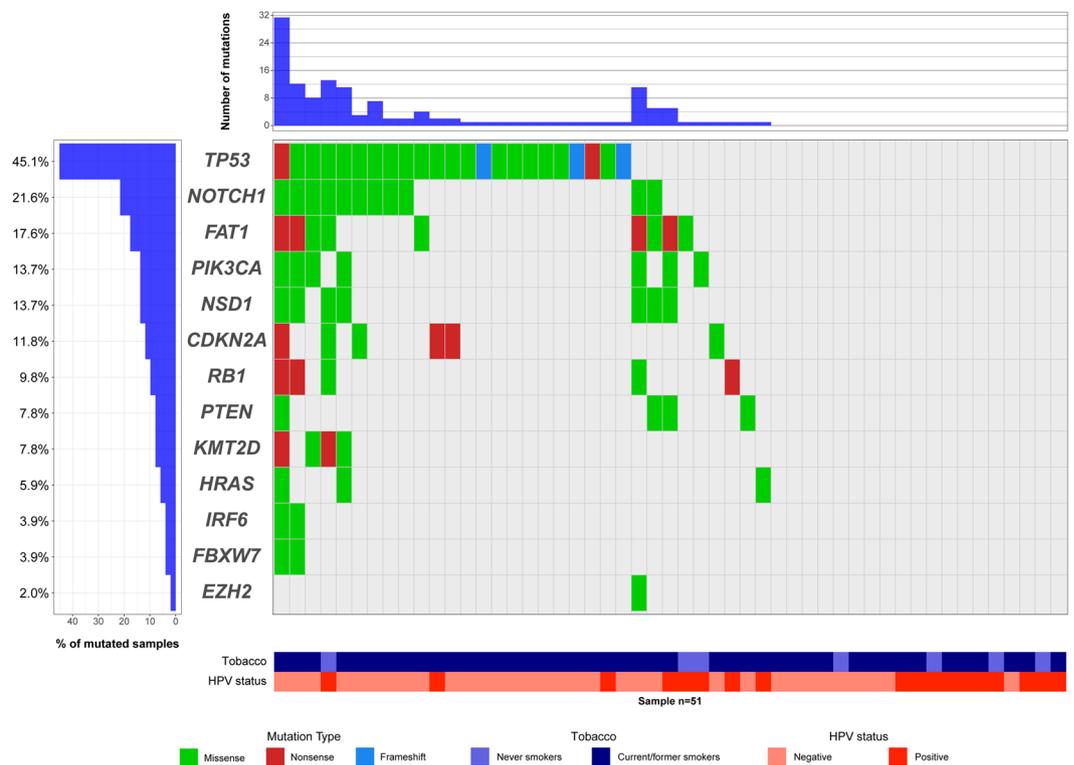


Figure 1. Oncoprint diagram with frequencies and types of mutation observed in OpSCC. Genes (rows) are sorted according to the frequency of mutation within samples ($n = 51$). Samples (columns) are further classified according to tobacco consumption and HPV status. The right panel represents mutations in the genes tested. Top, number of mutations per sample. Color codes indicate mutation type, tobacco and HPV status. The oncoprint was generated using the package GenVisR (version 1.14) in R 3.5.0 software.

		HPV status		Tobacco consumption			
Genes		Negative	Positive	p-value ⁺	Never	Current/Former	p-value ⁺
Any gene	WT	9 (27.3)	10 (55.6)	0.046*	4 (57.1)	15 (34.1)	0.402
	mutated	24 (72.7)	8 (44.4)		3 (42.9)	29 (65.9)	
CDKN2A	WT	29 (87.8)	16 (88.9)	1.000	6 (85.7)	39 (88.6)	1.000
	mutated	4 (12.2)	2 (11.1)		1 (14.3)	5 (11.4)	
FAT1	WT	27 (81.8)	15 (83.3)	1.000	5 (71.4)	37 (84.1)	0.592
	mutated	6 (18.2)	3 (16.7)		2 (28.6)	7 (15.9)	
NOTCH1	WT	23 (69.7)	17 (94.4)	0.072	6 (85.7)	34 (77.3)	1.000
	mutated	10 (30.3)	1 (5.6)		1 (14.3)	10 (22.7)	
PIK3CA	WT	28 (84.8)	16 (88.9)	1.000	6 (85.7)	38 (86.4)	1.000
	mutated	5 (15.2)	2 (11.1)		1 (14.3)	6 (13.6)	
TP53	WT	13 (39.4)	15 (83.3)	0.003*	6 (85.7)	22 (50.0)	0.112
	mutated	20 (60.6)	3 (16.7)		1 (14.3)	22 (50.0)	

Table 2. Results of the association between HPV status and tobacco consumption with the status of somatic gene mutation in the OpSCC patients. ⁺Results obtained using the Fisher's Exact test, except for* that were obtained with the Chi-square test.

Association between mutation status, HPV status and tobacco consumption. The average number of variants in HPV-negative samples was higher than in HPV-positive cases (3.33 ± 6.04 versus 1.44 ± 3.22 ; $p = 0.199$) (Supplementary Fig. 5); and the majority of mutated cases were HPV-negative (24/32, 75.0% versus 8/32, 25.0% for HPV-positive; Chi-square $p = 0.046$). As expected, genes commonly associated with tobacco-induced HNC carcinogenesis had a lower frequency of mutation in HPV-positive cases (Table 2): 1/18 (5.6%, $p = 1.000$) for *CDKN2A*, 3/18 (16.7%; $p = 1.000$) for *FAT1*, 1/18 (5.6%; $p = 0.077$) for *NOTCH1*, and 3/18 (16.7%; $p = 0.003$) for *TP53* (Fig. 1 and Supplementary Fig. 6). Interestingly, of the 8 HPV-positive cases harboring mutations, 5 were current or former smokers; and only 2/18 (11.1%) showed mutation in *PIK3CA*, reportedly the most frequently mutated gene in this group (Fig. 1 and Supplementary Fig. 6).

Characteristic	Number of cases	Number of events	5-year recurrence-free survival	p-value (log-rank)
HPV status				
Negative	33	11	48.8%	0.082
Positive	18	3	75.5%	
Tobacco history				
never	7	0	100.0%	0.045
current/former	44	14	50.2%	
Mutation status (any gene)				
WT	19	3	80.0%	0.156
mutated	32	11	47.9%	
PTEN status				
WT	47	11	64.5%	0.045
mutated	4	3	25.0%	
NOTCH1 status				
WT	40	8	68.2%	0.032
mutated	11	6	35.1%	
TP53 status				
WT	28	5	75.9%	0.137
mutated	23	9	42.6%	
EAp53				
WT/low-risk	37	7	71.8%	0.080
high-risk	11	6	38.4%	

Table 3. Recurrence-free survival information according to selected clinical and molecular factors.

Regarding tobacco consumption, the average number of variants was slightly higher in current or former tobacco smokers when compared with never smokers (2.75 ± 5.38 versus 2.14 ± 4.81 ; $p = 0.780$) (Supplementary Fig. 5). Additionally, most of the cases mutated for at least one of the genes tested were current or former smokers (29/32, 90.6% versus 3/32, 9.4% for never smokers; Fisher's Exact Test $p = 0.402$). Although not statistically significant, genes commonly associated with tobacco-induced HNC carcinogenesis were more frequently mutated in current or former smokers (Table 2): 10/44 (22.4%; $p = 0.592$) for *NOTCH1*, and 22/44 (50.0%; $p = 0.112$) for *TP53* (Fig. 1 and Supplementary Fig. 6). Although the overall frequency of mutations in *CDKN2A* and *FAT1* was slightly higher in never smokers, most of the mutated samples for these genes were current or former smokers: 5/6 (83.3%; $p = 1.000$) mutated for *CDKN2A* and 7/9 (77.8%; $p = 1.000$) for *FAT1* (Fig. 1 and Supplementary Fig. 6).

Interestingly, of the 19 cases for which no mutation was detected for the genes tested, all but one patient (18/19; 94.7%) were exposed to at least one of the risk factors evaluated (tobacco smoking and/or positivity for HPV infection): 3 (15.8%) cases were never-smokers, but HPV-positive; 8 were smokers, but HPV-negative; and 7 were smokers and HPV-positive (Fig. 1 and Supplementary Fig. 6).

Additionally, we compared the mutational rates of the 15 genes in our cohort and the 39 OpSCC from the TCGA study. Although the frequency of mutated samples was similar in both cohorts, as described earlier, significant differences in mutation frequencies were observed for specific genes (Supplementary Fig. 2). Genes commonly associated with tobacco-induced HNC carcinogenesis were less frequently mutated in the TCGA samples, namely *TP53* (9/39; 23.1%), *NOTCH1* (3/39; 7.7%), *FAT1* (2/39; 5.1%), and *CDKN2A* (1/39; 2.6%); while *PIK3CA* was mutated in 20.5% of the samples (8/39). This agrees with the lower prevalence of ever smokers (69.2% versus 86.3%) and higher prevalence of HPV-positive cases in the TCGA cohort (79.5% versus 35.3%; Supplementary Table 2).

Association between mutation status and clinical and outcome data. We tested whether the presence of mutation in the genes evaluated could be associated with the following demographical and clinical features: age at diagnosis, gender, tobacco consumption, drinking status, HPV status, T stage, N status and clinical stage. The only statistically significant association found was between *NOTCH1* mutation and tumor size: all tumors with *NOTCH1* mutation (11/11; $p = 0.023$) had a higher T-stage (T3/T4).

Next, we constructed Kaplan-Meier curves to assess the impact of gene mutations in recurrence-free survival of the patients tested. Although not statistically significant, the presence of mutation in at least one of the genes tested showed a decrease in recurrence-free survival (47.9% versus 80%; log rank p-value = 0.156), the same was observed for the presence of *TP53* (42.6% versus 75.9% for WT; log rank p-value = 0.137) (Table 3).

The presence of mutation in *NOTCH1*, significantly decreased the recurrence-free survival (RFS) of mutated cases (25.1% for mutated cases and 68.2% for WT; log-rank p-test = 0.032) (Table 3 and Fig. 2A). In a multivariate hazard ratio analysis for recurrence, adjusted by HPV and tobacco status, clinical stage, age at diagnosis, *TP53*, *PTEN* and *NOTCH1* mutation status, only *NOTCH1* mutation remained significant after stepwise selection, with a risk of recurrence of 4.503 for mutated cases (HR 95% CI = 1.112–14.572; Cox Regression $p = 0.034$) (Fig. 3).

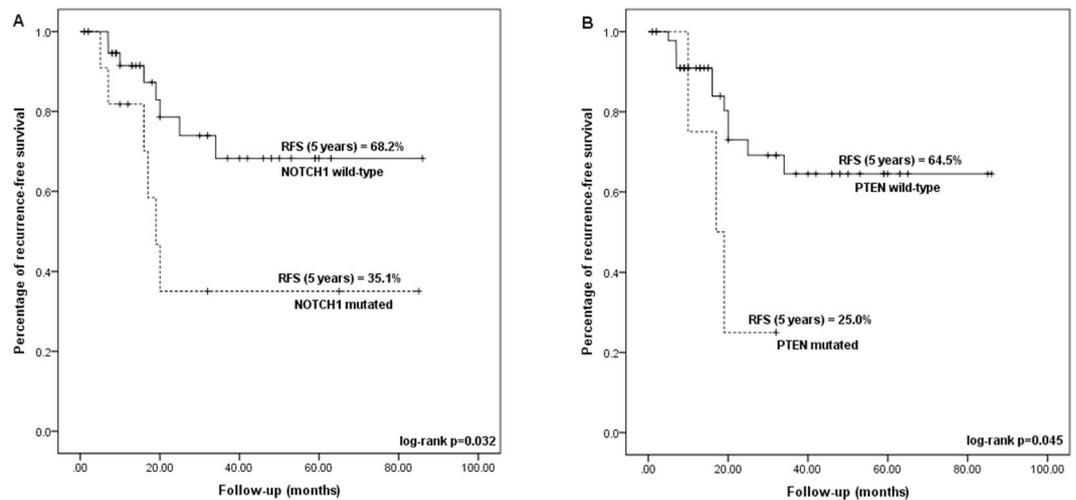


Figure 2. Kaplan Meier curves indicating the difference in 5-year recurrence-free survival (RFS) according to NOTCH1 (A) and PTEN (B) mutation status. The KM curves were generated using the software IBM SPSS Statistics Version 23.

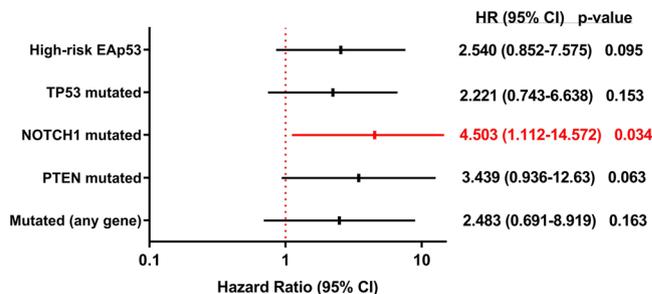


Figure 3. Forest plot with the hazard ratio values and 95% CI for selected molecular alterations (p-values were obtained through Cox regression test). The forest plot was generated using the software IBM SPSS Statistics Version 23.

The presence of *PTEN* mutation was also significantly associated with decreased recurrence-free survival (RFS: 25% versus 64.5% in WT; log rank p-test = 0.045) (Table 3 and Fig. 2B). However, the risk estimate by Cox regression was not statistically significant (Fig. 3).

Breakdown of TP53 effect in clinical outcome. As mentioned earlier, 45.1% (23/51) of the samples tested harbored *TP53* mutations. Although not statistically significant, we observed that the presence of *TP53* variants reduced recurrence-free survival (42.6% versus 75.9% for WT; log rank p-value = 0.137).

We further explored *TP53* missense mutations by using the Evolutionary Action score of *TP53*-coding variants (EAp53)^{38,39} to predict the impact of *TP53* mutations on outcome: 34.8% (8/23) of the mutated cases were scored as having low-risk variants, while 65.2% had high-risk variants. High-risk variants carriers had an increased risk of 2.540-fold of having a recurrence (HR 95% CI = 0.852–7.575; p-value = 0.095) and a marginally significant decreased recurrence-free survival (38.4% versus 71.8%; log rank p-test = 0.080) (Table 3).

Discussion

Oropharyngeal squamous cell carcinomas arise from epithelial cells of the mucosal lining of the upper aerodigestive tract and, despite all cases developing from one cell type in one tissue, these tumors are remarkably heterogeneous. This heterogeneity can be further explained by differences in etiology and in molecular alterations that drive carcinogenesis⁴².

OpSCC were classically associated with heavy tobacco and alcohol consumption; however, a significant decrease in cases related to tobacco/alcohol and an increase in cases related to infection by high-risk human papillomavirus (HPV) is being observed, especially in high and middle-income countries^{2,43}. In our cohort, most of the cases were current or former-smokers (86.3%) and only 35.3% were HPV-positive (by p16-IHC); even within the HPV-positive group, the majority of cases were smokers (66.7%). This information shows that tobacco smoking still seems to have a bigger impact in the onset of OpSCC treated at our institution. Despite the small number of cases evaluated in our study, tobacco smoking was significantly associated with reduced recurrence-free

survival (RFS), while HPV-status association with increased RFS was marginally significant. It is noteworthy that the majority of HPV-positive cases in our cohort were current or former smokers (12/18; 66.7%). HPV-positive OpSCC has a clearly improved outcome; however, smoking has a reportedly adverse effect on prognosis in both HPV-positive and HPV-negative cases⁶. This may have impacted the survival rates we observed, since although higher than in the HPV-negative group, the difference was not statistically significant.

Brazilian population has been frequently characterized by a considerable admixture of different ancestries in the genetic composition within individuals. Several studies have explored this admixture in large numbers of patients with the most prevalent tumor types (breast⁴⁴, colorectal³¹ and lung⁴⁵), as well as in the healthy population³⁴. All these studies found similar results, with a confirmed mixture of ancestry markers for the populations examined and a higher proportion of European ancestry markers, followed by African and then Asian and/or Native American in most of the patients. We were unable to test all our patients for genetic ancestry, but the small subset tested confirmed these results. Previous studies have shown differences in prevalence of mutation in specific genes according to genetic ancestry, such as an enrichment of *TP53* mutations in AFR and of *PIK3CA* in EUR⁴⁶. However, statistically significant confirmation of the genetic ancestry profile of the patients evaluated in this study and testing for any associations with the prevalence of gene mutations would require the evaluation of a substantially larger number of cases.

The most recent genomic progression model for head and neck cancer, resulting from new insights into the cancer genes that are commonly mutated in these tumors, points to different routes towards mucosal squamous cell carcinoma transformation. There seem to be three main genetic subgroups with two of them closely related to the etiology: (i) HPV-related; (ii) tobacco-related; and (iii) HPV-negative/tobacco-negative⁴². Our main findings will be discussed below based on these subgroups.

For HPV-positive cases, transcriptionally-active human papillomavirus leads to cell cycle deregulation through the abrogation of p53 and pRb pathways by the viral oncoproteins E6 and E7 (HPV E6 and E7). Further oncogenic events may lead to differentiation in HPV-KRT (HPV-keratinocyte differentiation and oxidative reduction process) or HPV-IMU (HPV-immune response and mesenchymal cell differentiation) tumors⁴⁷. p16 detection by immunohistochemistry (p16-IHC) was included in the clinical routine as the gold-standard test for HPV-positive oropharyngeal SCC with proved clinical impact^{18,48} due to its easy-of-use, low cost and high sensitivity rates (near 100%)^{41,49}. However, it frequently yields not ideal specificity rates (79% and 95%), detecting a positive p16 signal in tumors not associated with HPV^{48–50}. For this reason, several studies have suggested additional HPV-specific tests to ensure accurate classification of OpSCC as HPV-related^{48,51,52}. In our study, we performed HPV-DNA detection in a subset of the cases and observed a high agreement with p16-IHC results. Moreover, molecular characterization closely agreed with other studies that also used HPV-status to describe results. HPV-positive tumors usually have a lower mutational burden, are usually WT for *TP53* mutations and frequently harbor activating *PIK3CA* mutations and amplifications (mainly the HPV-KRT)^{20,22,47,53}. In accordance to this, in our cohort, HPV-negative cases had 2.5-times more variants than HPV-positive cases (3.27 ± 5.854 versus 1.33 ± 3.162 , respectively). Moreover, even though most of our HPV-positive cases were smokers (12/18; 66.7%), the frequency of *TP53* mutation in this subgroup was significantly lower (16.7%; $p = 0.003$). Conversely, we found 2 HPV-positive case with *PIK3CA* mutation (2/18; 11.1%); a classic C > T mutation in the helical domain, frequent in viral-associated tumors (p.E545K) and commonly associated with the apolipoprotein B mRNA-editing enzyme catalytic subunit (APOBEC)-induced mutational signature^{24,54}. Hayes and colleagues reported in a review article a frequency between 22–56% of *PIK3CA* activation in HPV-positive HNC; however, we only evaluated somatic mutations in the present study and did not explore copy-number alterations (CNA). This might explain the lower frequency in comparison to other studies that relate this activation both to mutation and amplification in HPV-positive cases⁵³. As expected, besides *TP53*, genes frequently mutated in tobacco-related HNC had a low frequency of mutation in the HPV-positive group, namely *NOTCH1* (1/18; 5.6%), *CDKN2A* (2/18; 11.1%) and *FAT1* (3/18; 16.7%)^{53,55}.

The second and most classic subgroup of HNC is associated with tobacco-smoking and also involves deregulation of the cell cycle mainly through loss-of-function (LoF) of two tumor suppressor genes: *TP53* and *CDKN2A*. *CDKN2A* encodes the p16INK4A protein and is lost through deletion, inactivating mutations and hypermethylation in 15–22% of HNC⁵³. Together with the frequent amplification of cyclin D1 (*CCND1*), also common in tobacco-associated HNC, *CDKN2A* LoF contributes to unscheduled DNA replication and leads to DNA damage and p53 activation⁵⁶; however, *TP53* inactivation is also found in 60–80% of HNC cases^{5,24}, thus abrogating cell cycle arrest and apoptosis⁵⁷. Consequently, these tumors have a higher mutational burden and frequent copy number alterations⁴². In our cohort, we found an overall rate of *CDKN2A* and *TP53* mutation of 9.8% and 45.1%, respectively. Dogan and colleagues found similar rates, when using a targeted exome sequencing approach to characterize 83 HPV-negative and 74 HPV-positive OpSCC⁵⁸: 49% of the cases were mutated for *TP53*, while 22% harbored *CDKN2A* mutations and deletions. Moreover, we did not test for *CDKN2A* inactivation through deletion or hypermethylation. However, among mutated cases, these genes were more frequently mutated in current/former smokers with rates of 83.3% for *CDKN2A* and 95.7% for *TP53*. The breakdown rates according to HPV status were: 78%/16% for *TP53* mutations in HPV-negative/positive cases and 39% versus 4% for *CDKN2A* mutations in HPV-negative/positive cases. It is already known that the presence of *TP53* mutations has an important impact in disease progression, treatment response, specially to platinum-based therapy, and survival^{59–62}. Although not statistically significant, we observed that the presence of *TP53* variants reduced recurrence-free survival (42.6% versus 75.9% for WT; log rank p -value = 0.137). Next, we explored whether the algorithm Evolutionary Action (EAp53), that stratifies patients according to *TP53* variants associated with especially poor outcomes^{38,39}, could be used as a classificatory in our cohort. The association of *TP53* variants classified with EAp53 still did not reach significance, but it is clear that, for our patients, high-risk variants carriers had an increased risk of 2.540-fold of having a recurrence (HR 95% CI = 0.852–7.575; p -value = 0.095) and a marginally significant decreased recurrence-free survival (38.4% versus 71.8%; log rank p -test = 0.080). Previous

studies have tested this algorithm in HPV-negative HNC and validated the association of high-risk *TP53* variants, as classified by EAp53, with decreased sensitivity to cisplatin, decreased survival and increased distant metastases in HNC^{38,39}. Here we focused our analysis in OpSCC patients with a significant representation of HPV-positive cases and treated by chemoradiation. These differences should be taken into consideration in the interpretation of the algorithm results.

After cell-cycle abrogation, HPV-negative/tobacco-positive cases usually acquire more alterations that drive different routes towards the progression of these tumors. These pathways usually involve squamous differentiation, oxidative stress and WNT signaling. Between 10–21% of HNC demonstrate LoF mutations in *NOTCH1*, impacting squamous differentiation and cell polarity, while 5–23% of cases harbor mutations in *FAT1*, an important component of the WNT signaling⁵³. This route seems to impact beta-catenin signaling and keratinocyte transformation, and tobacco-smoking is a known risk factor⁴². We found similar rates of mutated cases for these genes in our cohort, and the majority of cases mutated for *NOTCH1* and *FAT1* were current or former smokers. Moreover, we found an association between *NOTCH1* mutation and a higher T stage, consistent with previous studies showing that functional *NOTCH1* inhibits proliferation in SCC cells and that loss of canonical NOTCH increases tumorigenesis in both HPV-positive and p53-mutant mice^{21,63}. In addition, the presence of *NOTCH1* mutations significantly impacted recurrence-free survival. Emerging evidence indicates that Notch effects are dependent on the cellular context in which it is activated, with aberrant Notch signaling being associated with cancer recurrence, metastasis and treatment resistance in different tumor sites⁶⁴. In a recent study, Lim and colleagues tested the feasibility of using targeted NGS to guide treatment of HNC patients, among the genes with association with a poorer overall survival they found *NOTCH1* mutation as a predictor of worse outcome⁶⁵. Similarly, Dong and colleagues found that in HPV-positive OpSCC cases, the presence of *NOTCH1* mutations contributed to a worse overall-survival⁵⁸. Vettore and colleagues also found an association between the presence of mutations in genes of the Notch pathway and a lower disease-free survival in patients with oral tongue carcinoma⁶⁶. In a cohort of esophageal carcinomas from China, Song and colleagues found that patients with mutations in *NOTCH1* had shorter survival and failed to respond to chemotherapy⁶⁷. These results agree with our findings that the status of *NOTCH1* mutations is a promising predictive biomarker for patient outcome and treatment response in OpSCC.

Finally, *PTEN* is a tumor suppressor gene mutated in 9–23% of HPV-negative HNC, leading to oncogenic activity through the activation of the PI3K/AKT/mTOR signaling⁴². We found *PTEN* mutations in 7.8% of the cases tested; once again, this rate does not include CNAs. Interestingly, we observed a significant decrease in recurrence-free survival in the presence of *PTEN* mutation (25% versus 64.5% in WT; log rank p-test = 0.045). A recent study found that patients with high-expressing *PTEN* had an improved progression-free survival in response to cetuximab, in comparison to patients with low expression of this gene⁶⁸. A similar scenario was observed in a Brazilian study on HNC patients treated with cetuximab, with a worse progression-free and overall-survival in patients with loss of *PTEN*⁶⁹.

To our knowledge, this is the first study showing the association between *NOTCH1* and *PTEN* mutations and survival in OpSCC patients treated with platin-based chemotherapy plus radiation.

The third subgroup of HNC comes mainly from previous studies that identified a subset of HNC with a strong correlation between gain of function *HRAS* mutations and LoF mutation in *CASP8*^{21,42,70}. This subgroup is usually described as diploid, as having fewer CNAs, DNA mismatch repair proficient and to occur more frequently in women with oral cavity tumors without a history of alcohol and smoking consumption^{21,70}. The prevalence rates of this subgroup and the clinical impact is yet to be studied⁴². We did not detect any case with *CASP8* variants in our cohort; moreover, all 3 cases identified with *HRAS* mutation were current or former smokers and one of them was also HPV-positive.

Although differences in the mutation rates of specific genes were observed, mostly due to different prevalence of risk factors (HPV and tobacco smoking), when comparing the results on our cohort and the TCGA OpSCC samples explored, the frequency of cases mutated for at least one of the genes was similar (62.7% versus 64.1%). In the TCGA publication with genomic data on head and neck cancer, only 14% of the samples tested did not present mutation in any gene²⁴. Besides including data from all subsites within head and neck cancer, the mutation data reported in the article compiled results from whole exome sequencing, while we only tested a panel of a limited number of genes. This can explain the similarities between mutation rates in the two cohorts of OpSCC samples explored in this article in comparison to the frequency reported in the TCGA publication.

Recently, Perdomo and collaborators, evaluated a similar panel of genes in a cohort comprised of 180 HNC from 3 multicentric studies from South America and Europe³⁰. The study reports that 25% of the cases were from the oropharynx (45/180), and only 8% of the entire cohort was HPV-positive (15/180)³⁰. Therefore, their cohort was enriched for HPV-negative cases, and, although the number of OpSCC tested was similar to ours, only around 8% were HPV-positive, in contrast to our 35.3% frequency of HPV-positivity. This might explain the slightly lower rates of *TP53* mutations and other genes associated with tobacco-smoking in our cohort in comparison to their findings. In addition, they also explored somatic CNA, that are also included in their higher rates.

Our study has some limitations being the small number of cases evaluated and the short follow-up (median of 26 months) possibly the most important ones, directly impacting the statistical significance of the molecular findings in patient outcome. Besides that, we also did not evaluate copy-number alterations of the genes included, hindering comparisons to other genomic studies that include amplification and deletion in the mutational frequencies. It is important to point out that there is a lack of genomic studies including Brazilian cases; therefore, our study focused in the profile of somatic variants of a set of genes in a cohort of Brazilian OpSCC patients submitted to a homogeneous treatment and with sufficient and well-annotated clinical data. Despite of differences in geographic, economic and social habits when compared to other populations, the results were comparable to those reported before and new insights into possible molecular mechanisms associated with outcome were suggested.

Conclusions

In conclusion, we report here the results on the genetic characterization of a set of HNC-related genes in a cohort of OpSCC patients. Some of the findings suggest the clinical relevance of genomic approaches to better classify subgroups of patients with different outcomes. Validation of these data could lead to a more refined prognostic stratification and contribute to a personalized treatment approach of OpSCC patients.

Ethics approval and consent to participate. Informed consent was obtained from each individual prior to sample collection and the study protocol was approved by the Barretos Cancer Hospital Institutional Review Board.

Data availability

All data and materials are available upon request to the Corresponding Author.

Received: 18 February 2020; Accepted: 21 May 2020;

Published online: 19 June 2020

References

- Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, <https://doi.org/10.3322/caac.21492> (2018).
- Shield, K. D. *et al.* The global incidence of lip, oral cavity, and pharyngeal cancers by subsite in 2012. *CA: a cancer journal for clinicians* **67**, 51–64, <https://doi.org/10.3322/caac.21384> (2017).
- Gillison, M. L. *et al.* Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. *Journal of the National Cancer Institute* **100**, 407–420, <https://doi.org/10.1093/jnci/djn025> (2008).
- Leemans, C. R., Snijders, P. J. F. & Brakenhoff, R. H. Publisher Correction: The molecular landscape of head and neck cancer. *Nature reviews. Cancer* **18**, 662, <https://doi.org/10.1038/s41568-018-0057-9> (2018).
- Leemans, C. R., Braakhuis, B. J. & Brakenhoff, R. H. The molecular biology of head and neck cancer. *Nature reviews. Cancer* **11**, 9–22, <https://doi.org/10.1038/nrc2982> (2011).
- Ang, K. K. *et al.* Human papillomavirus and survival of patients with oropharyngeal cancer. *The New England journal of medicine* **363**, 24–35, <https://doi.org/10.1056/NEJMoa0912217> (2010).
- Brierley, J. e., Gospodarowicz, M. K. e. & Wittekind, C. e. *TNM classification of malignant tumours*. Eighth edition. edn.
- Dansky Ullmann, C., Harlan, L. C., Shavers, V. L. & Stevens, J. L. A population-based study of therapy and survival for patients with head and neck cancer treated in the community. *Cancer* **118**, 4452–4461, <https://doi.org/10.1002/cncr.27419> (2012).
- Haddad, R., Annino, D. & Tishler, R. B. Multidisciplinary approach to cancer treatment: focus on head and neck cancer. *Dental clinics of North America* **52**, 1–17, vii, <https://doi.org/10.1016/j.cden.2007.10.005> (2008).
- Cramer, J. D., Burtness, B., Le, Q. T. & Ferris, R. L. The changing therapeutic landscape of head and neck cancer. *Nature reviews. Clinical oncology* **16**, 669–683, <https://doi.org/10.1038/s41571-019-0227-z> (2019).
- Gatta, G. *et al.* Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: The EURO CARE-5 population-based study. *Eur J Cancer* **51**, 2130–2143, <https://doi.org/10.1016/j.ejca.2015.07.043> (2015).
- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J Clin* **68**, 7–30, <https://doi.org/10.3322/caac.21442> (2018).
- Abraham, R. *et al.* Predictors of Survival After Head and Neck Squamous Cell Carcinoma in South America: The InterCHANGE Study. *JCO Glob Oncol* **6**, 486–499, <https://doi.org/10.1200/GO.20.00014> (2020).
- Hoffman, H. T., Karnell, L. H., Funk, G. F., Robinson, R. A. & Menck, H. R. The National Cancer Data Base report on cancer of the head and neck. *Arch Otolaryngol Head Neck Surg* **124**, 951–962, <https://doi.org/10.1001/archotol.124.9.951> (1998).
- Boysen, M., Natvig, K., Winther, F. O. & Tausjo, J. Value of routine follow-up in patients treated for squamous cell carcinoma of the head and neck. *J Otolaryngol* **14**, 211–214 (1985).
- D'Souza, G. *et al.* Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med* **356**, 1944–1956, <https://doi.org/10.1056/NEJMoa065497> (2007).
- Benson, E., Li, R., Eisele, D. & Fakhry, C. The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas. *Oral Oncol* **50**, 565–574, <https://doi.org/10.1016/j.oraloncology.2013.09.008> (2014).
- Huang, S. H. & O'Sullivan, B. Overview of the 8th Edition TNM Classification for Head and Neck Cancer. *Curr Treat Options Oncol* **18**, 40, <https://doi.org/10.1007/s11864-017-0484-y> (2017).
- Lui, V. W. *et al.* Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. *Cancer discovery* **3**, 761–769, <https://doi.org/10.1158/2159-8290.CD-13-0103> (2013).
- Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157–1160, <https://doi.org/10.1126/science.1208130> (2011).
- Pickering, C. R. *et al.* Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers. *Cancer discovery* **3**, 770–781, <https://doi.org/10.1158/2159-8290.CD-12-0537> (2013).
- Agrawal, N. *et al.* Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* **333**, 1154–1157, <https://doi.org/10.1126/science.1206923> (2011).
- Mountzios, G., Rampias, T. & Psyrri, A. The mutational spectrum of squamous-cell carcinoma of the head and neck: targetable genetic events and clinical impact. *Annals of oncology: official journal of the European Society for Medical Oncology* **25**, 1889–1900, <https://doi.org/10.1093/annonc/mdu143> (2014).
- Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582, <https://doi.org/10.1038/nature14129> (2015).
- van Beers, E. H. *et al.* A multiplex PCR predictor for aCGH success of FFPE samples. *Br J Cancer* **94**, 333–337, <https://doi.org/10.1038/sj.bjc.6602889> (2006).
- Begum, S., Gillison, M. L., Ansari-Lari, M. A., Shah, K. & Westra, W. H. Detection of human papillomavirus in cervical lymph nodes: a highly effective strategy for localizing site of tumor origin. *Clinical cancer research: an official journal of the American Association for Cancer Research* **9**, 6469–6475 (2003).
- Singhi, A. D. & Westra, W. H. Comparison of human papillomavirus *in situ* hybridization and p16 immunohistochemistry in the detection of human papillomavirus-associated head and neck cancer based on a prospective clinical experience. *Cancer* **116**, 2166–2173, <https://doi.org/10.1002/cncr.25033> (2010).
- de Carvalho, A. C. *et al.* Clinical and Molecular Characterization of Surgically Treated Oropharynx Squamous Cell Carcinoma Samples. *Pathology oncology research: POR*, <https://doi.org/10.1007/s12253-018-0462-0> (2018).
- Hanna, G. J. *et al.* Plasma HPV cell-free DNA monitoring in advanced HPV-associated oropharyngeal cancer. *Ann Oncol* **29**, 1980–1986, <https://doi.org/10.1093/annonc/mdy251> (2018).
- Perdomo, S. *et al.* Genomic analysis of head and neck cancer cases from two high incidence regions. *PLoS one* **13**, e0191701, <https://doi.org/10.1371/journal.pone.0191701> (2018).

31. Duraes, R. O. *et al.* Role of Genetic Ancestry in 1,002 Brazilian Colorectal Cancer Patients From Barretos Cancer Hospital. *Front Oncol* **10**, 145, <https://doi.org/10.3389/fonc.2020.00145> (2020).
32. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular ecology notes* **7**, 574–578, <https://doi.org/10.1111/j.1471-8286.2007.01758.x> (2007).
33. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
34. Campanella, N. C. *et al.* Optimization of a pentaplex panel for MSI analysis without control DNA in a Brazilian population: correlation with ancestry markers. *Eur J Hum Genet* **22**, 875–880, <https://doi.org/10.1038/ejhg.2013.256> (2014).
35. Giolo, S. R. *et al.* Brazilian urban population genetic structure reveals a high degree of admixture. *European journal of human genetics: EJHG* **20**, 111–116, <https://doi.org/10.1038/ejhg.2011.144> (2012).
36. Tamborero, D. *et al.* Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome medicine* **10**, 25, <https://doi.org/10.1186/s13073-018-0531-8> (2018).
37. de Leng, W. W. *et al.* Targeted Next Generation Sequencing as a Reliable Diagnostic Assay for the Detection of Somatic Mutations in Tumours Using Minimal DNA Amounts from Formalin Fixed Paraffin Embedded Material. *PLoS one* **11**, e0149405, <https://doi.org/10.1371/journal.pone.0149405> (2016).
38. Neskey, D. M. *et al.* Evolutionary Action Score of TP53 Identifies High-Risk Mutations Associated with Decreased Survival and Increased Distant Metastases in Head and Neck Cancer. *Cancer research* **75**, 1527–1536, <https://doi.org/10.1158/0008-5472.CAN-14-2735> (2015).
39. Osman, A. A. *et al.* Evolutionary Action Score of TP53 Coding Variants Is Predictive of Platinum Response in Head and Neck Cancer Patients. *Cancer research* **75**, 1205–1215, <https://doi.org/10.1158/0008-5472.CAN-14-2729> (2015).
40. Amin, M. B. *et al.* *AJCC Cancer Staging Manual*. 8 edn, (Springer International Publishing, 2017).
41. Lewis, J. S. Jr. *et al.* Human Papillomavirus Testing in Head and Neck Carcinomas: Guideline From the College of American Pathologists. *Arch Pathol Lab Med* **142**, 559–597, <https://doi.org/10.5858/arpa.2017-0286-CP> (2018).
42. Leemans, C. R., Snijders, P. J. F. & Brakenhoff, R. H. The molecular landscape of head and neck cancer. *Nature reviews. Cancer* **18**, 269–282, <https://doi.org/10.1038/nrc.2018.11> (2018).
43. Castellsague, X. *et al.* HPV Involvement in Head and Neck Cancers: Comprehensive Assessment of Biomarkers in 3680 Patients. *Journal of the National Cancer Institute* **108**, djv403, <https://doi.org/10.1093/jnci/djv403> (2016).
44. Fernandes, G. C. *et al.* Prevalence of BRCA1/BRCA2 mutations in a Brazilian population sample at-risk for hereditary breast cancer and characterization of its genetic ancestry. *Oncotarget* **7**, 80465–80481, <https://doi.org/10.18632/oncotarget.12610> (2016).
45. Leal, L. F. *et al.* Mutational profile of Brazilian lung adenocarcinoma unveils association of EGFR mutations with high Asian ancestry and independent prognostic role of KRAS mutations. *Sci Rep* **9**, 3209, <https://doi.org/10.1038/s41598-019-39965-x> (2019).
46. Yuan, J. *et al.* Integrated Analysis of Genetic Ancestry and Genomic Alterations across Cancers. *Cancer Cell* **34**, 549–560 e549, <https://doi.org/10.1016/j.ccell.2018.08.019> (2018).
47. Zhang, Y. *et al.* Subtypes of HPV-Positive Head and Neck Cancers Are Associated with HPV Characteristics, Copy Number Alterations, PIK3CA Mutation, and Pathway Signatures. *Clinical cancer research: an official journal of the American Association for Cancer Research* **22**, 4735–4745, <https://doi.org/10.1158/1078-0432.CCR-16-0323> (2016).
48. Jordan, R. C. *et al.* Validation of methods for oropharyngeal cancer HPV status determination in US cooperative group trials. *The American journal of surgical pathology* **36**, 945–954, <https://doi.org/10.1097/PAS.0b013e318253a2d1> (2012).
49. Bishop, J. A., Lewis, J. S. Jr., Rocco, J. W. & Faquin, W. C. HPV-related squamous cell carcinoma of the head and neck: An update on testing in routine pathology practice. *Semin Diagn Pathol* **32**, 344–351, <https://doi.org/10.1053/j.semmp.2015.02.013> (2015).
50. Marur, S., D'Souza, G., Westra, W. H. & Forastiere, A. A. HPV-associated head and neck cancer: a virus-related cancer epidemic. *The Lancet. Oncology* **11**, 781–789, [https://doi.org/10.1016/S1470-2045\(10\)70017-6](https://doi.org/10.1016/S1470-2045(10)70017-6) (2010).
51. Prigge, E. S., Arbyn, M., von Knebel Doeberitz, M. & Reuschenbach, M. Diagnostic accuracy of p16(INK4a) immunohistochemistry in oropharyngeal squamous cell carcinomas: A systematic review and meta-analysis. *Int J Cancer* **140**, 1186–1198, <https://doi.org/10.1002/ijc.30516> (2017).
52. Westra, W. H. Detection of human papillomavirus (HPV) in clinical samples: evolving methods and strategies for the accurate determination of HPV status of head and neck carcinomas. *Oral Oncol* **50**, 771–779, <https://doi.org/10.1016/j.oraloncology.2014.05.004> (2014).
53. Hayes, D. N., Van Waes, C. & Seiwert, T. Y. Genetic Landscape of Human Papillomavirus-Associated Head and Neck Cancer and Comparison to Tobacco-Related Tumors. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **33**, 3227–3234, <https://doi.org/10.1200/JCO.2015.62.1086> (2015).
54. Henderson, S., Chakravarthy, A., Su, X., Boshoff, C. & Fenton, T. R. APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell reports* **7**, 1833–1841, <https://doi.org/10.1016/j.celrep.2014.05.012> (2014).
55. Seiwert, T. Y. *et al.* Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clinical cancer research: an official journal of the American Association for Cancer Research* **21**, 632–641, <https://doi.org/10.1158/1078-0432.CCR-13-3310> (2015).
56. Toledo, L., Neelsen, K. J. & Lukas, J. Replication Catastrophe: When a Checkpoint Fails because of Exhaustion. *Molecular cell* **66**, 735–749, <https://doi.org/10.1016/j.molcel.2017.05.001> (2017).
57. Levine, A. J. & Oren, M. The first 30 years of p53: growing ever more complex. *Nature reviews. Cancer* **9**, 749–758, <https://doi.org/10.1038/nrc2723> (2009).
58. Dogan, S. *et al.* Identification of prognostic molecular biomarkers in 157 HPV-positive and HPV-negative squamous cell carcinomas of the oropharynx. *International journal of cancer* **145**, 3152–3162, <https://doi.org/10.1002/ijc.32412> (2019).
59. Poeta, M. L. *et al.* TP53 mutations and survival in squamous-cell carcinoma of the head and neck. *The New England journal of medicine* **357**, 2552–2561, <https://doi.org/10.1056/NEJMoa073770> (2007).
60. Zhou, G., Liu, Z. & Myers, J. N. TP53 Mutations in Head and Neck Squamous Cell Carcinoma and Their Impact on Disease Progression and Treatment Response. *Journal of cellular biochemistry* **117**, 2682–2692, <https://doi.org/10.1002/jcb.25592> (2016).
61. Tandon, S., Tudur-Smith, C., Riley, R. D., Boyd, M. T. & Jones, T. M. A systematic review of p53 as a prognostic factor of survival in squamous cell carcinoma of the four main anatomical subsites of the head and neck. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **19**, 574–587, <https://doi.org/10.1158/1055-9965.EPI-09-0981> (2010).
62. Perrone, F. *et al.* TP53 mutations and pathologic complete response to neoadjuvant cisplatin and fluorouracil chemotherapy in resected oral cavity squamous cell carcinoma. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **28**, 761–766, <https://doi.org/10.1200/JCO.2009.22.4170> (2010).
63. Nyman, P. E., Buehler, D. & Lambert, P. F. Loss of Function of Canonical Notch Signaling Drives Head and Neck Carcinogenesis. *Clinical cancer research: an official journal of the American Association for Cancer Research* **24**, 6308–6318, <https://doi.org/10.1158/1078-0432.CCR-17-3535> (2018).
64. Brzozowa-Zasada, M. *et al.* Notch signalling pathway as an oncogenic factor involved in cancer development. *Contemporary oncology* **20**, 267–272, <https://doi.org/10.5114/wo.2016.61845> (2016).

65. Lim, S. M. *et al.* Investigating the Feasibility of Targeted Next-Generation Sequencing to Guide the Treatment of Head and Neck Squamous Cell Carcinoma. *Cancer research and treatment: official journal of Korean Cancer Association* **51**, 300–312, <https://doi.org/10.4143/crt.2018.012> (2019).
66. Vettore, A. L. *et al.* Mutational landscapes of tongue carcinoma reveal recurrent mutations in genes of therapeutic and prognostic relevance. *Genome medicine* **7**, 98, <https://doi.org/10.1186/s13073-015-0219-2> (2015).
67. Song, B. *et al.* Mutually exclusive mutations in NOTCH1 and PIK3CA associated with clinical prognosis and chemotherapy responses of esophageal squamous cell carcinoma in China. *Oncotarget* **7**, 3599–3613, <https://doi.org/10.18632/oncotarget.6120> (2016).
68. Eze, N. *et al.* PTEN loss is associated with resistance to cetuximab in patients with head and neck squamous cell carcinoma. *Oral oncology* **91**, 69–78, <https://doi.org/10.1016/j.oraloncology.2019.02.026> (2019).
69. da Costa, A. *et al.* The roles of PTEN, cMET, and p16 in resistance to cetuximab in head and neck squamous cell carcinoma. *Medical oncology* **36**, 8, <https://doi.org/10.1007/s12032-018-1234-0> (2018).
70. Smeets, S. J. *et al.* Genetic classification of oral and oropharyngeal carcinomas identifies subgroups with a different prognosis. *Cellular oncology: the official journal of the International Society for Cellular Oncology* **31**, 291–300, <https://doi.org/10.3233/CLO-2009-0471> (2009).

Acknowledgements

ACC, BPS, LMRA and LMJ were recipients of scholarship from São Paulo Research Foundation (FAPESP); WS was a recipient of a scholarship from Coordination for the Improvement of Higher Education Personnel (Capes); and RMR and ALC had a National Counsel for Technological and Scientific Development (CNPq) scholarship. This study was funded by São Paulo Research Foundation (FAPESP; grant # 2015/01286-0), National Counsel for Technological and Scientific Development (CNPq; grant # 445.053/2014-3), Barretos Cancer Hospital, and the Public Ministry of Labor Campinas (Research, Prevention, and Education of Occupational Cancer).

Author contributions

A.C.C. study design, sample selection, sample processing, sequencing experiments and analysis, clinical data collection, statistical analysis, manuscript preparation. S.P. study design, sequencing analysis, manuscript review. W.S. sequencing analysis, manuscript review. G.C.F. sequencing experiments and analysis. L.M.J. sample selection, sample processing, sequencing experiments, clinical data collection. R.S.C. sample selection, clinical data collection. C.S.N. and G.C.A. histopathological evaluations. B.P.S. sample processing and sequencing experiments. L.M.R.A. and M.E.M. study design and analysis. P.M. clinical data collection and interpretation. N.H. and R.M.R. study design and manuscript review. A.L.C. study design and coordination, manuscript review.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-66741-z>.

Correspondence and requests for materials should be addressed to A.L.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020