Oral Oncology 81 (2018) 35–44

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

Comparative genomic analysis of oral versus laryngeal and pharyngeal cancer

David M. Vossen[a,b], Caroline V.M. Verhagen[a,b], Marcel Verheij[a,c], Lodewyk F.A. Wessels[d,e], Conchita Vens[a,c,*], Michiel W.M. van den Brekel[b,f,g,*]

[a] Division of Cell Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
[b] Department of Head and Neck Oncology and Surgery, The Netherlands Cancer Institute, Amsterdam, The Netherlands
[c] Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
[d] Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Amsterdam, The Netherlands
[e] Department of EEMCS, Delft University of Technology, Delft, The Netherlands
[f] Institute of Phonetic Sciences, University of Amsterdam, Amsterdam, The Netherlands
[g] Department of Oral and Maxillofacial Surgery, Academic Medical Center, Amsterdam, The Netherlands

ARTICLE INFO

Keywords:
Head and neck squamous cell carcinoma
Oral cancer
Laryngeal neoplasms
Pharyngeal neoplasms
Genomics
DNA sequence analysis
Mutation
Homologous recombination
Chemoradiotherapy
General surgery

ABSTRACT

Objective: Locally advanced oral squamous cell carcinoma (OSCC) shows lower locoregional control and disease specific survival rates than laryngeal and pharyngeal squamous cell carcinoma (L/P-SCC) after definitive chemoradiotherapy treatment. Despite clinical factors, this can point towards a different tumor biology that could impact chemoradiotherapy response rates. This prompted us to compare the mutational profiles of OSCC with L/P-SCC.

Methods: We performed target capture DNA sequencing on 111 HPV-negative HNSCC samples (NKI dataset), 55 oral and 56 laryngeal/pharyngeal, and identified somatic point mutations and copy number aberrations. We next expanded our analysis with 276 OSCC and 134 L/P-SCC sample data from The Cancer Genome Atlas (TCGA dataset). We focused our analyses on genes that are frequently mutated in HNSCC.

Results: The mutational profiles of OSCC and L/P-SCC showed many similarities. However, OSCC was significantly enriched for *CASP8* (NKI: 15% vs 0%; TCGA: 17% vs 2%) and *HRAS* (TCGA: 10% vs 1%) mutations. *LAMA2* (TCGA: 5% vs 19%) and *NSD1* (TCGA: 7% vs 25%) mutations were enriched in L/P-SCC. Overall, we find that OSCC had fewer somatic point mutations and copy number aberrations than L/P-SCC. Interestingly, L/P-SCC scored higher in mutational and genomic scar signatures associated with homologous recombination DNA repair defects.

Conclusion: Despite showing a similar mutational profile, our comparative genomic analysis revealed distinctive features in OSCC and L/P-SCC. Some of these genes and cellular processes are likely to affect the cellular response to radiation or cisplatin. Genomic characterizations may guide or enable personalized treatment in the future.

Introduction

Definitive (chemo)radiotherapy (CRT) is a curative treatment option for inoperable, locally advanced oral squamous cell carcinoma (OSCC) [1]. However, it appears that definitive CRT in OSCC does not achieve similarly high locoregional control or disease specific survival rates as in laryngeal and pharyngeal squamous cell carcinoma (L/P-SCC) [2,3]. Current HNSCC treatment guidelines reflect this and L/P-SCC is preferably treated with definitive CRT and OSCC with surgery followed by postoperative radiotherapy with or without chemotherapy (S-PORT). Despite the influence of some clinical factors, the dissimilarity in outcome characteristics could be partly based on a different

Abbreviations: CNAs, copy number aberrations; CRT, chemoradiotherapy; FA, Fanconi anemia; HPV, human papillomavirus; HR, homologous recombination; HRD, Homologous Recombination Deficiency; L/P-SCC, laryngeal and pharyngeal squamous cell carcinoma; LST, Large Scale Transition; NtAI, Number of telomeric Allelic Imbalances; OSCC, oral squamous cell carcinoma; SPMs, somatic point mutations; S-PORT, surgery followed by postoperative radiotherapy with or without chemotherapy; TCGA, The Cancer Genome Atlas; TiTvs, transitions and transversions

* Corresponding authors at: Division of Cell Biology, The Netherlands Cancer Institute – Antoni van Leeuwenhoek (NKI-AVL), Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands (C. Vens). Department of Head and Neck Surgery, The Netherlands Cancer Institute – Antoni van Leeuwenhoek (NKI-AVL), Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands (M.W.M. van den Brekel).

E-mail addresses: c.vens@nki.nl (C. Vens), m.vd.brekel@nki.nl (M.W.M. van den Brekel).

https://doi.org/10.1016/j.oraloncology.2018.04.006
Received 24 December 2017; Received in revised form 28 March 2018; Accepted 7 April 2018

tumor biology that consequently also impacts CRT response. This led us to question whether the mutational profiles of OSCC and L/P-SCC differ.

Outcomes following S-PORT and CRT are comparable in locally advanced L/P-SCC [3–5], with the exception of T4 L-SCCs [6]. Since CRT preserves the larynx, tongue and tonsils in most patients, it is the preferred treatment for L/P-SCC. In contrast, in locally advanced OSCC, worse outcomes have been reported following CRT in comparison to S-PORT [1,2,7–10]. It should be noted that mainly inoperable OSCCs are treated with CRT, hence impeding any strong conclusion on CRT efficacy in this tumor site. Yet, some studies on operable HNSCC point towards a different CRT response of OSCC and L/PSCC [2,9]. As variation in genetic makeup in OSCC and L/P-SCC could result in altered biology and thereby CRT response, we compared their mutational profiles.

Genomics studies have identified the genes that are frequently mutated in HNSCC. Some focused exclusively on (forms of) OSCC [11–13], others analyzed HNSCC cohorts that comprised multiple subsites as a single entity [14–16]. A markedly different tumor biology was found by comparative genomics studies of HNSCC that focused on the differences between human papillomavirus (HPV)-positive and -negative oropharyngeal tumors [17,18]. Direct comparisons of the mutational profiles of OSCC and L/P-SCC have not yet been performed. Such an analysis could offer explanations for the difference in outcome of OSCC and L/P-SCC following CRT.

We therefore set out to investigate somatic point mutations (SPMs) and copy number aberrations (CNAs) in HPV-negative OSCC and L/P-SCC. We excluded HPV-positive tumors because these have a different genomic make-up, tumor biology and etiology. Specifically, we compared the total and gene-specific rates of SPMs and CNAs between OSCC and L/P-SCC. To this end we employed two datasets. The first dataset consists of targeted DNA sequencing data from 55 OSCC and 56 L/P-SCC tumor samples from our institute (NKI). The second dataset consists of 276 OSCC and 134 L/P-SCC samples, also HPV-negative, from The Cancer Genome Atlas (TCGA).

Patients and methods

Patients

We retrospectively analyzed fresh frozen pretreatment tumor samples from patients treated at our institute between 2001 and 2010. All patients gave informed consent to have biopsies stored in our tissue bank and used for scientific research. Only biopsies with at least 50% tumor cells as determined on H&E sections were selected for DNA extraction. Samples that were negative for HPV DNA, as determined by p16 staining, targeted DNA sequencing and PCR were included. Together 111 tumor samples ('NKI dataset'), of which 55 were OSCC ($OSCC_{NKI}$) and 56 L/P-SCC ($L/P\text{-}SCC_{NKI}$) were selected. Matched normal samples were unavailable for the majority of tumors and genomic analyses were therefore performed on tumor samples only. From the TCGA we collected data for all available HPV-negative OSCC (n = 276, $OSCC_{TCGA}$) and HPV-negative L/P-SCC (n = 134, $L/P\text{-}SCC_{TCGA}$) samples. NKI and TCGA patient and tumor characteristics are described in Table 1. Whereas the $L/P\text{-}SCC_{NKI}$ dataset consisted of hypo- and oropharyngeal cancers, the $L/P\text{-}SCC_{TCGA}$ dataset consisted mainly of laryngeal tumors (Table 1).

Sequencing and bioinformatics protocol of NKI dataset

Details of the sequencing and bioinformatics protocols applied in the NKI dataset are specified in the Supplementary Methods [19–27]. In short, we performed target capture DNA sequencing of 556 human genes (Supplementary Table 1). HPV gene baits, to capture HPV DNA in the samples, were included in order to determine the HPV status. We removed DNA sequence variants that were in any of three public SNP databases [25–27] and classified the remaining variants as SPMs (listed in Supplementary Table 2). Homozygous deletions and focal amplifications were detected using the R package PureCN [22].

The Cancer Genome Atlas data

We collected open access clinical, SPM and CNA data for the TCGA samples from the most recent available Firehose run (28-12-2016). SPMs of TCGA were detected by comparing whole exome sequencing data of tumors with their matched normal samples. For analyses of individual genes non-silent SPMs were selected. We included silent mutations for analyses on the total number of SPMs and the determination of transitions and transversions (TiTvs) rates. TiTvs were generated with the GenVisR package [28]. Through assessing the relative contribution of single-nucleotide polymorphism in a sample, a copy number profile can be generated using SNP array data [29,30]. CNAs were detected based on whole genome SNP6 arrays. These were available for the TCGA dataset [29,30]. From the gene level data, we selected CNAs that exceeded the chromosome arm aberrations in each sample. In TCGA data, these values are typically regarded as homozygous deletions and focal amplifications. We considered these CNAs to correspond best to those of the NKI dataset, because they both represent high amplitude CNAs. Furthermore, we performed an analysis to identify regions that were significantly amplified or deleted across all TCGA HNSCC samples. These were identified with the GISTIC2 algorithm [31] and are part of the open access data. Genomic scar signature scores were available for 141/276 OSCC and 76/134 L/P-SCC samples from the supplementary data of [32]. The codes to reproduce all analyses on TCGA data are available at https://github.com/dvossen/OSCC-versus-LPSCC.

Frequently mutated genes in HNSCC

Mutational profiles of OSCC and L/P-SCC tumors are based on a gene set of genes that are frequently mutated in HNSCC, as identified in [33]. These consisted of 168 genes with frequent SPMs ('$genes_{SPM}$', Supplementary Table 3) and 25 genes with frequent CNAs ('$genes_{CNA}$', Supplementary Table 4). To warrant a sufficient high statistical power, we limited our analyses to these genes. In addition, events will have to be frequent to explain a differential CRT response. In [33], '$genes_{SPM}$' were identified by algorithms that select genes with more SPMs than expected by chance given various background mutation rates and processes. The '$genes_{CNA}$' came from regions frequently affected by focal CNAs in HNSCC [33] and contain 25 genes that are annotated in the Cancer Gene Census [34] (Supplementary Table 4). The NKI targeted sequencing efforts captured 27 out of these 168 '$genes_{SPM}$' and 11 of the 25 '$genes_{CNA}$' (Supplementary Tables 3 and 4). NKI dataset analyses are based on this subset of genes and TCGA data on all '$genes_{SPM}$' and '$genes_{CNA}$'.

Statistical methods

Correlation coefficients refer to Spearman's rank correlation coefficient. We used Fisher's exact test to compare proportions between OSCC and L/P-SCC and the Wilcoxon rank-sum test to compare numerical variables. For the tests on the $genes_{SPM}$ and $genes_{CNA}$ we controlled the false discovery rate at 0.10 by correcting for multiple hypothesis testing with the Benjamini and Hochberg method. The corrected P-values are reported as Q-values. Error bars on proportions report the 95% confidence interval (Wilson score interval). We used a binomial mixed model to compare the proportion of each TiTv in OSCC and L/P-SCC, with subsite as a fixed and sample as a random effect. We used the log-rank test and Cox proportional hazards model to test for associations between clinical or genetic features and overall survival. All statistical analyses were performed in the R environment for statistical computing.