Comparative genomic analysis of oral versus laryngeal and pharyngeal cancer


DOI
10.1016/j.oraloncology.2018.04.006

Publication date
2018

Document Version
Final published version

Published in
Oral Oncology

License
Article 25fa Dutch Copyright Act (https://www.openaccess.nl/en/in-the-netherlands/you-share-we-take-care)

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
Comparative genomic analysis of oral versus laryngeal and pharyngeal cancer

David M. Vossena,b, Caroline V.M. Verhagena,b, Marcel Verheijac,d, Lodewyk F.A. Wesseld,e, Conchita Vensa,c, Michiel W.M. van den Brekelb,f,g,*

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...
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 dichotomy in this tumor site. Yet, some studies on operable HNSCC point towards a different CRT response of OSCC and L/P-SCC [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 (NIKI). 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 (NIKI dataset), of which 55 were OSCC (OSCCNIKI) and 56 L/P-SCC (L/P-SCCNKI) 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, OSCCTCGA) and HPV-negative L/P-SCC (n = 134, L/P-SCCTCGA) samples. NIKI and TCGA patient and tumor characteristics are described in Table 1. Whereas the L/P-SCCNKI dataset consisted of hypopharyngeal cancers, the L/P-SCCTCGA dataset consisted mainly of laryngeal cancers (Table 1).

Sequencing and bioinformatics protocol of NIKI dataset

Details of the sequencing and bioinformatics protocols applied in the NIKI 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 (TiTv) rates. TiTv 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 NIKI 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 GSTIC2 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 (‘genesSPM’, Supplementary Table 3) and 25 genes with frequent CNAs (‘genesCNA’, 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], ‘genesSPM’ were identified by algorithms that select genes with more SPMs than expected by chance given various background mutation rates and processes. The ‘genesCNA’ 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 NIKI targeted sequencing efforts captured 27 out of these 168 ‘genesSPM’ and 11 of the 25 ‘genesCNA’ (Supplementary Tables 3 and 4). NIKI dataset analyses are based on this subset of genes and TCGA data on all ‘genesSPM’ and ‘genesCNA’.

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 genesSPM and genesCNA 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 subite 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.
Results

Mutational profiles: major similarities and minor differences between subsites

We first determined the mutational profiles of the NKI tumor samples in terms of somatic point mutations (SPMs) and copy number aberrations (CNAs), and compared them to the TCGA. We restricted our analysis to gene sets with genes that were previously found to have frequent SPMs (\textit{genesSPM}) and CNAs (\textit{genesCNA}) in the TCGA HNSCC data (see Methods). Fig. 1A gives an overview of the frequently mutated genes in all samples. We then calculated the percentage of samples that carry a SPM in each individual gene\textit{SPM} and a CNA of each gene\textit{CNA} within OSCCNKI, OSCCTCGA, P-SCCNKI and L/P-SCCTCGA. These revealed many similarities across datasets and subsites, with frequent SPMs in \textit{TP53}, \textit{CDKN2A} and \textit{PIK3CA} and CNAs in \textit{CCND1}, \textit{CDKN2A} and \textit{EGFR} (Fig. 1B and C). When clustering on affected genes\textit{SPM}, or genes\textit{CNA} OSCC from different datasets clustered with each other (Supplementary Fig. 1). Overall similarities are further supported by the high correlations between the datasets and across subsites as illustrated in Supplementary Fig. 2 (\(\rho\) range 0.71–0.85).

In summary, the mutational profiles are remarkably similar. However, there are differences both between datasets and subsites. Therefore, to identify differences between subsites while excluding dataset bias, we compared the mutational profiles of subsites within datasets.

OSCCs and L/P-SCCs have distinct patterns of somatic point mutation affected genes

To expose the differences between OSCC and L/P-SCC, we next compared the percentage of samples with SPMs in the gene\textit{SPM} set. After correcting for multiple hypothesis testing, significance was reached for only one gene in the NKI and for four in the TCGA dataset (Fig. 2A). \textit{CASP8} mutations occurred almost exclusively in OSCC in both the NKI (15% vs 0%; \(Q < 0.1\)) and TCGA datasets (17% vs 2%; \(Q < 0.001\)). \textit{HRAS} mutations followed this trend, which reached significance in the TCGA dataset (10% vs 1%; \(Q < 0.01\)). In contrast, \textit{LAMA2} (5% vs 19%; \(Q < 0.01\)) and \textit{NSD1} mutations (7% vs 25%; \(Q < 0.001\)) occurred less frequently in OSCC than L/P-SCC. \textit{LAMA2} and \textit{NSD1} were not sequenced in the NKI dataset.

Next, we compared the total number of SPMs (including silent mutations, see Methods) per sample between subsites. This was performed in the TCGA dataset only, since the total number of mutations was too small in the NKI dataset due to the limited targeted sequencing (556 genes). We found that L/P-SCC had significantly more mutations per sample than OSCC (median 191 vs 146 per sample; \(P < 0.001\)) (Fig. 2B). We then investigated the distribution of the different possible transitions and transversions (TiTvs) in the SPMs. Each individual TiTv was more frequent in L/P-SCC samples (Supplementary Fig. 1B and C). When clustering on affected genes\textit{SPM}, or genes\textit{CNA} OSCC from different datasets clustered with each other (Supplementary Fig. 1). Overall similarities are further supported by the high correlations between the datasets and across subsites as illustrated in Supplementary Fig. 2 (\(\rho\) range 0.71–0.85).

In summary, the mutational profiles are remarkably similar. However, there are differences both between datasets and subsites. Therefore, to identify differences between subsites while excluding dataset bias, we compared the mutational profiles of subsites within datasets.

Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>OSCCNKI (n = 55)</th>
<th>P-SCCNKI (n = 56)</th>
<th>OSCCTCGA (n = 276)</th>
<th>L/P-SCCTCGA (n = 134)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>63</td>
<td>(12)</td>
<td>60</td>
<td>(9)</td>
</tr>
<tr>
<td>Median (range)</td>
<td>64 (43–93)</td>
<td>58 (37–78)</td>
<td>62 (19–90)</td>
<td>61 (38–85)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>40</td>
<td>73</td>
<td>39</td>
<td>70</td>
</tr>
<tr>
<td>Female</td>
<td>15</td>
<td>27</td>
<td>17</td>
<td>30</td>
</tr>
<tr>
<td>Primary site</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypopharynx</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>50</td>
</tr>
<tr>
<td>Larynx</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oral cavity</td>
<td>55</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oropharynx</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>50</td>
</tr>
<tr>
<td>T stage*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T1</td>
<td>9</td>
<td>16</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>T2</td>
<td>27</td>
<td>49</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>T3</td>
<td>7</td>
<td>13</td>
<td>24</td>
<td>43</td>
</tr>
<tr>
<td>T4</td>
<td>12</td>
<td>22</td>
<td>23</td>
<td>41</td>
</tr>
<tr>
<td>Tx</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N stage*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>28</td>
<td>51</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>N1</td>
<td>9</td>
<td>16</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>N2</td>
<td>17</td>
<td>31</td>
<td>35</td>
<td>63</td>
</tr>
<tr>
<td>N3</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Nx</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Smoker</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current or former</td>
<td>35</td>
<td>64</td>
<td>50</td>
<td>89</td>
</tr>
<tr>
<td>Never</td>
<td>12</td>
<td>22</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Unknown</td>
<td>8</td>
<td>15</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>30</td>
<td>55</td>
<td>51</td>
<td>91</td>
</tr>
<tr>
<td>No</td>
<td>18</td>
<td>33</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Unknown</td>
<td>7</td>
<td>13</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

* Pathologic stages are reported for OSCCNKI, OSCCTCGA and L/P-SCCTCGA, clinical stages for P-SCCNKI.
numbers and altered mutation spectrum, we assessed which of 30 COSMIC mutational signatures (http://cancer.sanger.ac.uk/cosmic/signatures, 18-12-2017) was present in each tumor sample in order to infer the cause of these differences (Supplementary Methods). We found five signatures with a significant difference in the prevalence of these signatures (Fig. 2C). Signatures 1 and 15, associated with age and DNA mismatch repair respectively, are more prevalent in OSCC. L/P-SCC, in contrast, are characterized by a prevalence of signatures 3, 4 and 24, that are associated with homologous recombination (HR) deficiency, smoking and aflatoxin exposure respectively. The prevalence of the age and smoking associated signatures is consistent with the relative contribution of smokers and younger patients within the different patient populations (Table 1).

In summary, OSCC is characterized by frequent CASP8 and HRAS mutations and relatively infrequent LAMA2 and NSD1 mutations. The mutational burden is higher in L/P-SCC and they show more frequently signs of HR DNA repair defects.

Common copy number aberrations occur less frequently in OSCCs

To characterize larger genomic aberrations, we next compared the percentage of OSCC and L/P-SCC samples with CNAs in the genesCNA.
The occurrence of CNAs was significantly lower in OSCC compared to L/P-SCC for \textit{CCND1} and \textit{TP63}. \textit{CCND1} was less frequently amplified in OSCC than L/P-SCC in both the NKI (24\% vs 70\%; Q < 0.001) and TCGA datasets (23\% vs 37\%; Q < 0.1) (Fig. 2D). \textit{TP63} (14\% vs 35\%; Q < 0.001) was also less frequently amplified in OSCC than L/P-SCC in the TCGA dataset. The trend was the same in the NKI dataset, but did not reach significance. Interestingly, both genes were less frequently amplified in OSCC than L/P-SCC. This led us to question whether OSCCs have in general fewer CNAs than L/P-SCCs.

First, for each sample, we considered the total number of CNAs amongst the genes\textsubscript{CNA} set. We compared OSCC with L/P-SCC samples and found that OSCC samples have fewer CNAs in both the NKI (P < 0.001) and TCGA (P < 0.001) datasets (Supplementary Table 5). Others have shown that \textit{CASP8}-mutated HNSCC harbor fewer CNAs than \textit{CASP8}-wildtype HNSCC [11]. Given the enrichment for \textit{CASP8} mutations in OSCC, we also compared the frequency of CNAs while excluding \textit{CASP8}-mutated samples. We still found fewer CNAs in OSCC in both, the NKI (P < 0.01) and TCGA (P < 0.05) datasets (Supplementary Table 5). Finally, per sample, we considered the total number of CNAs amongst regions that are significantly often amplified or

Fig. 2. Genetic mutations that occur with significant different frequencies between HNSCC subsites. (A) Genes\textsubscript{SPM} with a significant different frequency of somatic point mutations between oral squamous cell carcinomas (OSCC) and laryngeal and pharyngeal squamous cell carcinomas (L/P-SCC). Bars show the frequency of somatic point mutations in these HNSCC subsites, in the TCGA (top) and NKI (bottom) dataset. (B) Frequency of transitions and transversions (TiTvs) (y-axis) in each TCGA sample (x-axis). One sample without somatic mutations and four samples with more than 1000 somatic mutations are not shown. Inlay boxplot compares the distribution of total number of TiTvs per sample between subsites (outliers not shown, but included in statistical testing). (C) COSMIC mutational signatures that were present in significantly higher proportion of OSCC or L/P-SCC. (D) Genes\textsubscript{CNA} with significant different frequency of copy number aberrations between OSCC and L/P-SCC.
deleted across all TCGA HNSCC samples. This analysis was performed on whole genome copy number data that was possible in the TCGA data set. OSCC harbored fewer regions of amplification and deletion than L/P-SCC, both including \( (P < 0.001) \) and excluding \( (P < 0.01) \) \( \text{CASP8} \)-mutated samples (Fig. 3A and B). As \( \text{CASP8} \) and \( \text{HRAS} \) mutations frequently co-occur in OSCC \([11]\) (in OSCC\textsubscript{TCGA} 40\% of \( \text{CASP8} \)-mutated versus 4\% of \( \text{CASP8} \)-wildtype tumors harbored \( \text{HRAS} \) mutations, \( P < 0.001 \)), we further stratified by \( \text{HRAS} \) mutation status as well. Also \( \text{HRAS} \)-wildtype OSCC showed fewer regions of amplification and deletion than their \( \text{HRAS} \)-wildtype L/P-SCC counterparts (Fig. 3C).

Taken together, these copy number data show that OSCC harbor fewer CNAs and in particular less frequent amplification of \( \text{CCND1} \) and \( \text{TP63} \) than L/P-SCC.

**OSCCs have fewer genomic scars**

Chromosomal instability processes may underlie the observed increase in CNAs in L/P-SCC. HR deficiency can cause chromosomal instability that results in gross CNAs \([35]\), often referred to as ‘genomic scars’. Three SNP array-based genomic scar signatures were applied to the TCGA HNSCC dataset in a recent study \([32]\). These signatures quantify the level of a particular pattern of genomic scarring and each signature is associated with HR deficiency (i.e. \( \text{BRCA1} \) or \( \text{BRCA2} \) mutations) or markers thereof (cisplatin sensitivity). The signatures are:

- Number of telomeric Allelic Imbalances (NtAI) \([36]\),
- Large Scale Transition (LST) \([37]\) and
- Homologous Recombination Deficiency (HRD) score \([38]\).

Here we compared signature scores between OSCC and L/P-SCC, in order to probe for possible HR deficiency differences. We find that OSCCs had significantly lower genomic scar scores than L/P-SCCs. This was true for all tested signatures (Fig. 4A). Stratifying subsites according to \( \text{CASP8} \) mutation status revealed a consistent pattern across all three signatures: \( \text{CASP8} \)-mutated OSCC had the lowest score, then \( \text{CASP8} \)-wild type OSCC and finally \( \text{CASP8} \)-wild type L/P-SCC (Fig. 4B). There were no \( \text{CASP8} \)-mutated L/P-SCC samples with available signatures scores in this dataset. The pattern was similar when stratifying by \( \text{HRAS} \) mutation status (Fig. 4C).

To further quantify potential HR deficiency across subsites, we split samples at the median HRD score following the example of Abkevich \ et al. \([38]\). The proportion of samples with a high HRD score was substantially higher in the L/P-SCCs than OSCCs, regardless whether \( \text{CASP8} \)-mutated tumors were included (76\% vs 45\%; \( P < 0.001 \)) or excluded (76\% vs 50\%; \( P < 0.001 \)). These results are in line with our observation that L/P-SCCs have a higher prevalence of an HR deficiency associated mutational signature (Fig. 2C). Samples with presence of the HR deficiency associated mutational signature had higher genomic scar signature scores, thereby further sustaining this finding (Supplementary Fig. 4).

In a previous study (manuscript submitted), we found functional HR and Fanconi anemia (FA) pathway defects in a panel of HNSCC cell lines that were accompanied by FA/HR gene SPMs. As SPMs in HR/FA genes (Supplementary Table 6) may have caused the observed HR deficiencies, we tested for an association between such mutations and the signature scores in all samples with available scar signature scores. HR/FA-mutated tumor samples had significantly higher NtAI and LST scores (Fig. 5A). HRD scores were, however, similar. This pattern was also still present when analyzing OSCC and L/P-SCC separately (Fig. 5B and C), although not always significant, possibly due to the smaller sample sizes. Finally, the proportion of HR/FA-mutated OSCCs was lower than that of L/P-SCC in the TCGA (\( P < 0.05 \)), but this could not be confirmed in the smaller NKI dataset (Supplementary Fig. 5).

Overall, our results indicate that OSCC harbor less HR defects as determined by genomic scar signatures. This holds true when excluding the CNA-silent subgroup of \( \text{CASP8} \)-mutated tumors, enriched in OSCCs. Thus, a lower incidence of HR deficiency in OSCC than L/P-SCC might underlie the lower incidence of CNAs in OSCC.

**Associations of mutational features with clinical outcomes**

OSCC, compared to L/P-SCC, showed largely similar but also distinctively different mutational features. Contemporary OSCC treatment regimens prioritize surgery over CRT, partly due to the absence of a pronounced CRT response in OSCC in the past. Thus, this resulted in a lack of CRT-treated OSCC patients available for genomic CRT response association studies. In an attempt to investigate the potential clinical relevance of the here observed biology characteristic to L/P-SCC or OSCC, i.e. genetic aberrations, we explored alternative data bases.
Fig. 4. OSCC has fewer genomic scars. (A) Boxplot of genomic scar signature scores in 141 oral squamous cell carcinoma (OSCC) and 76 laryngeal and pharyngeal squamous cell carcinoma (L/P-SCC) samples from the TCGA dataset. The signatures are: number of telomeric Allelic Imbalances (NtAI), Large Scale Transition (LST), Homologous Recombination Deficiency (HRD) score. (B) Same as (A), but subsites are stratified according to CASP8 mutation status. (C) Same as (A), but subsites are stratified according to HRAS mutation status.

Fig. 5. Somatic point mutations in Homologous recombination and Fanconi anemia genes are associated with genomic scar signature scores. (A) Genomic scar signature scores in 141 oral squamous cell carcinoma (OSCC) and 76 laryngeal and pharyngeal squamous cell carcinoma (L/P-SCC) samples from the TCGA dataset. Boxplots compare score signatures of samples with a somatic mutation in any of the HR/FA genes (Mut) and those wildtype for all HR/FA genes (Wt). (B) Same as (A), but only showing OSCC samples. (C) Same as (A), but only showing L/P-SCC samples.
hazard ratios and P with a better overall survival (Fig. 6D). Here we observe that mutational burden, NtAI and LST are associated with OS as a continuous variable (Fig. 6E, F). Subsequently, these variables were split for illustration purposes, and the corresponding log-rank test P-values are shown in-figure. Similarly, in the TCGA ovarian cancer dataset (D) mutational burden, (E) NtAI score and (F) LST score were each associated with OS as a continuous variable (P = 0.005, P = 0.029 and P < 0.001 in univariable Cox models, respectively). Each was subsequently median split for illustration purposes. Supplementary Tables 7 and 8 contain the hazard ratios and P-values of all uni- and multivariable Cox models for the TCGA OSCC and ovarian cancer datasets, respectively.

Primarily set up for genomic studies, the TCGA data contains only surgically treated cases and clinical data are incomplete. To test adequacy, we first analyzed whether the two HNSCC datasets display the known impact of strong clinical factors (T-stage, N-stage and age) on overall survival (Supplementary Fig. 6). As the OSCC clinical data behaved as expected, we next tested whether the individual mutational features are associated with clinical outcome (Supplementary Table 7). We find that mutational burden, NtAI score and CASP8 mutations are connected to poor prognosis in this patient group (Fig. 6A-C). To further deduce a role in platinum-based treatment responses, we investigated the relevance of the genomic alterations in the TCGA ovarian cancer dataset (Supplementary Fig. 6 and Supplementary Table 8). These patients received adjuvant platinum-based chemotherapy [39]. Here we observe that mutational burden, NtAI and LST are associated with a better overall survival (Fig. 6D–F and Supplementary Table 8), suggesting that patients with tumors with such characteristics may benefit from platinum-based treatments.

Together, these observations could point to a favorable platinum response of tumors with the genetic characteristics of L/P-SCC, in particular repair-defect associated signatures and scars. They are less frequent in OSCC, however relevant to outcome if present. This might explain the lack of a pronounced CRT response in this tumor site.

**Discussion**

Prompted by clinical indications that OSCC may respond differently to CRT than L/P-SCC, we compared the mutational profiles of OSCC and L/P-SCC. We identified six genes that are more frequently mutated in either OSCC or L/P-SCC. We find that OSCC has fewer SPMs and CNAs than L/P-SCC. Important in the context of CRT response, the subsites specifically in mutation and genomic scar signatures associated with HR deficiency.

Our analyses show that OSCC scores lower than L/P-SCC on three HR deficiency associated scar signatures. These signatures scores were positively associated with SPMs in HR/FA genes. These results thus strongly suggest that L/P-SCC harbor HR defects, and that those are less common in OSCC. It should be noted that these signatures were developed for breast and ovarian cancer. Two of these signatures are associated with mutations in the HR genes BRCA1 and BRCA2, in basal-like breast cancer [37] and ovarian tumors [38]. The third signature is associated with cisplatin sensitivity in breast cancer cell lines [36]. In our analysis in HNSCC, we found, however, overall a good interrelation between genomic scar signatures, COSMIC HR signatures and the presence of SPMs in FA/HR genes, together supporting biological significance of the scoring. Importantly, cells and tumors with HR defects
are particularly sensitive to platinum-based chemotherapeutics such as cisplatin [35]. Clinical association studies based on TCGA data should be considered with caution. In addition, TCGA patients all underwent surgery, hampering a comparison to definite CRT. Yet, our observations do illustrate a potential link between the genetic characteristics related to HR defects and the reported superior CRT response in L/P-SCC. Indicating relevance in HNSCC, they depicted poor prognosis in the OSCC patient dataset. A prolonged survival was evident in patients with ovarian cancers that showed high mutational burden or signs of HR defects. Treated with platinum-based adjuvant chemotherapy [39], this could point to a favorable treatment setting for such HR-affected tumors. With a decreased incidence in OSCC, this would be therefore consistent with an overall decrease in the CRT response. Taken together, our findings could therefore provide the basis for a differential CRT response and explain why the reported outcomes of L/P-SCC following platinum-based CRT may be better than those of OSCC [2,7,8,10].

The HNSCC genes\textsuperscript{Spam} and genes\textsuperscript{CNA} sets where originally identified in the TCGA dataset [33]. The TCGA HNSCC dataset however mostly consists of OSCC tumor samples (roughly two thirds), thereby possibly causing some bias in the gene set selection. Genes which were frequently mutated in OSCC were therefore well captured. However, we acknowledge that due to the smaller proportion of L/P-SCC samples in the TCGA dataset, genes that are frequently mutated in L/P-SCC but not OSCC might have remained undetected. Yet, while also including our NKI data set, we were able to assess well whether or not these genes\textsuperscript{Spam} and genes\textsuperscript{CNA} were also frequently mutated in L/P-SCC.

The results of our analysis show that OSCCs are enriched for CASP8 and HRAS mutations and are consistent with other studies that identified a subset of OSCCs characterized by frequent CASP8 mutations co-occurring with HRAS mutations and infrequent CNAs [11,14]. One of these studies found that CASP8-mutated tumors were more tumorigenic and larger in an orthotopic model of OSCC [11]. Another study demonstrated that CASP8 mutations promote migration and invasion in HNSCC cell lines and tumor growth in mouse xenografts [40]. These experimental results seem in contradiction with clinical data, which indicate that HNSCC with infrequent CNAs have a relatively good outcome [14,41]. CASP8 has a role in the extrinsic apoptotic pathway that may also influence treatment resistance [42,43]. Cisplatin- and radiotherapy-induced apoptosis signals have been correlated to tumor response in some cancer types [44,45]. Whether inactivating CASP8 mutations in OSCC [40] could contribute to a poor CRT response, will have to be evaluated in future studies. The finding that CASP8-mutated OSCCs harbor fewer CNAs than other OSCC is supported both by our and other’s data [11,14]. Our results further demonstrate that OSCC is a subtype with fewer CNAs even when CASP8 wild-type. Another finding of this study was enrichment of LAMA2 and NSD1 mutations in L/P-SCCs. NSD1 encodes a histone methyltransferase. An epigenome deregulation study conducted on TCGA data confirms the enrichment of NSD1 mutations in laryngeal carcinomas [46]. The enrichment of LAMA2 mutations in L/P-SCC has not yet been reported to our knowledge. LAMA2 (encoding a subunit of laminin) germline loss has been reported in neurofibromatosis type 1 associated tumors [47] and hypermethylation in colorectal carcinoma [48]. As positive selection of LAMA2 mutations suggests functional involvement of this gene, it raises the question why this is in particular the case in L/P-SCC. Future research may answer the subtype-specific role of these mutated genes.

### Conclusion

There are genetic differences between OSCC and L/P-SCC with respect to several genes that are frequently mutated. COSMIC signatures, genomic scars and frequent CNAs reveal possible differences in HR deficiency. This reflects a different biology that could provoke an altered CRT response. Our findings could stimulate future investigations into more individualized treatment protocols for different subtypes of head and neck cancer.

### Conflict of interest

The authors declare no conflicts of interest.

### Acknowledgements

This study was sponsored by grants from the Verwielens fund, Brunel International, the EU 7th framework program (ARTFORCE) and Alpe d’HuZes-KWF (DESIGN). No grant number is applicable. The authors thank members of the Genomics Core Facility for their assistance and Jos Elbers, Paul Essers and Martijn van der Heijden for carefully reading the manuscript.

### Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.oraloncology.2018.04.006.

### References


