Integrative Analysis of Head and Neck Cancer Identifies Two Biologically Distinct HPV and Three Non-HPV Subtypes

Michaela K. Keck1,2, Zhixiang Zuo1, Arun Khattri1, Thomas P. Stricker3, Christopher D. Brown4, Matin Imanguli5, Damian Rieke1, Katharina Endhardt1, Petra Fang1, Johannes Brägelmann1, Rebecca DeBoer1, Mohamed El-Dinai1, Serdal Aktolga1, Zhendeng Lei6, Patrick Tan6,7, Steve G. Rozen6, Ravi Salgia1,8, Ralph R. Weichselbaum1,8, Mark W. Lingen3,8, Michael D. Story9, K. Kian Ang10, Ezra E.W. Cohen11, Kevin P. White8,12, Everett E. Vokes1,8, and Tanguy Y. Seiwert1,8,12

Abstract

Purpose: Current classification of head and neck squamous cell carcinomas (HNSCC) based on anatomic site and stage fails to capture biologic heterogeneity or adequately inform treatment. Experimental Design: Here, we use gene expression-based consensus clustering, copy number profiling, and human papillomavirus (HPV) status on a clinically homogenous cohort of 134 locoregionally advanced HNSCCs with 44% HPV+ tumors together with additional cohorts, which in total comprise 938 tumors, to identify HNSCC subtypes and discover several subtype-specific, translationally relevant characteristics.

Results: We identified five subtypes of HNSCC, including two biologically distinct HPV subtypes. One HPV+ and one HPV− subtype show a prominent immune and mesenchymal phenotype. Prominent tumor infiltration with CD8+ lymphocytes characterizes this inflamed/mesenchymal subtype, independent of HPV status. Compared with other subtypes, the two HPV subtypes show low expression and no copy number events for EGFR/HER ligands. In contrast, the basal subtype is uniquely characterized by a prominent EGFR/HER signaling phenotype, negative HPV-status, as well as strong hypoxic differentiation not seen in other subtypes.

Conclusion: Our five-subtype classification provides a comprehensive overview of HPV+ as well as HPV− HNSCC biology with significant translational implications for biomarker development and personalized care for patients with HNSCC. Clin Cancer Res; 21(4); 870–81. ©2014 AACR.

Introduction

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common nonskin cancer worldwide with an annual incidence of approximately 600,000 cases and a mortality rate of 40% to 50% (1, 2). The major known risk factors are environmental exposures to tobacco products, alcohol, and infection with high-risk human papillomaviruses (HPV). The incidence of HPV+ tumors is rising rapidly and HPV status is now the strongest prognostic marker (3).

Although HNSCC is now widely viewed as composed of two distinct clinical entities, HPV+ and HPV−, a therapeutically relevant molecular classification system remains elusive. Prior attempts at profiling this clinically heterogeneous disease have been hampered by lack of information about HPV status and/or absence of significant numbers of HPV− cases, small sample size, and lack of annotation (4, 5). Currently, neither molecular classification nor validated biomarkers are used in clinical practice. Instead all patients with HNSCC are treated independent of the underlying biology based on stage and anatomic location, typically using a combination of surgery, radiation, and chemotherapy (6). Cetuximab, an anti-EGFR antibody, is the only approved targeted therapy for HNSCC, but predictive biomarkers remain to be identified and overall cetuximab has a response rate of 13% (7).

In the current study, we investigate a fully clinically annotated, very homogeneous patient cohort of locoregionally advanced HNSCC (including 44% HPV+) tumors, all treated uniformly with organ-preserving chemoradiotherapy together with all available other HNC cohorts, including the cancer genome atlas head
and neck cancer cohort (TCGA), to identify HNSCC subtypes and their biologic and prognostic characteristics. We integrate gene expression-based consensus clustering, cross-platform approaches for validation, copy number profiling, and HPV status to discover subtype-specific, translationally relevant biology.

The resulting new taxonomy of HNSCC closely correlates with multiple HNC-relevant oncogenic processes and a biology-based taxonomy of HNSCC. In particular, we identify two subtypes of HPV+ head and neck cancer and two subtypes (one HPV− and one HPV+) with a prominent immune phenotype characterized by CD8+ T-cell infiltration, as well as a subtype with near exclusive HER ligand/EGFR copy number-driven signaling. As such, our classification is broadly applicable to biomarker and therapy development for head and neck cancer and will inform future clinical and preclinical studies, biomarker development, and personalized HNC care.

Materials and Methods

Patient samples

OCT blocks of frozen tissue samples were obtained from the University of Chicago Head and Neck Cancer tissue bank (Chicago, IL; Institutional Review Board-approved protocol UCCCC#8980; Supplementary Experimental Procedures and Supplementary Fig. S1).

Nucleic acid extraction

A total of 171 locoregionally advanced HNSCC specimens obtained before treatment with concurrent chemoradiotherapy (all patients received organ-preserving chemoradiotherapy with curative intent) were selected for this study. Our aim was to select at least 100 cases to be able to perform consensus clustering later. Samples were collected and banked between 1997 and 2010, but our focus was on recent cases. A section was cut from OCT-frozen blocks and stained with hematoxylin and eosin (H&E). An expert HNC pathologist reviewed slides to determine and circle the area with the highest tumor content. Samples with microscopic tumor or tumor content <60% were excluded from further analysis. Guided by the H&E-stained slides, the region with the highest tumor content was cut from the OCT blocks, pulverized using a CryoPrep (Covaris) and homogenized in lysis buffer from an AllPrepRNA/DNA/Protein Mini kit (Qiagen) or from an RNA/DNA/Protein Purification Kit (Norgen Biotek) using an Ultrasonicator (Covaris). DNA, RNA, and protein were isolated from each sample using the respective kit and following the manufacturer’s protocol.

Gene expression and copy number analysis

Gene expression profiling was done using Agilent 4 × 44 Kv2 expression arrays on 134 samples (Supplementary Fig. S1 and Supplementary Experimental Procedures). To evaluate copy number aberration (CNA) in HNSCC, we selected a panel of 86 and subsequently 147 candidate genes known to have frequent CNA in cancer and analyzed these using the Nanostring nCounter assay (NanoString Technologies; ref. 8; Supplementary Experimental Procedures). Data and materials availability: Gene Expression Omnibus Accession number: GSE40774 (Embargoed until publication; http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=jsjnbqwcyoiiygrac=GSE40774).

Discovery of subtypes

To discover gene expression-based subtypes, we used newly generated gene expression data (Agilent, n = 134), as well as a data set based on Illumina BeadArrays (n = 131; Yordy and colleagues; in preparation), and publicly available HNSCC microarray data from two studies based on Affymetrix HG133 arrays (n = 106; refs. 9, 10; see Supplementary Experimental Procedures for details).

IHC and immunofluorescence

IHC and immunofluorescence were performed using routine methods (Supplementary Experimental Procedures).

Results

Identification and validation of three HNSCC supergroups

Microarray gene expression profiles of 371 HNSCC samples were derived from several datasets (Supplementary Table S1), including our cohort (n = 134, Agilent; Supplementary Fig. S1), an Illumina cohort (Yordy and colleagues; unpublished results), and two Affymetrix cohorts (9, 11). These cross-platform datasets were used as the discovery set for the classification of HNSCC samples. All datasets were analyzed as described in the Supplementary Experimental Procedures and outlined in Supplementary Fig. S2A. This resulted in 821 reliable and informative genes (Supplementary Table S2), based on which five subtypes were derived separately in each dataset (Supplementary Fig. S2B). These five subtypes slightly differ across the three datasets, but fall into three very stable cross-platform groups (termed supergroups) on the basis of hierarchical clustering of their centroids (Fig. 1A, B, and C). All centroids in the three cross-platform supergroups show a strongly positive silhouette width (average = 0.47), indicating that centroids fit their respective cluster (Fig. 1B; ref. 12). The high correlation of our three supergroups and previously

Translational Relevance

Head and neck cancer (HNC) is composed of human papillomavirus HPV+ and HPV− tumors. However, there remains significant heterogeneity in clinical behavior (e.g., response to therapies, including anti-EGFR or anti–PD-1 immunotherapy) and further biologic subclassification and elucidation of specific biologic characteristics are lacking. One limitation of available classifications is the absence of any larger (and representative) cohort of HPV− tumors or the lack of validation across multiple datasets/platforms. We identify five HNC subgroups, including two distinct HPV subgroups with differential biology across multiple HNC cohorts, and show validity of subtypes and associated biology across all currently available HNC datasets, including The Cancer Genome Atlas (TCGA) HNC cohort: In particular, we report (i) an immune (and mesenchymal) phenotype present in a group of HNC tumors independent of HPV status and (ii) a group with non–HPV-associated tumors showing a prominent HER-driven phenotype as well as hypoxia, which are candidate biomarkers for respective therapies (e.g., PD-1, EGFR/HER, or hypoxia-targeting agents).
identified subtypes of lung SCC (13) and HNSCC (14) confirms the validity of our classification (Fig. 1C). We applied the nearest centroid approach, a standard classification method, to assign new samples from the validation datasets (14–16) to the supergroup with closest matching centroid. Supergroups were validated in two independent validation datasets from different platforms, including Affymetrix and RNA-Seq, suggesting broad applicability (Fig. 1E). According to prior nomenclature in other cancer types and molecular characteristics (12, 17, 18), we named the three supergroups inflamed/mesenchymal (IMS), basal (BA), and classical (CL), respectively.

Two distinct HPV HNSCC subtypes and three non-HPV HNSCC subtypes

HNSCC tumors are composed of two distinct disease entities: HPV− and HPV+ tumors. We investigated the relationship between HPV status and the three supergroups (basal (BA), classical (CL), and inflamed/mesenchymal (IMS)). First, we determined the HPV status on our cohort using HPV DNA-based PCR and RNA-based qPCR (see Supplementary Experimental Procedure); for the TCGA dataset, we used exome-sequencing data and RNA-Seq data to determine the DNA and RNA of HPV viruses (see Supplementary Experimental Procedure). The HPV status of other public datasets was determined using our unpublished HPV gene

Figure 1.
Identification and validation of HNSCC expression subtypes. A, hierarchical clustering of subtype centroids from the last iteration of consensus clustering on the three discovery datasets (Supplementary Fig. S2B). The heatmap shows the Pearson correlation of the centroids. Centroids of the subtypes from the three datasets cluster together into three supergroups: Inflamed/mesenchymal (IMS), basal (BA), and classical (CL). B, Silhouette plot of the centroids (n = 15) for three supergroups shows a positive width for all centroids. C, principal component analysis plot shows the correlation between three supergroups and previously published subtypes for lung SCC (17) and HNSCC (14). D, heatmap shows the 821 predictive gene signature in the three discovery datasets (Agilent: n = 130, Illumina: n = 128, Affymetrix: n = 104). Samples were ordered according to group predictions, and genes were clustered using the three discovery datasets. E, heatmap shows the 821 predictive gene signature in two validation datasets (Affymetrix: n = 136, RNA-Seq: n = 414): Samples in the validation dataset were assigned to groups using a nearest centroid algorithm. Gene order from the training set was maintained.
expression signature. In all the datasets, HPV+ tumors were not gathered into one group, but fall into two distinct groups: (i) the inflamed/mesenchymal (IMS) supergroup and (ii) the classical (CL) supergroup. No HPV+ tumors fell in the basal supergroup (Fig. 2A). It is evident that HPV+ HNSCCs are composed of two distinct gene expression subtypes, namely, IMS-HPV and CL-HPV. Interestingly, there is biologic overlap of these HPV+ samples in a supergroup and HPV- tumors in the same supergroup (e.g., immune or mesenchymal differentiation). The non-HPV subtypes were named as basal, CL-nonHPV, and IMS-nonHPV.

**Figure 2.**
Stratification of HNC tumors by supergroups and HPV status. A, the HNC tumors in both the discovery datasets (Agilent, Illumina, Affymetrix; Discovery) and validation datasets (RNA-Seq, Affymetrix; Validation) are stratified into five subtypes according to supergroups and HPV status. HPV tumors are subdivided into classical (CL) and inflamed/mesenchymal (IMS), and completely absent in basal (BA). B, selection of HNC-relevant genes/pathways according to ingenuity pathway annotation shows that supergroups/subtypes are characterized by distinct and differential biologic processes. Smoking history and differential expression of genes related to smoking are also displayed. C, significantly associated pathways with subtypes. The pathway activity of hypoxic (up and down), epithelial, mesenchymal, and proliferation for each patient was obtained by single sample gene set enrichment analysis (ssGSEA). Heatmaps are used to show the difference of pathway activities between subtypes in Agilent and TCGA cohorts.

www.aacrjournals.org Clin Cancer Res; 21(4) February 15, 2015 873

Published OnlineFirst December 9, 2014; DOI: 10.1158/1078-0432.CCR-14-2481
Molecular and genetic characteristics of HNSCC subtypes

**Basal subtype—BA.** One distinctive feature of the basal subtype is the significant enrichment for hypoxia signaling, represented by hypoxia-responsive genes such as *HIF1A*, *CA9*, and *VEGF* (Fig. 2B and Supplementary Fig. S3A). In addition, the gene expression profile of the basal subtype closely matches a published hypoxia gene signature (Fig. 2C) which validates the hypoxic phenotype (19). The second key characteristic of the basal subtype is strong enrichment for neuregulin signaling, including *EGFR*, *AREG* (amphiregulin), and *NRG1* (neuregulin 1/hergulin; Fig. 2B and Supplementary Fig. S3A). Overexpression of epithelial markers such as P-cadherin (*CDH3*) and cytokeratins (*KRT1*, *KRT9*) is another distinctive characteristic of the basal subtype (Fig. 2B and Supplementary Fig. S3A) similar to basal breast cancer (20). Moreover, a published EMT signature also demonstrates the elevated epithelial pathway activities in the basal subtype (Fig. 2C; ref. 21). Consistent with the high expression of cytokeratins in the basal subtype, the morphology data also demonstrated that basal tumors are highly keratinizing and well differentiated (Supplementary Fig. S4).

**Classical subtypes—CL-HPV and CL-nonHPV.** The most distinctive feature of the classical supergroup is the significant enrichment for putrescine (polyamine) degradation pathway (Supplementary Fig. S3B), which is relevant for detoxification, for example, related to tobacco use. Increased polyamine levels are associated with increased cell proliferation (22). Consistently, a published proliferation signature also indicates that the classical supergroup has a higher proliferation rate compared with the other groups (Supplementary Fig. S4; ref. 23).

Although the CL-HPV and CL-nonHPV subtypes share similarities, they are still two distinct disease entities, reflected in many biologic pathways. Cell-cycle genes, such as mini-chromosome maintenance proteins (*MCM2* and *MCM10*), cell division cycle protein kinase (*CDC7*), and -related genes (*CDKN2A*, *E2F2*, and *RPA2*) are overexpressed in the CL-HPV subtype. The two subtypes also show significant difference in tobacco use with 42% heavy smokers in CL-nonHPV compared with 74% heavy smokers in CL-HPV (Fig. 2B and Table 1, *P* = 0.01, Fisher exact test). Consistent with smoking status, xenobiotic metabolism pathway genes *AKR1C1*, *AKR1C3*, and *ALDH3A1*, which are known to be associated with smoking, are enriched in CL-nonHPV (Fig. 2B; refs. 24, 25).

**Inflamed/mesenchymal subtypes—IMS-HPV and IMS-nonHPV.** The distinguishing features of the IMS group is expression of immune response genes like *CD8*, *ICOS*, *LAG3*, and *HLA-DRA* (Fig. 2B) related to the infiltration of CD8⁺ T lymphocytes in tumors.

### Table 1. Patient and sample overview

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Basal</th>
<th>CL-HPV</th>
<th>CL-nonHPV</th>
<th>IMS-HPV</th>
<th>IMS-nonHPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients, n</td>
<td>130</td>
<td>43</td>
<td>19</td>
<td>19</td>
<td>36</td>
<td>13</td>
</tr>
<tr>
<td>Age, y (median)</td>
<td>57</td>
<td>56</td>
<td>56</td>
<td>60</td>
<td>59</td>
<td>57</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>23</td>
<td>10</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Male</td>
<td>107</td>
<td>33</td>
<td>17</td>
<td>12</td>
<td>33</td>
<td>12</td>
</tr>
<tr>
<td>% Tumor Median</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>60–77.5</td>
<td>60–72.5</td>
<td>60–70</td>
<td>60–80</td>
<td>60–80</td>
<td>60–70</td>
</tr>
<tr>
<td>HPV status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>55</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>Negative</td>
<td>75</td>
<td>43</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Anatomic site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larynx</td>
<td>29</td>
<td>10</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Oral cavity</td>
<td>25</td>
<td>18</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Oropharynx</td>
<td>75</td>
<td>14</td>
<td>18</td>
<td>5</td>
<td>35</td>
<td>3</td>
</tr>
<tr>
<td>Other*</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tobacco use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>22</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Light</td>
<td>36</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Heavy</td>
<td>72</td>
<td>27</td>
<td>8</td>
<td>14</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Alcohol use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>16</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Light</td>
<td>35</td>
<td>16</td>
<td>4</td>
<td>3</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Heavy</td>
<td>79</td>
<td>23</td>
<td>12</td>
<td>16</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>T stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0–T2</td>
<td>43</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>T3–T4</td>
<td>87</td>
<td>34</td>
<td>13</td>
<td>16</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>N stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>12</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N1</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>N2</td>
<td>90</td>
<td>22</td>
<td>12</td>
<td>15</td>
<td>31</td>
<td>10</td>
</tr>
<tr>
<td>N3</td>
<td>19</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Tumor stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I–II</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>III</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>IV</td>
<td>124</td>
<td>40</td>
<td>18</td>
<td>19</td>
<td>34</td>
<td>13</td>
</tr>
</tbody>
</table>

**NOTE:** 4 out of 134 samples (3%) were undefined and did not meet criteria to allow classification into only one subtype.

*Paranasal sinus.*
Mesenchymal genes such as vimentin (VIM), matrix metalloproteinases (MMP9), and S100A4 also show increased expression in the IMS group (Fig. 2B), which typically associates with increased metastatic risk (8, 26). Epithelial markers such as P-cadherin (CDH3) and cytokeratins (KRT1, KRT9) are downregulated, suggesting epithelial-to-mesenchymal transition (EMT; Fig. 2B; ref. 21). Consistently, a published EMT signature (21) is upregulated in the IMS group and there is downregulation of epithelial differentiation markers (Fig. 2C).

The two subtypes in the IMS group show a significant difference in cell-cycle pathways and smoking-associated pathways (Fig. 2B). Similar to the CL-HPV subtype, the IMS-HPV subtype has significantly higher cell-cycle pathway activities, in which HPV is known to play a critical role. The IMS-HPV subtype shows a higher proliferation rate indicated by a published proliferation signature (Fig. 2C; ref. 23) and IMS-HPV subtype tumors are nonkeratinizing and poorly differentiated according to the morphology review by light microscopy (Supplementary Fig. S3).

Patient characteristics and survival

Our cohort of 134 HNSCC samples with 44% HPV+ is representative for the patient population presenting to a large referral center for advanced HNC. There is an increasing proportion of HPV+ cases (Table 1). All patients in this new cohort had locoregionally advanced disease and were treated with concurrent chemoradiotherapy. Consistent with previous reports, patients with oropharyngeal tumors are mainly in the two HPV subtypes:
IMS-HPV and CL-HPV (71%, $P = 4.19 \times 10^{-11}$). In contrast, oral cavity tumors are overrepresented in the basal group (72%, $P = 1.04 \times 10^{-10}$). Seventy-four percent of the CL-nonHPV tumors and 85% of IMS-nonHPV are heavy smokers, which is significantly higher than other subtypes.

Five-year survival rate was assessed for our Agilent, TCGA, and Illumina cohorts, respectively. In all the cohorts, the HPV subtypes have a significantly higher 5-year survival than non-HPV subtypes. Furthermore, between the two HPV subtypes, the IMS-HPV subtype shows a trend toward higher 5-year survival than CL-HPV (Fig. 3A) consistent across all cohorts. We used Kaplan–Meier estimator to measure the overall survival (OS for the subtypes in the above three cohorts, separately. The log-rank test was used to evaluate the Kaplan–Meier curve difference among the subtypes. We found that OS differs significantly among the five subtypes in our Agilent cohort ($P = 0.037$) and the TCGA cohort ($P = 0.003$) but not in the Illumina cohort ($P = 0.589$; Fig. 3B). Furthermore, the Kaplan–Meier analysis on the combined dataset of the three cohorts showed a clearer difference in OS of the five subtypes ($P = 8e-06$), which was more significant than any cohort independently (Fig. 3B). Importantly, the IMS-HPV subtype is different from the CL-HPV subtype by better survival (Fig 3A and B) suggesting a potential positive impact of immune response.

Copy number analysis exhibits distinct alterations per subtype

Copy number analysis was executed in all samples. Initially, 101 samples were analyzed targeting 75 genes previously described to have CNA (27–29). A GISTIC-derived strategy was used to identify significant CNA. Significant copy number gains were detected in 12 genes, including $PIK3CA$ (3q26.3), $VEGFA$ (6p21), $EGFR$ (7p11), $MYC$ (8q24), and $CCND1$ (11q13; Fig. 4A and B and Supplementary Table S3). Significant copy number losses were detected in 16 genes, including $CDKN2A$ (9p21) and $RB1$ (13q14; Fig. 4B and Supplementary Tables S3). Subsequently, we aimed to further investigate these aberrant regions by scanning 72 additional genes within copy number altered regions in 55 of the samples. We identified that $TP63$, $SOX2$ (3q26-28), and $PIK3CA$ (3q26.3; Supplementary Table S4) are coamplified in most tumors.

Furthermore, CNAs for a lot of oncogenes and tumor suppressors are found to have strong associations with subtypes (Supplementary Table S5). Amplification of $EGFR$ (7p12), $CCND1$ (11q13), and $FADD$ (11q13), as well as deletion of $FHIT$ (3p14) and $CDKN2A$ (9p2) with few exceptions are only observed in the non-HPV subtypes, particularly in the basal subtype (Supplementary Table S5). Amplification of $MYC$ (8q24) and $ITGB4$ (17q25) is significantly enriched in the basal subtype (Supplementary Table S5). $ITGB4$ is also overexpressed in the basal subtype, and it is reported that $ITGB4$ mediates cell adhesion and plays a crucial role in the initiation, progression, and metastasis of solid tumors (16, 30).

The copy number gains of genes ($PRKCI$, $PIK3CA$, and $DCLUN1D1$) in the 3q26-28 region can be seen in all subtypes, but are significantly higher in the classical subtypes (Supplementary Table S5). Although the frequency of $E2F3$ (6p22) amplification in our cohort is only 5%, most of the amplification events occur in the classical subtypes, including both CL-HPV and CL-nonHPV subtypes (Supplementary Table S5). Consistently, $E2F3$ is overexpressed in classical subtypes. $E2F3$ encodes a transcription factor important for cell-cycle regulation and DNA replication, and its amplification and overexpression are known to associate with invasive tumor growth and rapid tumor cell proliferation in urinary bladder cancer (31). Importantly, the two HPV subtypes differ significantly in these regions, suggesting different biology of two HPV subtypes. The amplification of 3q26-28 was validated by different markers such as $TP63$ and $SOX2$ (Fig. 4C).

Using IHC, we validated the ability of key markers ($SOX2$, $Cyclin D1/CCND1$, p16/CDKN2A, and $TP63$) to differentiate subtypes (Supplementary Fig. S5).

Cytotoxic T-cell infiltration and subtypes

One of the key findings based on gene expression is the discovery of immune-related marker expression ($CD8A/B$) in IMS tumors (Fig. 5A). We performed multicolor immunofluorescence to identify tumor-infiltrating CD8+ lymphocytes in these tumors (Fig. 5B). The enrichment of cytotoxic T-cell infiltration in mesenchymal tumors is present even when considering oropharynx tumors only, suggesting independence of anatomic location and lack of contamination from normal lymphoid tissue (Fig. 5C).

Discussion

Although previous attempts at expression profiling of HNSCC identified distinct tumor subtypes (11, 14, 19), a therapeutically relevant molecular classification remains elusive due to lack of information about HPV status and/or absence of significant numbers of HPV+ cases, small sample size, and lack of predictive biomarkers. The HNSCC subtypes proposed in our study are identified in an unsupervised, cross-cohort and cross-platform way, which is supported by four discovery cohorts and four independent validation cohorts, which together total 938 patients. We showed that our subtypes are remarkably similar to those found in LIISC (17) and also have a strong correlation with the previous identified HNSCC subtypes (14). Moreover, our expression profiling includes a large and representative number of HPV+ tumors, allowing us to identify two biologically distinct HPV− subtypes.

Consequently, we now propose that HNSCC can be classified into five distinct subtypes—two HPV and three non-HPV subtypes. The close correlation of the five HNSCC subtypes with morphologic characteristics, molecular processes, survival, and copy number changes supports a biologic and clinical basis for

Figure 4.

Copy number (CN) analysis in the five subtypes. A, genome-wide plot of copy number G-scores across the samples is shown for each subtype. The subtypes are represented by different colors. B, copy number for 75 genes in 101 samples (discovery set). Heatmap shows samples in columns, genes in rows ordered by chromosomal region. The copy number is normalized to $-2$ to $2$, where $-2$ means high-level deletion, $-1$ means low-level deletion, $1$ means low-level amplification, and $2$ means high-level amplification. C, box plots show validation of chr3q26 amplification using different marks in this region. $PIK3CA$ is used as markers in a discovery dataset with 101 samples. $TP63$ and $SOX2$ are used as the markers in a validation dataset with 55 samples.

www.aacnjournals.org

Clin Cancer Res; 21(4) February 15, 2015 877

Published OnlineFirst December 9, 2014; DOI: 10.1158/1078-0432.CCR-14-2481

Downloaded from clincancerres.aacnjournals.org on January 5, 2018. © 2015 American Association for Cancer Research.
this classification, making it unlikely that subtypes are attributable to chance, artifact, or bias.

The most important finding of this report is the identification of two distinct HPV subtypes. Previous clinical observations suggest that HPV+ tumors are diverse, for example, the subgroup of patients with HPV+ tumors do not respond well to therapy (3, 32). Our classification into two distinct HPV subtypes provides a biologic basis for clinical heterogeneity and suggests that differential treatment approaches might be required for HPV+ tumors. CL-HPV and IMS-HPV subtypes exhibit significant differences in many aspects such as morphology, molecular processes, CNAs, and clinical features. More than 40% of the CL-HPV tumors show keratinization, whereas none of the IMS-HPV tumors are keratinizing. Accordingly, IMS-HPV subtype tumors are more poorly differentiated compared with CL-HPV subtype tumors. The IMS-HPV subtype exhibits significantly elevated CD8 T cell infiltration differs between HNC subtypes. A, boxplots showing the difference of the mRNA level expression of CD8 in HNC subtypes for Agilent, TCGA, and Illumina cohorts. Mesenchymal subtypes show significantly higher level of CD8 mRNA expression, independent of HPV status. B, immunofluorescence showing diffuse CD8 T-cell infiltration in an IMS tumor (right panel; CD8+ cells are shown in green), absence thereof in a basal (BA) tumor (left panel). C, the difference of CD8 mRNA expression in HNC subtypes considering oropharynx tumor only, for Agilent and TCGA cohorts. The CD8 mRNA expression are compared within HPV subtypes and within non-HPV subtype.

Figure 5.
expression of mesenchymal markers, whereas the CL-HPV subtype shows higher proliferation. Compared with the IMS-HPV subtype, the CL-HPV subtype presents significantly more canonical genomic aberrations associated with squamous cell carcinoma, such as amplification of 3q26-27. For example, EZF3 amplification, which is associated with cell proliferation, is only found in the CL-HPV subtype but not in the IMS-HPV subtype, which correlates with the proliferation signature seen by expression in the CL-HPV subtype. The differences in the biologic patterns found between the two HPV subtypes are supported by a trend for favorable prognosis of IMS-HPV subtype tumors, and consistent with findings in other cancer types. Although further investigations are needed to explain the difference in immune response between the two HPV subtypes, one can hypothesize that differences in the host, tumor, and viral genetic background, as well as their interplay, are etiologic.

Given the widespread use of IHC for diagnostic purposes in routine pathology practice (e.g., p16, p63), we evaluated the ability of protein expression to differentiate subtypes (Supplementary Fig. S5). Although additional optimization and validation are necessary, development of either an IHC or a qPCR-based assay could potentially also allow classification of HNSCC similar to what is used for breast cancer (47).

In conclusion, we propose a new taxonomy of HNSCC based on five subtypes with a comprehensive overview of HPV+ as well as HPV− HNSCC biology validated across 938 HNSCC patient tumors. There are significant translation implications with respect to immunotherapy, presence of two biologically distinct HPV subtypes, and HER-targeted therapies that will inform biomarker development and personalized care efforts that are already being pursued.

Disclosure of Potential Conflicts of Interest

T.Y. Seiwert is a consultant/advisory board member for Merck/MSD and Bayer/Onyx. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions


Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M.K. Keck, T.P. Stricker, M. Imanaguli, P. Fang, R.R. Weichselbaum, M.W. Lingen, M.D. Story, K.K. Ang, E.E.W. Cohen, K.P. White, E.E. Vokes, T.Y. Seiwert


Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M.K. Keck, D. Rieke, K. Endhardt, R. Deboer, M. El Dinali, S. Aktopla, T.Y. Seiwert

Study supervision: E.E. Vokes, T.Y. Seiwert

Other (funding): T.Y. Seiwert

Acknowledgments

The authors thank Drs. Janet Rowley and Yusuke Nakamura for their critical feedback, suggestions, and guidance, the Gleason Family, and the Achatz Research Fund for their support of HNC research (to T.Y. Seiwert).

Grant Support

This work was supported by an ASCO Translational Professorship (to E.E. Vokes). Dr. T.Y. Seiwert was supported by The Flight Attendant Medical Research Institute (FAMRI) Young Investigator Award.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 14, 2014; accepted November 7, 2014; published OnlineFirst December 9, 2014.
References


HPV-positive (+) and HPV-negative (−) recurrent/metastatic squamous cell carcinoma of the head and neck (R/M SCCHN): Analysis of the global phase III SPECTRUM trial. J Clin Oncol 30, 2012 (suppl; abstr 5504).  