Integrative analysis of Head and Neck Cancer identifies two biologically distinct HPV and three non-HPV subtypes

Michaela K. Keck\textsuperscript{1,2}, Zhixiang Zuo\textsuperscript{1}, Arun Khattri\textsuperscript{1}, Thomas P. Stricker\textsuperscript{3}, Christopher Brown\textsuperscript{7}, Matin Imanguli\textsuperscript{4}, Damian Rieke\textsuperscript{1}, Katharina Endhardt\textsuperscript{1}, Petra Fang\textsuperscript{1}, Johannes Brägelmann\textsuperscript{1}, Rebecca DeBoer\textsuperscript{1}, Mohamed El-Dinali\textsuperscript{1}, Serdal Aktolga\textsuperscript{1}, Zhengdeng Lei\textsuperscript{5}, Patrick Tan\textsuperscript{5,6}, Steve G. Rozen\textsuperscript{5}, Ravi Salgia\textsuperscript{1,11}, Ralph R. Weichselbaum\textsuperscript{1,11}, Mark W. Lingen\textsuperscript{3,11}, Michael D. Story\textsuperscript{8}, Kie Kian Ang\textsuperscript{12}, Ezra E.W. Cohen\textsuperscript{9}, Kevin P. White\textsuperscript{10,11}, Everett E. Vokes\textsuperscript{1,11}, Tanguy Y. Seiwert\textsuperscript{1,10,11}\textsuperscript{*}

\textsuperscript{1}The University of Chicago, Department of Medicine, Chicago, IL, USA.
\textsuperscript{2}The University of Hohenheim, Institute of Biological Chemistry and Nutrition, Stuttgart, Germany.
\textsuperscript{3}The University of Chicago, Department of Pathology, Chicago, IL, USA.
\textsuperscript{4}The University of Texas Southwestern Medical Center, Department of Otolaryngology-Head and Neck Surgery, Dallas, TX, USA.
\textsuperscript{5}Duke-NUS Graduate Medical School, Singapore.
\textsuperscript{6}Genome Institute of Singapore, Singapore.
\textsuperscript{7}The University of Chicago, Department of Human Genetics, Chicago, IL, USA.
\textsuperscript{8}The University of Texas, Southwestern Medical Center, Department of Radiation Oncology, Dallas, TX, USA.
\textsuperscript{9}University of California San Diego, La Jolla, CA
\textsuperscript{10}The University of Chicago, Institute for Genomics and Systems Biology, Chicago, IL, USA.
\textsuperscript{11}The University of Chicago Comprehensive Cancer Center, Chicago, IL, USA.
\textsuperscript{12}The University of Texas, MD Anderson Cancer Center, Houston, TX, USA.
\textsuperscript{†}Authors contributed equally

\textbf{Keywords:} Head and Neck Cancer; Integrated Genomic Analysis; Human papillomavirus; Immune Phenotype, HER signaling, oligoclonal T-cell infiltration

\textbf{Financial Support:}
This work was supported by an ASCO Translational Professorship (EEV). We would like to thank the Achatz Research Fund for its support of HNC research. Dr. Seiwert was supported by The Flight Attendant Medical Research Institute (FAMRI) Young Investigator Award.

\textbf{*Corresponding Author Address:}
Tanguy Seiwert
Knapp Center for Biomedical Discovery at the University of Chicago
900 E. 57\textsuperscript{th} Street, 7110
Chicago, IL 60637
United States
Tel: 773-702-2452
Fax: 773-702-3002
Email: tseiwert@medicine.bsd.uchicago.edu
The authors disclose no potential conflict of interest.

**Word count (excluding references): 3731**

**Number of figures and tables: 6**

**Statement of Translational Relevance:**

Head and Neck Cancer (HNC) is comprised of 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 sub-classification and elucidation of specific biologic characteristics is 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 subtypes including two distinct HPV-subtypes with differential biology across multiple HNC cohorts, and show validity of subtypes and associated biology across all currently available HNC datasets including the TCGA HNC cohort: In particular we report 1) an immune (and mesenchymal) phenotype present in a group of HNC tumors independent of HPV-status, 2) 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).
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 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 subtypes - independent of HPV status. Compared to other subtypes, the two HPV subtypes show low expression and no copy number events for EGFR/HER-ligands and absence of EGFR/HER-ligand expression. By contrast the Basal subtype is uniquely characterized by a prominent EGFR/HER signaling phenotype driven by multiple HER ligands, 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 HNSCC patients.
Introduction

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common non-skin cancer worldwide with an annual incidence of ~600,000 cases and a mortality rate of 40-50% (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).

While HNSCC is now widely viewed as comprised 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 HNSCC patients 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 7-13% (7).

In the current study, we investigate a fully clinically annotated, very homogenous 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 (TCGA) head and neck cohort, in order 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 in order 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(+), 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 (IRB approved protocol UCCCC#8980)(Supplemental Experimental Procedures and Supplementary Figure S1).

Nucleic acid extraction

A total of 171 locoregionally advanced HNSCC specimens obtained prior to 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 CryoPrep (Covaris, Woburn, MA) and homogenized in lysis buffer from an AllPrepRNA/DNA/Protein Mini kit (Qiagen, Valencia, CA) or from an RNA/DNA/Protein Purification Kit (Norgen Biotek, Thorold, Canada) using an Ultrasonicator (Covaris, Woburn, MA). DNA, RNA and protein were isolated from each sample using the respective kit and following manufacturer’s protocol.

Gene expression and copy number analysis
Gene expression profiling was done using Agilent 4x44Kv2 expression arrays on 134 samples (Supplementary Figure S1, Supplemental Experimental Procedures). To evaluate 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 (NanoString Technologies, Seattle, WA) (8)(Supplemental Experimental Procedures). Data and materials availability: GEO Accession number: GSE40774 (Embargoed until publication) (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=jjsnbqwecyoirg&acc=GSE40774).

**Discovery of subtypes**

To discover gene expression-based subtypes we used newly generated gene expression data (Agilent, n=134), as well as a dataset based on Illumina BeadArrays (n=131)(Yordy et al, under review), and publically available HNSCC microarray data from two studies based on Affymetrix HG133 arrays (n=106) (9,10) (see Supplemental Experimental Procedures for details).

**Immunohistochemistry (IHC) and Immunofluorescence (IF)**

IHC and IF were performed using routine methods (Supplemental 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 Figure S1), an Illumina cohort (Yordy et al, 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 Supplemental Experimental Procedures and outlined in Supplementary Figure S2A. This resulted in 821 reliable and informative genes (Supplementary Table S2), based on which five subtypes were derived separately in each dataset (Supplementary Figure S2B). These five subtypes slightly differ across the three datasets, but fall into three very stable cross-platform groups (termed ‘super-groups’) based on the hierarchical clustering of their centroids (Figure 1A/B). All centroids in the three cross-platform super-groups show a strongly positive silhouette width (average = 0.47), indicating that centroids fit their respective cluster (Figure 1B) (12). The high correlation of our three super-groups and previously identified subtypes of lung SCC (13) and HNSCC (14) confirms the validity of our classification (Figure 1C). We applied the nearest centroid approach, a standard classification method, to assign new samples from the validation data sets (TCGA, Slebos et al(5), Rickman et al(15), Ye et al(16)) to the super-group with closest matching centroid. Super-groups were validated in two independent validation datasets from different platforms including Affymetrix and RNA-Seq suggesting broad applicability (Figure 1D). According to prior nomenclature in other cancer types and molecular characteristics (12,17,18), we named the three super-groups Inflamed/mesenchymal (IMS), Basal (BA) and Classical (CL), respectively.
Two distinct HPV HNSCC subtypes and three non-HPV HNSCC subtypes

HNSCC tumors are comprised of two distinct disease entities: HPV negative and HPV positive tumors. We investigated the relationship between HPV status and the three super-groups (BA, CL and MS). Firstly, 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 expression signature. In all the datasets HPV positive tumors were not gathered into one group, but were fall into two distinct groups, 1) the Inflamed/Mesenchymal supergroup (IMA) and 2. the Classical super-group (CL). No HPV(+) tumors fell in the Basal super-group (Figure 2A). It is evident that HPV positive HNSCCs are comprised of two distinct gene expression subtypes, namely, inflamed/mesenchymal-HPV (IMS-HPV) and Classical-HPV (CL-HPV). Interestingly there is biologic overlap of these HPV(+) samples in a supergroup and HPV(-) negative tumors in the same supergoup (e.g. immune or mesenchymal differentiation). The non-HPV subtypes were named as Basal (BA), Classical-nonHPV (CL-nonHPV) and inflamed/mesenchymal-nonHPV (IMS-nonHPV).

Molecular and genetic characteristics of HNSCC subtypes

Basal Subtype – BA

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

\textit{Classical Subtypes – CL-HPV and CL-nonHPV}

The most distinctive feature of the CL super-group is the significant enrichment for putrescine (polyamine) degradation pathway (\textit{Supplementary Figure S3B}), which is relevant for de-toxification e.g related to tobacco use. Increased polyamine levels are associated with increased cell proliferation (22). Consistently, a published proliferation signature also indicates that the CL super-group has a higher proliferation rate compared to the other groups (\textit{Supplementary Figure S4}) (23).

Although the CL-HPV and CL-nonHPV subtypes share similarities, they are still two distinct disease entities, reflected in many biological pathways. Cell cycle genes, such as mini-chromosome maintenance proteins (\textit{MCM2, MCM10}), cell division cycle protein kinase (\textit{CDC7}) and HPV related genes (\textit{CDKN2A, E2F2 and RPA2}) are overexpressed in the CL-H subtype. The two subtypes also show significant difference in tobacco use with 74% heavy smokers in CL-N compared to 42% heavy smokers in CL-H (\textit{Figure 2B, Table 1, P=0.01, Fisher’s 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-N (*Figure 2B*) (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, HLA-DRA* (*Figure 2B*) related to the infiltration of CD8+ T-lymphocytes in tumors.

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

The two subtypes in the IMS group show a significant difference in cell cycle pathways and smoking associated pathways (*Figure 2B*). Similar to the CL-H subtype, the IMS-H subtype has significantly higher cell cycle pathway activities, in which HPV is known to play a critical role. IMS-H subtype shows a higher proliferation rate indicated by a published proliferation signature (*Figure 2C*) (23) and IMS-H subtype tumors are non-keratinizing and poorly differentiated according to the morphology review by light microscopy (*Supplementary Figure 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%, $P=4.19\times10^{-11}$). By contrast oral cavity tumors are overrepresented in the BA group (72%, $P=1.04\times10^{-5}$). 74% of the CL-nonHPV tumors and the 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 five-year survival than non-HPV subtypes. Furthermore between the two HPV subtypes, the IMS-HPV subtype shows a trend towards higher five-year survival than CL-HPV (Figure 3A) consistent across all cohorts. We used Kaplan-Meier estimator to measure the overall survival for the subtypes in the above three cohorts, separately. The log-rank test was used to evaluate the KM curve difference among the subtypes. We found that overall survival (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 ($p=0.589$) (Figure 3B). Furthermore, the KM 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 (Figure 3B). Importantly the IMS-H subtype is different from the CL-H subtype by better survival (Figure 3A/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 copy number aberrations (CNA) (27-29). A GISTIC-derived strategy was used to identify significant copy number aberrations (CNA). Significant copy number gains were detected in 12 genes including PIK3CA (3q26.3), VEGFA (6p21), EGFR (7p11), MYC (8q24) and CCND1 (11q13) (Figure 4A/B, Supplementary Table S3). Significant copy number losses were detected in 16 genes including CDKN2A (9p21) and RB1 (13q14) (Figure 4/B, 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 co-amplified in most tumors.

Furthermore, copy number aberrations 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 BA subtype (Supplementary Table S5). Amplification of MYC (8q24) and ITGB4 (17q25) is significantly enriched in the basal (BA) subtype (Supplementary Table S5). ITGB4 is also overexpressed in the BA subtype, and it is reported that ITGB4 mediates cell adhesion and plays a crucial role on the initiation, progression and metastasis of solid tumors (16,30).

The copy number gains of genes (PRKCI, PIK3CA and DCUN1D1) in the 3q26-28 region can be seen in all subtypes, but are significantly higher in the classical (CL) 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 is 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* (*Figure 4C*).

Using immunohistochemistry we validated the ability of key markers (*SOX2, Cyclin D1/CCND1, p16/CDKN2A, and TP63*) to differentiate subtypes (*Supplementary Figure 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 inflamed/mesenchymal tumors (*Figure 5A*). We performed multicolor immunofluorescence to identify tumor infiltrating CD8+ lymphocytes in these tumors (*Figure 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 (*Figure 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 LUSC (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 morphological characteristics, molecular processes, survival and copy number changes supports a biologic and clinical basis for 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 – e.g. 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 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, copy number aberrations and clinical features. More than 40% of the CL-
HPV tumors show keratinization, while none of the IMS-HPV tumors are keratinizing. Accordingly, IMS-HPV subtype tumors are more poorly differentiated compared to CL-HPV subtype tumors. The IMS-HPV subtype exhibits significantly elevated expression of mesenchymal markers, while the CL-HPV subtype shows higher proliferation. Compared to 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. E.g. $E2F3$ 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 biological patterns found between the two HPV subtypes are supported by a trend towards better survival in the IMS-HPV subtype compared to the CL-HPV subtype, and this observation holds true across all cohorts, although it does not correlate as strongly with prognosis as e.g. tobacco use.

Another key finding of this report is the recognition of the differential expression of immune markers in the subtypes. Cancer immune surveillance is considered to be a factor in the body’s ability to prevent cancer (33). Evading immune surveillance has been considered an emerging hallmark of cancer (34). Importantly activity of checkpoint blockade was recently reported in HNC using anti-PD-1 therapies, and PD-L1 expression correlates with an inflamed phenotype that is consistent with and validates our IMS subtypes (35,36).

Furthermore accumulating evidence indicates that cancer immune suppression can be driven by hypoxia (37,38). We detected that the BA subtype is highly enriched for HIF1A signaling. Immunosuppressive factors such as VEGF (39) are upregulated in the Basal (BA) subtype, and a lack of immune related markers is evident. Alternative to hypoxia induced activation, HIF1A can be activated under normoxic conditions by EGFR signaling (40,41),
which is also found to be significantly activated in the BA subtype tumors. EGFR amplification and MYC amplification are found in the BA subtypes. Taken together, the immunosuppression of the BA subtype could be driven by either hypoxia and potentially also EGFR signaling, and further validation will be important as immune targeted therapies are being developed for HNC.

Currently no clinically relevant EGFR biomarkers exist and anti-EGFR therapy is administered indiscriminately despite a low single agent response rate of 7-13% (7,42). Our results suggest that the BA subtype could serve as a candidate predictive biomarker in evaluating the benefit of an anti-EGFR therapy, as well as a hypoxia-targeting therapy.

In contrast to the absence of immune markers of the BA subtype, the IMS subtypes including both the IMS-HPV and the IMS-nonHPV subtypes exhibit a strongly activated immune phenotype, validated by CD8+ T cell infiltration on immunohistochemistry. Infiltration of tumors with CD8+ cytotoxic T-lymphocytes has been associated with a favorable prognosis in several tumor types and may be a predictive biomarker for cancer immunotherapy (43). Furthermore, novel immunomodulatory therapies (e.g. PD-1 checkpoint targeting drugs) have shown activity in squamous cell carcinomas (44) and evidence of activity in HNSCC was recently reported. We demonstrate that both IMS subtypes independent of HPV status are characterized by CD8+ T-cell infiltration (Figure 5A/B) irrespective of anatomic location (Figure 5C). Immune response against HPV(+) is of particular interest and evidence of immune escape was reported recently (45,46). This is the first report that HPV(+) tumors have two subtypes that exhibit a different immune phenotype. CD8+ T-cell infiltration is therefore one possible explanation for the trend towards a more 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 immunohistochemistry (IHC) for diagnostic purposes in routine pathology practice (e.g. p16, p63), we evaluated the ability of protein expression to differentiate subtypes (Supplemental Figure S5). While additional optimization and validation are necessary, development of either an IHC or a quantitative PCR-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 translational 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.

**Acknowledgments:**

- We would like to thank Drs. Janet Rowley and Yusuke Nakamura for their critical feedback, suggestions, and guidance.
- We would like to thank the Gleason Family for their support of HNC research (TYS).
References:


Table 1. Patient and sample overview.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>BA</th>
<th>CL-HPV</th>
<th>CL-nonHPV</th>
<th>IMS-HPV</th>
<th>IMS-nonHPV</th>
<th>UD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients No.</td>
<td>130</td>
<td>43</td>
<td>19</td>
<td>19</td>
<td>36</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>Age Median (Yrs)</td>
<td>57</td>
<td>56</td>
<td>56</td>
<td>60</td>
<td>59</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></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>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>107</td>
<td>33</td>
<td>17</td>
<td>12</td>
<td>33</td>
<td>12</td>
<td></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>
<td></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>
<td></td>
</tr>
<tr>
<td>HPV Status Positive</td>
<td>55</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>36</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>HPV Status Negative</td>
<td>75</td>
<td>43</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Anatomic Site</td>
<td></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>
<td></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>
<td></td>
</tr>
<tr>
<td>Oropharynx</td>
<td>75</td>
<td>14</td>
<td>18</td>
<td>5</td>
<td>35</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Other*</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Tobacco Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>22(17%)</td>
<td>5(11%)</td>
<td>4(21%)</td>
<td>1(5%)</td>
<td>11(31%)</td>
<td>1(7.5%)</td>
<td></td>
</tr>
<tr>
<td>Light</td>
<td>36(28%)</td>
<td>11(26%)</td>
<td>7(37%)</td>
<td>4(21%)</td>
<td>13(36%)</td>
<td>1(7.5%)</td>
<td></td>
</tr>
<tr>
<td>Heavy</td>
<td>72(55%)</td>
<td>27(63%)</td>
<td>8(42%)</td>
<td>14(74%)</td>
<td>12(33%)</td>
<td>11(85%)</td>
<td></td>
</tr>
<tr>
<td>Alcohol Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>16(12%)</td>
<td>4(9%)</td>
<td>3(16%)</td>
<td>0(0%)</td>
<td>8(22%)</td>
<td>1(7.5%)</td>
<td></td>
</tr>
<tr>
<td>Light</td>
<td>35(27%)</td>
<td>16(37%)</td>
<td>4(21%)</td>
<td>3(16%)</td>
<td>11(31%)</td>
<td>1(7.5%)</td>
<td></td>
</tr>
<tr>
<td>Heavy</td>
<td>79(61%)</td>
<td>23(54%)</td>
<td>12(63%)</td>
<td>16(84%)</td>
<td>17(47%)</td>
<td>11(85%)</td>
<td></td>
</tr>
<tr>
<td>T Stage</td>
<td></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>
<td></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>
<td></td>
</tr>
<tr>
<td>N Stage</td>
<td></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>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>N2</td>
<td>90</td>
<td>22</td>
<td>12</td>
<td>15</td>
<td>31</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>N3</td>
<td>19</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Tumor Stage</td>
<td></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>
<td></td>
</tr>
<tr>
<td>III</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>124</td>
<td>40</td>
<td>18</td>
<td>19</td>
<td>34</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>
*paranasal sinus
Figure Legends:

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 (Figure S2B). The heatmap shows the Pearson correlation of the centroids. Centroids of the subtypes from the three datasets cluster together into three super-groups: Inflamed/mesenchymal (IMS), basal (BA), classical (CL) (B) Silhouette Plot of the centroids (n=15) for three super-groups shows a positive width for all centroids (C) Principal Component Analysis plot shows the correlation between three super-groups and previously published subtypes for lung SCC 4 and HNSCC 5 (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.

Figure 2 Stratification of HNC tumors by super-groups and HPV status

(A) The HNC tumors in the both the discovery datasets (Agilent, Illumina, Affymetrix(Discovery)) and validation datasets (RNA-Seq, Affymetrix(Validation)) are stratified into five subtypes according to super-groups 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 super-groups/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.

**Figure 3** Survival analysis

(A) Five-year survival rate for the five subtypes (BA, CL-nonHPV, CL-HPV, IMS-nonHPV and IMS-HPV) in Agilent cohort (n=129), TCGA cohort (n=350) and Illumina cohort (n=99). (B) Overall survival for the five subtypes in Agilent cohort, TCGA cohort, Illumina cohort, as well as the combined dataset was estimated using Kaplan-Meier analysis. P-values from log-rank tests evaluate significance of differences in survival between subtypes.

**Figure 4** Copy number (CN) analysis in the five subtypes

(A) Genome-wide plot of CN G-scores across the samples is shown for each subtype. The subtypes are represented by different colors. (B) CN 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 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.
**Figure 5** 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 (IF) showing diffuse CD8+ T-cell infiltration in a basal tumor (BA) (left panel), absence thereof in an inflamed/mesenchymal (IMS) tumor (right 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 3

A

<table>
<thead>
<tr>
<th></th>
<th>Agilent</th>
<th>TCGA</th>
<th>Illumina</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMS-HPV</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>IMS-nonHPV</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>CL-HPV</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>CL-nonHPV</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>BA</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

5 year survival

B

<table>
<thead>
<tr>
<th></th>
<th>Agilent Overall Survival</th>
<th>TCGA Overall Survival</th>
<th>Combined Overall Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA (42)</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>CL-nonHPV (19)</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>CL-HPV (19)</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>CL-nonHPV (13)</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>CL-nonHPV (36)</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>P = 0.036</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

Survival Time in Days

<table>
<thead>
<tr>
<th></th>
<th>Illumina Overall Survival</th>
<th>Combined Overall Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA (44)</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>CL-nonHPV (18)</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>CL-HPV (9)</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>CL-nonHPV (22)</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>CL-nonHPV (9)</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>P = 0.589</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

Survival Time in Days

P = 8e-06

Downloaded from clincancerres.aacrjournals.org on April 20, 2017. © 2014 American Association for Cancer Research.
Figure 4

A: Loss Gscore vs Gain Gscore

B: Heatmap of gene expression

C: Copy number analysis

Chr3q26 (gain): PIK3CA, TP63, SOX2

Discovery Set (PIK3CA)

Validation Set (TP63)

Validation Set (SOX2)
Figure 5

A

Agilent

RNA-Seq (TCGA)

Illumina

CD8 mRNA expression

BA CL-HPV IMS-HPV CL-nonHPV IMS-nonHPV

CD3/Pan-CK/CD8/DAPI

CD8 mRNA expression

CL IMS BA/CL IMS

HPV nonHPV

Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.
# Integrative analysis of Head and Neck Cancer identifies two biologically distinct HPV and three non-HPV subtypes


*Clin Cancer Res* Published OnlineFirst December 9, 2014.

<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-14-2481</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://clincancerres.aacrjournals.org/content/suppl/2015/01/14/1078-0432.CCR-14-2481.DC2">http://clincancerres.aacrjournals.org/content/suppl/2015/01/14/1078-0432.CCR-14-2481.DC2</a></td>
</tr>
<tr>
<td>Author Manuscript</td>
<td>Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.</td>
</tr>
</tbody>
</table>

---

Sign up to receive free email-alerts related to this article or journal.

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.