

**Analysis of types of Tumor-Infiltrating Immune Cells in Oral Squamous Cell Carcinoma (OSCC) and  
their associations with survival of OSCC patients.**

Manali V. Vora

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Public Health

University of Washington

2017

**Committee:**

Chu Chen

Noel S. Weiss

Wei Sun

Program Authorized to Offer Degree:

Epidemiology

©Copyright 2017

Manali V. Vora

University of Washington

**Abstract**

**Analysis of types of Tumor-Infiltrating Immune Cells in Oral Squamous Cell Carcinoma (OSCC) and their associations with survival of OSCC patients.**

Manali V. Vora

Chair of the Supervisory Committee:

Chu Chen, PhD, NRCC, DABCC

Affiliate Professor,

Department of Epidemiology

**Background:** The 5-year survival associated with oral cancer (OSCC) is relatively low. Recent investigations have implicated the presence of tumor infiltrating lymphocytes (TILs) as independent prognostic markers in several malignancies, including oral cancer. However, the studies of oral cancer evaluated only a limited number of immune cell types as prognostic markers. CIBERSORT, a computational software for in silico analysis of tumor microenvironment, provides an opportunity to measure 22 cell types of TILs in two previously-assembled cohorts of OSCC patients.

**Methods:** We used CIBERSORT to evaluate Gene Expression Profiles (GEPs) of OSCC tumor samples using data from The Cancer Genomic Atlas (TCGA) and Fred Hutchinson Cancer Research Center's Oralchip study to infer the types of immune cells that are present in the tumor tissue. Tumor samples were analyzed to evaluate a difference in immune cell composition of OSCC tumors by HPV status, location and by tumor stage. We also assessed the independent prognostic value of individual immune cell types by means of a multivariable Cox regression model which adjusted for patient's age at diagnosis, HPV status, tumor location, tumor stage, and smoking and alcohol consumption history. In addition, we did a subset analysis on 50 HPV-positive oropharyngeal (OPC) cases where we assessed the independent prognostic value of individual immune cell types through the same multivariable Cox regression model.

**Results:** The immune microenvironment of HPV-positive OSCC cases was significantly different compared to HPV-negative OSCC cases. The proportion of TILs such as naïve B cells, cytotoxic T cells, activated memory T cells, follicular helper and regulatory T cells was higher among HPV-positive OSCC

cases and in HPV-positive OPC cases. Composition of some immune cell types differed by tumor stage. No significant association was observed between higher composition of individual immune cell types and overall survival in OSCC patients. Subset analysis showed that HPV-positive OPC cases with higher concentration of cytotoxic CD8+ T cells and activated memory CD4+ T cells were associated with better prognosis. On the other hand, higher concentrations of activated dendritic cells, mast cells and neutrophils were associated with relatively poor overall survival.

**Conclusion:** Our study suggests immune system is predictive of survival in HPV-positive OPC patients, and that the composition of immune cells in HPV-positive OPC is different from that of HPV-negative OPC. These findings may help guide clinical management of these patients.

## **INTRODUCTION:**

Oral and oropharyngeal squamous cell carcinomas (OSCC) are among the most common cancers worldwide, and have a relatively high case-fatality. Oral cavity and pharynx cancer represent 2.9% of all new cancer cases in the US [1]. In the US, the major risk factors for OSCC are tobacco use, excessive alcohol intake and Human Papillomavirus (HPV) type 16 infection. Although there have been advances in both surgical and adjuvant therapies, the overall 5-year survival for OSCC has remained low (64.5% in the US [1]) for the past few decades. 5-year survival differs by tumor stage [1], but the established TNM clinical cancer staging system, by itself, is an imperfect predictor of survival in OSCC patients. Recent investigations show that HPV-positive oropharyngeal cancer patients have a much better survival compared to HPV-negative oropharyngeal cancer patients, [53] but there is a need to search for additional prognostic biomarkers to improve clinical management of patients with this cancer. While extensive analyses of the phenotype of tumor cells by molecular approaches have been undertaken, few studies have looked at the impact of immune infiltrate or tumor related immune genes on survival.

In some malignant tumors, levels of infiltrating immune cells are associated with tumor growth, cancer progression, and patient outcome. [3,4] The presence of tumor Infiltrating leukocytes (TILs) is prognostic in several malignancies, such as ovarian, breast, and colorectal carcinoma. [5,6]

Recently, gene expression profiles (GEP) were used to compute the relative fractions of leukocytes in complex tissues, such as solid tumors. This computational method is known as Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT). Cell compositions characterized from GEPS of solid tumors using CIBERSORT have strong agreement with flow cytometry assessment of immune subsets in bulk tumors. This method is highly robust to noise or variation in gene expression data due to stochastic fluctuation, which is commonly encountered when dealing with solid tumors. The developers of CIBERSORT have designed and validated a leukocyte gene signature matrix, termed LM22 for leukocyte deconvolution from bulk tumors. It contains 547 genes that distinguish 22 human hematopoietic cell phenotypes, including seven T cell types, naïve and memory B cells, plasma cells and myeloid subsets. [9]

Another paper by the same authors analyzed the association between clinical outcomes and abundance of diverse tumor-infiltrating leukocyte (TIL) subsets elicited by CIBERSORT across 39 different malignancies, including head and neck cancer. They observed that the prognostic value of some TILs on survival in patients was independent of patient's general immune response. [7] Such large-scale analysis of collection of human tumors has allowed the identification of components of the immune contexture that predict relatively favorable and unfavorable outcomes. However, tumor infiltrates are heterogeneous between tumor types, and it is possible their impact on overall survival may differ by cancer type. [15] Thus, the presence of immune cell infiltration and its clinical implications needs to be assessed separately in OSCC. Most research that has studied the immune microenvironment of OSCC tumors addressed but one or two types of the immune cell population. [8, 11-13] Also, these studies typically had a small sample size and used traditional methods to quantify immune cells, such as examining H&E stained FFPE tumor samples, immunohistochemistry (IHC), and flow cytometry, which can be subjective, labor-intensive and somewhat qualitative in nature [14]. Only a few among these have examined the impact of TIL on survival in OSCC patients [16-20]. Cytotoxic T cells (CD8+); Helper T Cells (CD4+ etc.); Suppressor T cells (Tregs) and antigen presenting Dendritic Cells (DC) are the most commonly studied immune cell population in oral cancer. CIBERSORT provides us with an opportunity to evaluate OSCC tumor samples for 22 different type of immune cell types at once.

Our objective is to evaluate both the detailed immune microenvironment of OSCC and its association with survival. We also aim to assess whether this 'immunome' differs by given clinical stage of the tumor and HPV positivity status.

## **METHODS:**

### **Genomic and clinical data acquisition:**

The Cancer Genome Atlas (TCGA) contains publically available raw microarray and gene expression data from patient-derived tumor samples of OSCC. These samples are collected from different clinical sites, from all over the US. At present, gene expression data for 500 U.S. patients with primary head and neck squamous cell cancer (HNSC), diagnosed between the years 1992 to 2014, can be found in the Genomic Data Commons data portal (<https://portal.gdc.cancer.gov>). Whole transcriptome RNA-sequencing (RNA-seq) data were available for 546 tumor samples in total (including secondary tumors and metastatic

tumors) and we used the open access, normalized Fragments per Kilobase of transcript per Million mapped reads (FPKM) count files. Clinical data for these cases – such as patient demographics, medical and lifestyle history, vital status, time to death/time to last follow-up, and tumor characteristics (such as site, stage and HPV status) - are also openly accessible and were downloaded through TCGA. The GEPs of the 546 tumor samples were merged to produce a mixture file required for running CIBERSORT.

TCGA's sequencing data is annotated using GENCODE v22, and the given Ensembl ids were substituted with corresponding HUGO Gene Nomenclature Committee (HGNC) symbols for the mixture file.

Additionally, Fred Hutchinson Cancer Research Center's (FHCRC) OralChip dataset [51] comprising 166 OSCC tumors' gene expressions and associated clinicopathological data was used as supplementary and validation data for the two-different analysis. The 166 gene expression profiles from FHCRC were obtained using the Affymetrix HG-U133 2.0 platform instead of RNA-seq like TCGA. They were derived from OSCC patients treated at University of Washington Medical Center (Seattle, WA), the Harborview Medical Center (Seattle, WA), or the Veterans Affairs Puget Sound Health Care System during 2003 to 2007.

Normalization of the CEL files was done to match the normalization of the TCGA samples. Approval of the FHCRC's Institutional Review Board was acquired to access and analyze this dataset.

### **Identification of immune cells and their estimated fractions**

The CIBERSORT website (<http://cibersort.stanford.edu>) includes tutorials for the use of the computational tool and preparation of input data. The input file (also known as mixture file) containing gene expression data was formatted according to the instructions in the manual and uploaded on the website. For leukocyte deconvolution, the mixture file was used with their extensively benchmarked LM22 gene signature. The absolute and relative modes of CIBERSORT were selected to run together; the no. of statistical permutations was set to 1000 (>100 recommended) and quantile normalization was disabled. Given the data and the selected parameters, fractional representation of 22 different cell types was generated as a readily downloadable heat map tables.

We had two different outputs for absolute and relative modes. The relative output was such that for each tumor sample, the sum of all estimated immune cell-type fractions equals 1. The absolute estimates of immune cells were derived from their respective relative fractions. Absolute mode, in CIBERSORT, scales estimated relative cellular fractions into a score that reflects the absolute proportion of each cell type in the

tumor. The absolute immune score is estimated by the median expression level of all genes in the signature matrix (LM22) divided by the median expression level of all genes in the tumor sample. Although it is not expressed as a fraction, the absolute score can be directly compared across cell types (i.e., relative differences between cell types are maintained).

The output provides relative fractions and absolute scores for the following cells: B cells (naïve and memory), Plasma cells, T cells CD4 (naïve, memory resting, memory activated), T cells CD8, T cells follicular helper, T cells regulatory, T cells gamma delta, NK cells (resting and activated), Monocytes, Macrophages (M0, M1, M2), Dendritic cells (resting and activated), Mast cells (resting and activated), Eosinophils, Neutrophils. Additionally, it also provides a p-value (reflecting the statistical significance of the deconvolution result); Pearson's correlation coefficient (R), generated from comparing the leukocyte gene signature (original mixture) with the estimated immune fractions (imputed mixture) and root mean squared error between the original mixture and the imputed mixture.

In TCGA, GEPs were available for 386 oral cavity or oropharyngeal cancers. In the FHCRC dataset, 6 of the 166 samples overlapped between TCGA and FHCRC datasets. Therefore, 386 tumor samples from TCGA and 160 tumors from the Oralchip study were included in the analysis.

### **Estimation of HPV positivity through CDKN2a expression**

The TCGA Network had estimated HPV status for only 204 of 386 included cases. [21] Studies have found that p16 protein coded by gene CDKN2a (cyclin-dependent kinase Inhibitor 2A) is overexpressed in HPV positive oral cancer tumors. [22, 23] We graphed the CDKN2a counts for known HPV positive and negative samples. (Fig A) We could see this in our sample as well. Thus, we used logistic regression to predict HPV status for the remaining 182 cases with CDKN2a and tumor location (oral/ oropharyngeal) as the predictors, and the HPV status of 204 cases as outcome variable. We graphed the sensitivity and specificity of our logistic model (Fig B) and selected 0.25 as our cutoff (highest sensitivity and specificity). All samples with probability value equal to or greater than 0.25 were assumed to be HPV positive. Such estimation was not required for the Oralchip data, which had HPV status information based on results of PCR and a Roche HPV Linear Array testing for all 160 included samples.



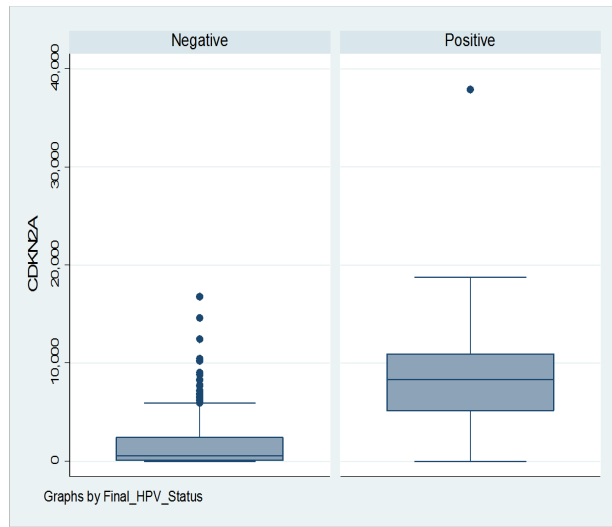


Fig A: Box plots of CDKN2a expression by HPV positivity status.

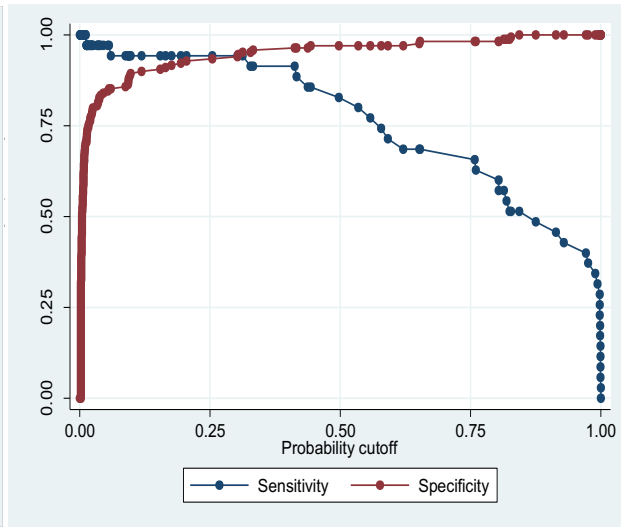


Fig B: Sensitivity and specificity plot of the logistic model fit to estimate HPV positivity for 182 samples

### Statistical analysis

Both absolute and relative estimates of the 22 immune cell types (exposure variable) were used to evaluate our study aims.

Both the TCGA and Oralchip samples, with a combined total of 546 samples, were initially divided into four categories: HPV+ Oral cavity tumors (OCC) (n=67), HPV- OCC tumors (n=376), HPV+ Oropharyngeal tumors (OPC) (n=69) and HPV- OPC tumors (n=34). Kaplan Meier survival curves were plotted to observe the difference in survival by tumor's HPV status and location in OSCC patients. To assess the difference between the immunome of HPV-positive OPC patients and other OSCC patients, we used non-parametric Wilcoxon rank sum test.

The Kruskal Wallis nonparametric test was used to compare cell distributions by tumor stage for samples in which this information was available (n=493).

Median relative fractions and absolute scores were reported for all 22 cell types.

Independent prognostic value of estimated immune cells fractions and scores of TCGA samples was assessed by a multivariable Cox regression analysis with time to death as outcome variable and age at diagnosis, HPV status, tumor stage and location, smoking and alcohol history as covariates. We used listwise deletion method for handling of missing data. With this method, an entire sample is excluded from

analysis if any single value is missing for the variables used in the multivariable Cox regression. Hazard ratios for all 22 immune cell populations were estimated. In our analysis, we multiplied the relative fractions or absolute scores of each cell type by 10 so that each unit change of these quantities is equivalent to a change of 0.1. This facilitates the interpretation of regression coefficient. For example, a regression coefficient of 0.5 for relative fraction of B cell means when the relative fraction of B cell increases 0.1, the hazard increases  $\exp(0.5) = 1.65$ . The OralChip study samples were used to validate this prognostic model. We also performed subset analysis of 50 HPV-positive oropharyngeal tumors in which we evaluated through Cox regression analysis if any of the 22 immune cells are associated with survival independent of stage, age at diagnosis, smoking and alcohol history.

Benjamin-Hochberg (BH) step up procedure was used to control the false discovery rate at 5%. The overall corrected p values were reported for each statistical test performed. They were very similar for the same tests but different exposure (relative or absolute scale). However, they were quite different for the same test but different subpopulation comparison (comparison between HPV+ and HPV- cases will have different corrected p-value than the comparison test between HPV+ OPC and HPV- OPC- cases).

All statistical tests were two-sided. All analyses were run using STATA 14.

## **RESULTS:**

The characteristics of the TCGA and Oralchip study populations are shown in Tables 1a and 1b, respectively. In general, the OSCC cases tended to be in the oral cavity. Most persons with these tumors were white, older than 50 years at diagnosis, male, and current smokers, and most had advanced disease.

### **Composition of tumor infiltrating immune cells:**

The most frequently encountered immune cell populations in the TCGA tumors are illustrated in Figure 1a. According to the bar plot, Macrophages (M0, M1, and M2) were present in almost all samples. Plasma cells, cytotoxic T cells (CD8+) and T CD4 memory cells were also found in most samples. T gamma delta cells, on the other hand, were absent from all but 10% of the sample. Memory B cells and naïve CD4 T cells along with monocytes and eosinophils were more often absent in these tumors.

**Table 1a:** Baseline patient and primary tumor characteristics of TCGA samples.

Variable	No. of samples (386)	%	Valid %
<b>Age at diagnosis (years)</b>			
<i>&lt;=50</i>	65	16.8	16.9
<i>&gt;50</i>	320	82.9	83.1
<i>Missing</i>	1	0.3	
<b>Gender</b>			
<i>Male</i>	273	70.7	70.7
<i>Female</i>	113	29.3	29.3
<b>Tumor location</b>			
<i>Oral cavity</i>	330	85.5	85.5
<i>Oropharyngeal region</i>	56	14.5	14.5
<b>Tumor stage</b>			
<i>Stage I</i>	23	6	6.9
<i>Stage II</i>	61	15.8	18.3
<i>Stage III</i>	64	16.6	19.2
<i>Stage IV</i>	186	48.2	55.7
<i>Missing</i>	54	13.5	
<b>HPV status (Oral Cavity)</b>			
<i>Positive</i>	19	5.8	10.4
<i>Negative</i>	163	49.4	89.6
<i>Missing</i>	148	45	
<b>HPV status (Oropharyngeal)</b>			
<i>Positive</i>	16	28.6	72.7
<i>Negative</i>	6	10.7	27.3
<i>Missing</i>	34	60.7	
Variable	No. of samples (386)	%	Valid %
<b>Race</b>			
<i>White</i>	337	87.3	89.9
<i>African American</i>	28	7.3	7.5
<i>Asian</i>	9	2.3	2.4
<i>American Indian</i>	1	0.3	0.2
<i>Missing</i>	11	2.9	
<b>Ethnicity</b>			
<i>Hispanic</i>	19	5	5.3
<i>Non-Hispanic</i>	341	88.3	94.7
<i>Missing</i>	26	8.7	
<b>Smoking history</b>			
<i>Yes</i>	276	71.5	72.8
<i>No</i>	103	26.7	27.2
<i>Missing</i>	7	1.8	
<b>Alcohol history</b>			
<i>Yes</i>	260	67.4	69
<i>No</i>	117	30.3	31
<i>Missing</i>	9	2.3	
<b>History of prior/other malignancy</b>			
<i>Yes</i>	22	5.7	5.7
<i>No</i>	364	94.3	94.3

**Table 1b: Baseline patient and tumor characteristics of Oralchip**

Variable	No. of samples (166)	%	Variable	No. of samples (386)	%
<b>Age at diagnosis (years)</b>			<b>Race</b>		
<=50	38	22.9	<i>White</i>	151	91
>50	128	77.1	<i>Non-white</i>	15	9
<b>Gender</b>			<b>Smoking history</b>		
<i>Male</i>	119	71.7	<i>Yes</i>	130	78.3
<i>Female</i>	47	28.3	<i>No</i>	36	21.7
<b>Tumor location</b>			<b>Alcohol history</b>		
<i>Oral cavity</i>	115	69.3	<i>Yes</i>	147	9.8
<i>Oropharyngeal region</i>	51	30.7	<i>No</i>	16	90.2
<b>Tumor stage</b>			<i>Missing</i>		
<i>Stage I</i>	36	21.8	3		
<i>Stage II</i>	16	9.7			
<i>Stage III</i>	22	13.3			
<i>Stage IV</i>	91	55.2			
<b>HPV status (Oral Cavity)</b>					
<i>Positive</i>	19	16.5			
<i>Negative</i>	96	83.5			
<b>HPV status (Oropharyngeal)</b>					
<i>Positive</i>	34	66.7			
<i>Negative</i>	17	33.3			

Fig 1b is a similar plot corresponding to the Oralchip data. Table 2 and Table 3 provide details about the median fractions and scores of 22 immune cell subpopulations. The column “A” presents the absolute score while column “R” displays the relative fractions.

Figure 1a on the left shows a stacked bar graph of percentage of tumor samples by 22 different immune cell types. The dark grey bar depicts the no. of samples in which an immune cell was present and the light grey bar shows the no. of samples in which it was absent. For example, naïve B-cell were elicited in only 70% of sample whereas Macrophages M0 were present in 99.5% tumor samples. 1b is a similar graph of Oralchip data.

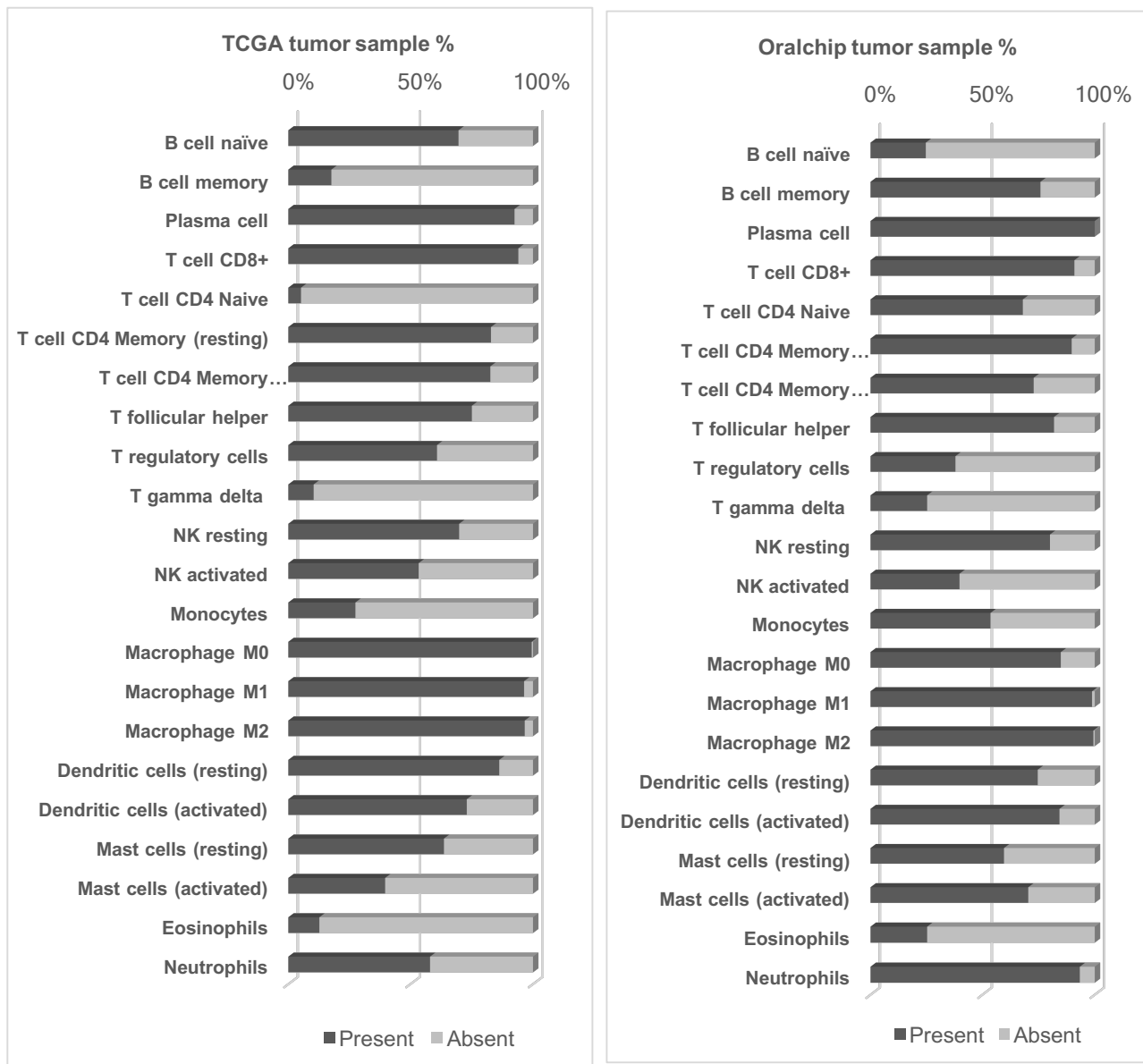


Figure 2a on the top left shows a stacked bar plot of percentage of HPV+ tumor samples by 22 different immune cell types in TCGA data while the figure 2b on top right shows HPV- tumor samples. Figure 2c (bottom left) and 2d (bottom right) are similar plots of HPV positive and negative samples from the Oralchip data.

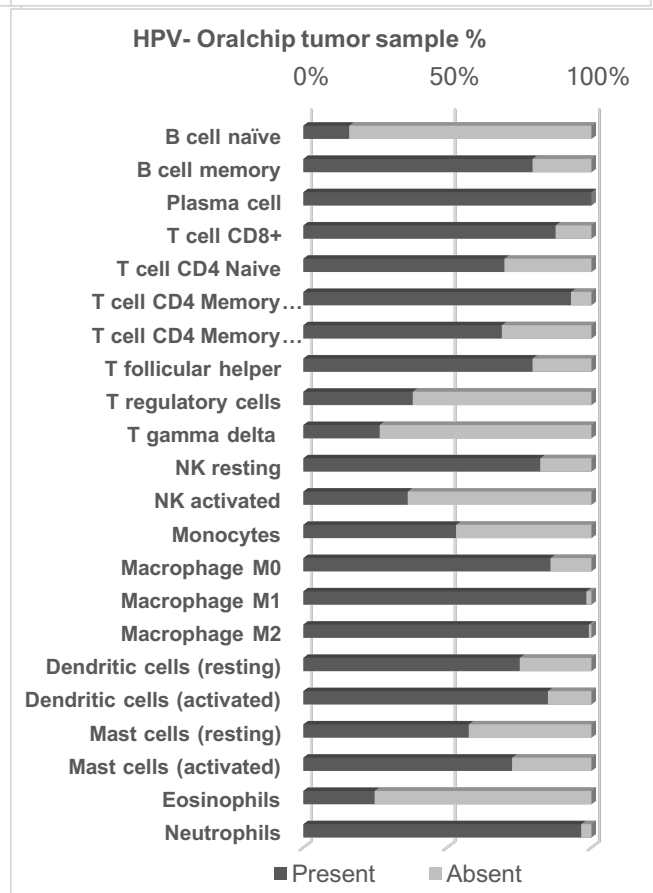
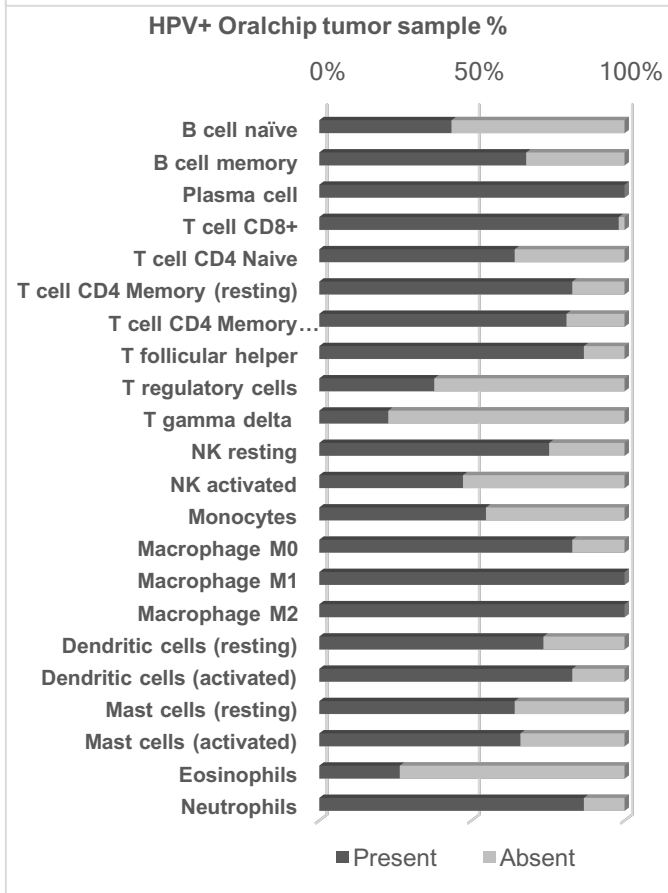
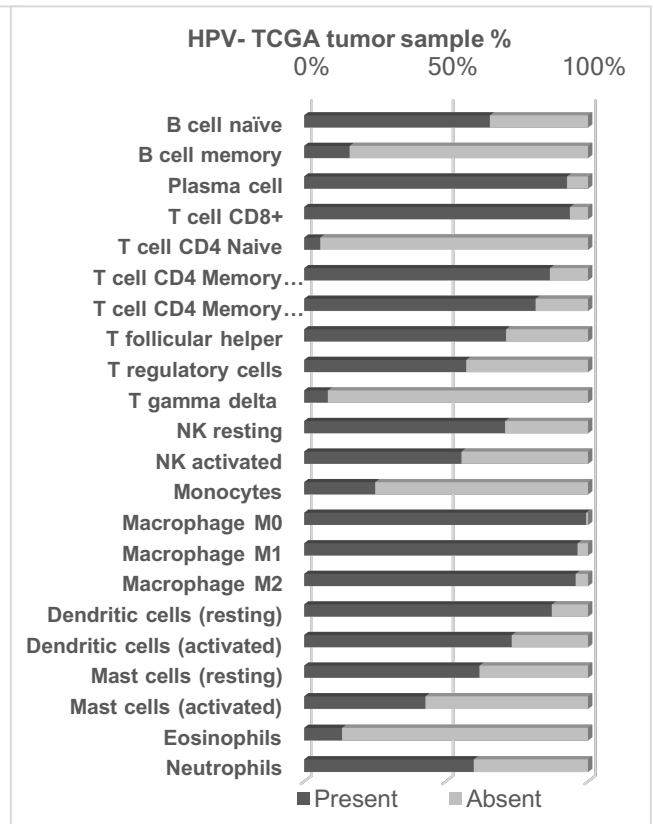
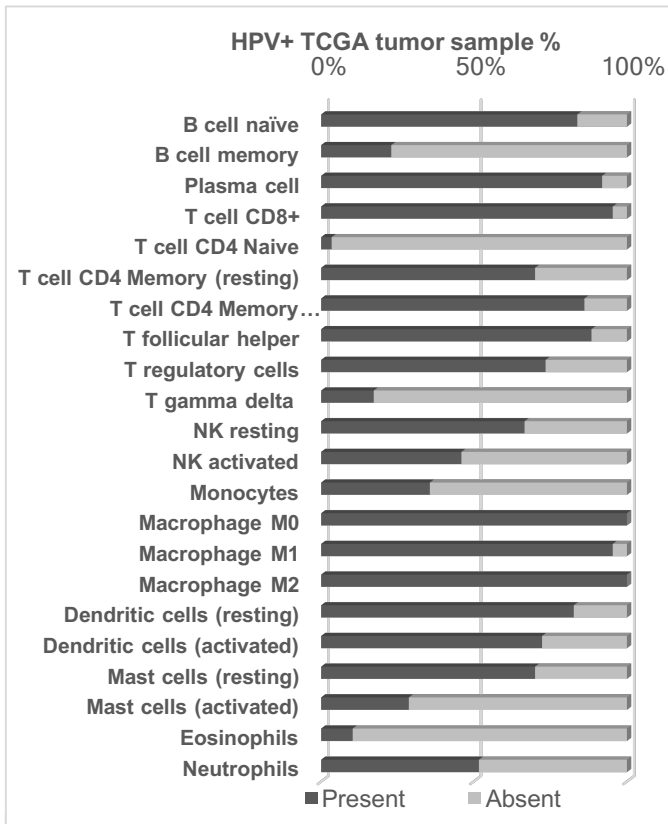


Table 2: Difference in the immune cell **absolute scores (A)** and **relative fractions (R)** by HPV status and location. P values calculated from nonparametric Wilcoxon ranksum test which compared HPV+ tumor samples to HPV- tumors; HPV+ OPC vs. HPV- OPC cases; and HPV+ OCC vs HPV- OCC cases. These results pertain to the combined tumor samples of TCGA and Oralchip data (n=546).

Median proportions	B cells naive		B cells memory		Plasma cells		T cells CD8		T cells CD4 naive		CD4 memory (resting)	
	A	R	A	R	A	R	A	R	A	R	A	R
OPC HPV-positive cases	0.1	0.01	0	0	1	0.1	2.7	0.12	0	0	0.24	0.05
OPC HPV-negative cases	0	0	0	0	0.5	0.13	0.78	0.05	0	0	0.27	0.06
OCC HPV-positive cases	0.07	0.01	0	0	0.16	0.03	1.01	0.08	0	0	0.3	0.05
OCC HPV-negative cases	0.008	0.001	0	0	0.15	0.02	0.65	0.06	0	0	0.37	0.07
Total	0.02	0.003	0	0	0.25	0.03	0.83	0.07	0	0	0.34	0.06
P-value (HPV+ vs HPV-)	<b>0.0006</b>	<b>0.0021</b>	0.09	0.2	<b>0.00001</b>	<b>0.002</b>	<b>0.00001</b>	<b>0.00001</b>	0.6	0.6	0.02	<b>0.001</b>
P-value (OPC+ vs OPC-)	<b>0.007</b>	<b>0.003</b>	0.9	0.5	0.08	0.8	<b>0.00001</b>	<b>0.00001</b>	0.8	0.8	0.7	0.3
P-value (OCC+ vs OCC-)	0.2	0.2	0.7	0.7	0.8	0.9	0.01	0.017	0.7	0.7	0.05	0.019

Median proportions	CD4 memory (activated)		T cells follicular helper		T cells regulatory		T cells gamma delta		NK cells resting		NK cells activated	
	A	R	A	R	A	R	A	R	A	R	A	R
OPC HPV-positive cases	0.32	0.05	0.24	0.04	0.06	0.01	0	0	0.08	0.01	0	0
OPC HPV-negative cases	0.12	0.02	0.1	0.02	0	0	0	0	0.09	0.02	0	0
OCC HPV-positive cases	0.28	0.06	0.18	0.03	0.02	0.002	0	0	0.1	0.02	0.004	0.0008
OCC HPV-negative cases	0.17	0.03	0.1	0.02	0.006	0.001	0	0	0.09	0.02	0.008	0.002
Total	0.19	0.03	0.12	0.02	0.02	0.002	0	0	0.09	0.02	0	0
P-value (HPV+ vs HPV-)	<b>0.004</b>	<b>0.001</b>	<b>0.00001</b>	<b>0.00001</b>	<b>0.0006</b>	<b>0.0005</b>	0.1	0.1	0.5	0.1	0.7	0.6
P-value (OPC+ vs OPC-)	<b>0.004</b>	<b>0.01</b>	<b>0.0002</b>	<b>0.007</b>	<b>0.0021</b>	<b>0.004</b>	0.8	0.8	0.7	0.3	0.9	0.8
P-value (OCC+ vs OCC-)	0.04	0.016	0.0015	0.0002	0.5	0.4	0.08	0.08	0.6	0.3	0.5	0.6

Median proportions	Monocytes		Macrophages M0		Macrophages M1		Macrophages M2		Dendritic cells (resting)		Dendritic cells (activated)	
	A	R	A	R	A	R	A	R	A	R	A	R
OPC HPV-positive cases	0	0	0.6	0.08	0.6	0.08	0.6	0.08	0.07	0.01	0.06	0.01
OPC HPV-negative cases	0	0	0.5	0.08	0.2	0.05	0.6	0.12	0.08	0.02	0.11	0.02
OCC HPV-positive cases	0	0	0.8	0.11	0.52	0.1	0.6	0.11	0.2	0.03	0.16	0.02
OCC HPV-negative cases	0	0	1.1	0.17	0.49	0.09	0.6	0.11	0.14	0.02	0.10	0.02
Total	0	0	1	0.15	0.5	0.09	0.6	0.11	0.14	0.02	0.1	0.02
P-value (HPV+ vs HPV-)	0.02	0.03	<b>0.0034</b>	<b>0.0001</b>	0.05	0.5	0.3	<b>0.002</b>	0.47	0.04	0.9	0.5
P-value (OPC+ vs OPC-)	0.8	0.8	0.32	0.8	<b>0.001</b>	<b>0.01</b>	0.8	<b>0.002</b>	0.9	0.35	0.08	<b>0.01</b>
P-value (OCC+ vs OCC-)	0.09	0.1	0.05	0.01	0.4	0.3	0.4	0.5	0.7	0.9	0.3	0.6

Median proportions	Monocytes		Macrophages M0		Macrophages M1		Macrophages M2		Dendritic cells (resting)		Dendritic cells (activated)	
	A	R	A	R	A	R	A	R	A	R	A	R
OPC HPV-positive cases	0	0	0.6	0.08	0.6	0.08	0.6	0.08	0.07	0.01	0.06	0.01
OPC HPV-negative cases	0	0	0.5	0.08	0.2	0.05	0.6	0.12	0.08	0.02	0.11	0.02
OCC HPV-positive cases	0	0	0.8	0.11	0.52	0.1	0.6	0.11	0.2	0.03	0.16	0.02
OCC HPV-negative cases	0	0	1.1	0.17	0.49	0.09	0.6	0.11	0.14	0.02	0.10	0.02
Total	0	0	1	0.15	0.5	0.09	0.6	0.11	0.14	0.02	0.1	0.02
P-value (HPV+ vs HPV-)	0.02	0.03	<b>0.0034</b>	<b>0.0001</b>	0.05	0.5	0.3	<b>0.002</b>	0.47	0.04	0.9	0.5
P-value (OPC+ vs OPC-)	0.8	0.8	0.32	0.8	<b>0.001</b>	<b>0.01</b>	0.8	<b>0.002</b>	0.9	0.35	0.08	<b>0.01</b>
P-value (OCC+ vs OCC-)	0.09	0.1	0.05	0.01	0.4	0.3	0.4	0.5	0.7	0.9	0.3	0.6

BH corrected critical p-value for comparing absolute scores (A) and relative fractions (R) of all HPV+ OSCC with all HPV – OSCC cases and HPV+ OPC with HPV- OPC cases was 0.02. BH corrected p-value for comparing absolute scores (A) and relative fractions (R) of comparison between HPV+ OCC with HPV- OSCC was 0.002.

### Composition of tumor infiltrating immune cells in OSCC tumors with different HPV status and

#### location:

All 546 samples from TCGA and Oralchip study were included in this analysis. Results from table 2 indicate that the immune cell composition of HPV-positive tumors is distinct from HPV-negative tumors. On an absolute scale, higher concentration of naïve B cells, plasma cells, cytotoxic T cells, activated memory CD4+ T cells, follicular helper and regulatory T cells and a lower concentration of undifferentiated macrophages, neutrophils is seen in HPV-positive cases. Results were similar on the relative scale with a few differences: lower concentration of M2 phase macrophages in HPV-positive cases. When we further stratified by location and compare HPV+ OPC with HPV- OPC, we observed a higher concentration of naïve B cells, cytotoxic T cells, activated memory CD4+ T cells, follicular helper and regulatory T cells, M1 macrophages was seen in HPV+ OPC cases. A lower concentration of activated dendritic cells, M2 macrophages and neutrophils was observed in HPV+ OPC cases.



Table 3: Difference in the immune cell absolute scores (A) and relative fractions (R) among cases by tumor stage. P values calculated from nonparametric Kruskal-Wallis test. (TCGA and Oralchip data combined samples (n= 498))

Median proportions by:	B cells naive		B cells memory		Plasma cells		T cells CD8		T cells CD4 naive		T cells CD4 memory(resting)	
	A	R	A	R	A	R	A	R	A	R	A	R
Stage I	0	0	0.08	0.016	0.5	0.12	1.44	0.06	0	0	0.24	0.06
Stage II	0.065	0.009	0	0	0.15	0.02	0.8	0.06	0	0	0.36	0.07
Stage III	0.047	0.008	0	0	0.17	0.02	0.64	0.06	0	0	0.26	0.04
Stage IV	0.009	0.001	0	0	0.22	0.03	0.8	0.06	0	0	0.32	0.07
<b>P value</b>	<b>0.006</b>	<b>0.004</b>	<b>0.002</b>	<b>0.001</b>	<b>0.006</b>	<b>0.001</b>	0.02	0.7	<b>0.001</b>	<b>0.001</b>	0.4	0.6

Median proportions	T cells CD4 memory(activated)		T cells follicular helper		T cells regulatory		T cells gamma delta		NK cells resting		NK cells activated	
	A	R	A	R	A	R	A	R	A	R	A	R
Stage I	0.09	0.02	0.11	0.3	0.03	0.008	0	0	0.9	0.02	0.003	0.0008
Stage II	0.27	0.04	0.15	0.3	0.03	0.005	0	0	0.06	0.01	0.056	0.01
Stage III	0.15	0.03	0.13	0.2	0.006	0.001	0	0	0.09	0.02	0.007	0.0004
Stage IV	0.16	0.03	0.09	0.2	0	0	0	0	0.1	0.02	0	0
<b>P value</b>	0.2	0.4	0.02	0.01	0.05	0.09	0.2	0.2	0.6	0.3	0.02	0.04

Median proportions	Monocytes		Macrophages M0		Macrophages M1		Macrophages M2		Dendritic cells (resting)		Dendritic cells (activated)	
	A	R	A	R	A	R	A	R	A	R	A	R
Stage I	0	0	0.5	0.1	0.5	0.1	0.6	0.12	0.1	0.03	0.08	0.02
Stage II	0	0	1.1	0.15	0.06	0.09	0.7	0.09	0.2	0.04	0.1	0.03
Stage III	0	0	1.1	0.15	0.5	0.09	0.6	0.11	0.1	0.02	0.1	0.02
Stage IV	0	0	1	0.16	0.4	0.08	0.6	0.11	0.1	0.02	0.08	0.02
<b>P value</b>	0.3	0.2	<b>0.002</b>	<b>0.0005</b>	0.2	0.3	0.7	0.2	0.2	0.2	0.06	0.2

Median proportions	Mast cells (resting)		Mast cells (activated)		Eosinophils		Neutrophils	
	A	R	A	R	A	R	A	R
Stage I	0.18	0.03	0.03	0.007	0	0	0.08	0.02
Stage II	0.16	0.03	0	0	0	0	0.04	0.004
Stage III	0.08	0.02	0.03	0.005	0	0	0.03	0.005
Stage IV	0.08	0.01	0.006	0.001	0	0	0.03	0.005
<b>P value</b>	0.17	0.2	0.15	0.1	0.03	0.02	<b>0.002</b>	<b>0.004</b>

BH corrected p-value for comparing absolute scores (A) by tumor stage was 0.015. Thus, all tests with p values less than 0.015 were considered significant. BH corrected p-value for comparing relative fractions (R) by tumor stage was 0.02. Thus, all tests with p values less than 0.02 were considered significant.

### Composition of tumor infiltrating immune cells in OSCC tumors by tumor stage:

493 cases that had tumor stage information were analyzed to assess difference in the concentration of immune cells by cancer stage. The concentrations of memory B cells, plasma cells, cytotoxic T cells, activated memory T cells, follicular helper and regulatory T cells were higher in earlier stage tumor sample, whereas the levels of undifferentiated macrophages and neutrophils decreased with advancing tumor stage. Similar results were obtained for the separate TCGA and Oralchip datasets.

### Subpopulations of immune cells as predictors of survival in OSCC:

To assess the individual prognostic value of each immune phenotype we divided the samples into training and validation sets. TCGA data were used as the training set. Of the 386 patient samples, 2 were excluded

Table 4: Prognostic value of each immune cell type for overall survival of OSCC patients adjusted for age at diagnosis, HPV status, tumor location, tumor stage, smoking and alcohol history (TCGA data, Complete case analysis, n = 319)

Immune cell type	Absolute scores as exposure		Relative fractions as exposure	
	Hazard ratio(95% CI)	P-value	Hazard ratio (95% CI)	P-value
B cell naïve	0.79 (0.54-1.18)	0.2	0.76 (0.49-1.20)	0.2
B cell memory	0.45 (0.03-6.60)	0.6	1.65 (0.44-6.14)	0.5
Plasma cell	0.60 (0.35-1.05)	0.08	0.72 (0.45-1.16)	0.183
T cell CD8+	0.83 (0.69-0.99)	0.04	0.87 (0.70-1.06)	0.184
T cell CD4 Naive	9.96 (0.21-463.6)	0.2	3.63 (0.96-13.6)	0.053
T cell CD4 Memory (resting)	0.93 (0.76-1.13)	0.5	1.02 (0.83-1.26)	0.8
T cell CD4 Memory (activated)	0.72 (0.51-1.03)	0.07	0.86 (0.61-1.21)	0.4
T follicular helper	0.44(0.22-0.88)	0.02	0.61 (0.31-1.18)	0.143
T regulatory cells	0.32 (0.11-0.88)	0.03	0.38 (0.13-1.07)	0.07
T gamma delta	0.06 (0.0003-9.52)	0.3	1.07 (0.05-23.07)	0.9
NK resting	0.60 (0.32-1.13)	0.1	1.04 (0.60-1.83)	0.8
NK activated	1.65 (0.66-4.13)	0.3	2.07 (0.99-4.33)	0.05
Monocytes	0.86 (0.06-12.56)	0.9	1.29 (0.27-6.16)	0.75
Macrophage M0	0.85 (0.73-0.99)	0.03	1.00 (0.90-1.11)	0.05
Macrophage M1	0.95 (0.75-1.21)	0.7	1.14 (0.89-1.45)	0.3
Macrophage M2	0.92 (0.69-1.22)	0.6	1.19 (0.85-1.67)	0.3
Dendritic cells (resting)	0.98(0.64-1.52)	0.9	1.07 (0.78 -1.45)	0.7
Dendritic cells (activated)	0.86 (0.43-1.74)	0.7	1.5 (0.74-1.78)	0.5
Mast cells (resting)	0.66 (0.37-1.19)	0.2	0.78 (0.44 -1.37)	0.4
Mast cells (activated)	1.61 (0.96-2.68)	0.07	1.42 (1.04-1.95)	0.03
Eosinophils	75.43 (0.29-19591.72)	0.1	22.68(1.21-425.5)	0.04
Neutrophils	0.60(0.14-2.44)	0.5	0.82 (0.31-2.20)	0.7

BH corrected critical p-value for both absolute scores and relative fractions is 0.0022

because of missing follow-up information. Age at diagnosis was unavailable for 1 case, tumor stage was missing for 52 cases. Smoking history was missing for 7 patients and alcohol history was missing for 9

patients. These subjects, too, were excluded from the analysis. Of the 319 included in the multivariable Cox regression analysis, 140 died during follow up. The follow up period was measured in days and it ranged from 2 to 5480 days with the mean being 845 days. Table 4 lists the hazard ratios estimated from our multivariate model in the TCGA data set.

We found that after adjusting for age at diagnosis, tumor stage, tumor location, HPV status, smoking and alcohol history on both absolute and relative scale, patients with 0.1 higher value of absolute score or relative fractions of undifferentiated macrophages, cytotoxic T cells, follicular helper and regulatory T cells were associated with better overall survival (p-value<0.05), though these associations were not statistically significant using the BH corrected p-value of 0.002. Similarly, a 0.1 higher value of absolute score or relative fractions of activated mast cells and eosinophils were associated with a poor prognosis among these patients, but these associations also were not significant by the corrected p-value (p value < 0.002).

**Validation of results:**

Table 5: Prognostic value of each immune cell type for overall survival among OSCC patients adjusted for age at diagnosis, HPV status, tumor stage location, Smoking and alcohol history (Oralchip study, n = 156)

Immune cell type	Absolute scores as exposure		Relative fractions as exposure	
	Hazard ratio(95% CI)	P-value	Hazard ratio (95% CI)	P-value
B cell naïve	0.03 (0.0004-1.94)	0.09	0.21 (0.03-1.37)	0.1
B cell memory	0.77 (0.29-2.04)	0.6	0.94 (0.63-1.41)	0.7
Plasma cell	1.06 (0.70-1.62)	0.8	1.08 (0.91-1.28)	0.4
T cell CD8+	0.92 (0.83 -1.01)	0.1	0.71 (0.45-1.12)	0.1
T cell CD4 Naive	1.11 (0.24-4.98)	0.1	1.13 (0.60-2.13)	0.7
T cell CD4 Memory (resting)	1.39 (0.42-4.56)	0.6	1.17 (0.73-1.86)	0.5
T cell CD4 Memory (activated)	0.26 (0.05-1.42)	0.1	0.54 (0.24-1.19)	0.1
T follicular helper	0.034 (0.002-0.59)	0.02	0.21 (0.06-0.74)	0.01
T regulatory cells	0.004 (1.65e-06 - 15.14)	0.2	0.18 (0.008-3.84)	0.3
T gamma delta	0.31 (0.02-3.46)	0.3	0.54 (0.17-1.7)	0.2
NK resting	1.55 (0.14-17.33)	0.4	1.34 (0.51-3.51)	0.6
NK activated	2.77 (0.07-105.51)	0.6	1.83 (0.43-7.81)	0.4
Monocytes	14.06 (0.41-481.5)	0.1	2.9 (0.68-12.33)	0.1
Macrophage M0	1.69 (0.88-3.28)	0.1	1.23 (0.95-1.60)	0.1
Macrophage M1	0.64 (0.25-1.68)	0.4	0.85 (0.55-1.30)	0.4
Macrophage M2	0.52 (0.19-1.40)	0.2	0.84 (0.57-1.25)	0.4
Dendritic cells (resting)	0.17 (0.02-1.28)	0.08	0.53 (0.24-1.14)	0.1
Dendritic cells (activated)	0.86 (0.24-3.15)	0.8	0.91 (0.53-1.55)	0.7
Mast cells (resting)	0.27 (0.07-0.97)	0.04	0.61 (0.36-1.03)	0.06
Mast cells (activated)	2.1 (0.99-4.45)	0.05	1.30 (0.97-1.74)	0.08
Eosinophils	47.07(0.42-5273.8)	0.1	4.27 (0.60-30.18)	0.1
Neutrophils	1.85 (0.54-6.39)	0.3	1.24 (0.75-2.08)	0.4

BH corrected critical p-value for both absolute scores and relative fractions is **0.0022**

For three patients in the FHCRC's Oralchip study there was no information regarding alcohol use, and on another there was no information regarding tumor stage. After excluding these cases, 156 cases were included in the Cox regression model used to validate the results. Of the 156, 104 patients died during follow up. The follow up period in days ranged from 14 to 4434 days with the mean value being 1962 days. The hazard ratio estimates and their 95% confidence intervals are presented in Table 5.

Analysis of the Oralchip data identified the same associations (except for regulatory T cells) seen in training dataset (TCGA sample), and also failed to reject the hypothesis of no association between higher concentration of immune cell phenotypes and overall survival.

**Subset analysis:**

Of the 69 persons with HPV-positive samples, 17 did not have information regarding tumor stage, and 2 had missing information about smoking and alcohol history. The remaining 50 were included in the regression analysis. The follow up period ranged from 2 to 4391 days. 20 cases died during follow up.

Table 6 gives the hazard ratios estimates and their 95% confidence intervals.

Table 6: Prognostic value of each immune cell type, adjusted for age at diagnosis, tumor location, tumor stage, smoking and alcohol history, in HPV-positive OPC samples from TCGA and Oralchip data combined. (Complete case analysis, n = 50)

Immune cell type	Absolute scores as exposure		Relative fractions as exposure	
	Hazard ratio(95% CI)	P-value	Hazard ratio (95% CI)	P-value
B cell naïve	0.006 (0.00003-1.23)	0.06	0.017038 ( 0.0004 -0.78)	0.04
B cell memory	2.02 (0.44-9.17)	0.4	1.426359 ( 0.66-3.07)	0.4
Plasma cell	0.73 (0.28-1.89)	0.5	1.09678 ( 0.75-1.61)	0.6
T cell CD8+	0.64 (0.49-0.83)	<b>0.001</b>	0.1088112 (0.034-0.35)	<b>&lt;0.0001</b>
T cell CD4 Naive	61.14 ( 1.08-3458.37)	0.04	8.992302 (1.85-43.6)	0.006
T cell CD4 Memory (resting)	1.38 (0.37-5.15)	0.6	1.897521 (0.91 -3.95)	0.09
T cell CD4 Memory (activated)	0.005 (0.0002-0.17)	<b>0.003</b>	0.0234634 (0.003-0.17)	<b>&lt;0.0001</b>
T follicular helper	0.06 ( .003-1.70)	0.1	0.2793594 (0.03- 2.35)	0.2
T regulatory cells	0.0006 ( 1.02e-07-3.73)	0.09	0.0012 (4.80e-06-0.327)	0.02
T gamma delta	0.28 (0.005-15.59)	0.5	0.63 (0.07-5.308)	0.7
NK resting	0.24 (0.012-4.6)	0.3	1.874838 (0.38-9.19)	0.4
NK activated	0.98 (0.07-13.87)	0.9	1.132388 (0.12-10.12)	0.9
Monocytes	164.88 (.0002-1.63e+08)	0.5	235.4322 (0.43-126111.7)	0.08
Macrophage M0	0.83 (0.31-2.201)	0.7	1.327616 (0.93-1.90)	0.1
Macrophage M1	0.12 (0.02-0.78)	0.03	0.3200701 (0.11-0.90)	0.032
Macrophage M2	0.07 (0.004-1.18)	0.07	0.8348502 (0.28-2.49)	0.746
Dendritic cells (resting)	1.80e-06 ( 1.78e-12-1.8)	0.06	0.0015661 (5.35e-06-0.46)	0.026
Dendritic cells (activated)	95.78 ( 5.24-1749.48)	<b>0.002</b>	15.43 (3.7-63.6)	<b>&lt;0.0001</b>
Mast cells (resting)	0.23 (.01-4.80)	0.3	0.67 (0.09-4.90)	0.7
Mast cells (activated)	7.96 (1.7-35.89)	<b>0.007</b>	2.36 (1.36-4.10)	<b>0.002</b>
Eosinophils	1.08e-08 ( 1.93e-24-6.07e+07)	0.3	0.003 (7.37e-10-12833.31)	0.5
Neutrophils	1542.90 ( 9.17-259371.7)	<b>0.005</b>	35.62 (3.38-387.06)	<b>0.003</b>

BH corrected critical p-value for both absolute scores and relative fractions is **0.01**

We found statistically significant associations with overall survival in five immune cell populations in the HPV-positive OPC tumors subset. Tumor sample groups that had higher concentration of tumor infiltrating immune cells like cytotoxic T cells, activated CD4+ T cells had lower risk of dying. On the other hand, higher concentrations of antigen presenting dendritic cells, activated mast cells and neutrophils were predictive of poor prognosis.

We did a similar analysis for 30 HPV-negative OPC samples. Their follow up time ranged from 37 to 4409 days. 16 died during follow-up. We saw no significant association between higher concentration of immune cell populations and survival in these patients (Supplementary Table 1).

## **DISCUSSION:**

We described the immune microenvironment of a large set of oral squamous cell carcinomas through in silico analysis using data from two independent prospective cohorts. We also assessed the independent prognostic value of 22 immune phenotypes while controlling for important clinicopathological parameters. We observed that the immune cell composition is different for HPV-positive and HPV-negative tumors. When we further stratified HPV-positive and HPV-negative samples by location, we saw that the immune cell composition of HPV+ OPC tumors was significantly different from HPV- OPC tumors. However, even though certain immune cell fractions and scores differed between HPV+ OCC and HPV- OCC, these differences were not statistically significant. We also observed that composition of certain immune cells differed by stage. Although we did not identify a statistically significant association between higher concentration of immune cell types and survival, we saw the potential predictive capacity of these immune cells which can help future researchers investigate this question.

In our study, we considered both the relative and absolute proportion of immune cell phenotypes. Absolute immune score scales linearly with tumor purity. On the other hand, relative fraction estimates the percentage of a given cell population in the total tumor infiltrate. In our sample, macrophages, cytotoxic T cells, plasma cells and activated CD4 memory cells were the most frequently observed immune phenotypes. In our study, both relative and absolute measurements provided similar results with respect to associations with HPV status, clinical stage, and survival.

When we stratified the tumor samples by HPV status and location, we found that oropharyngeal HPV-positive tumors had a significantly different immune composition when compared to their HPV-negative

counterparts and to oral cavity tumors. Krupar et al [27] observed a higher amount of lymphocytic infiltrate in HPV-positive oropharyngeal samples compared to HPV-negative tumors. They found a higher concentration of cytotoxic and activated helper T cells, which is similar to what was observed in our study. In contrast to our finding, they found no difference in concentration of regulatory T cells between HPV-positive and HPV-negative lesions. Other studies by Ward et al [28] and Badoual et al [29] also corroborated our findings with respect to higher concentration of cytotoxic, activated helper and regulatory T cells in OSCC tumors. No study (to our knowledge) has examined the concentrations of plasma cells, follicular helper cells, naïve B-cells, mast cells or macrophages with regards to HPV status. Previous studies have observed a higher proportion TILs present in early stage tumors [15,30,32], which might be expected given that immune suppression is a hallmark of advanced stage tumors. We saw a similar pattern among cell populations such as naïve B cells, plasma cells, cytotoxic T cells, regulatory T cells. However, composition of undifferentiated macrophages was higher among late stage tumors. This might be expected since tumor associated macrophages are often associated with neoplastic transformation, tumor immune evasion and subsequent metastasis [52]. Further examination by others with large data sets would be warranted.

Our survival analysis failed to identify a significant association between individual immune cell types and overall survival in OSCC patients after controlling for important clinical parameters such as age at diagnosis, tumor stage, tumor location, HPV status, and smoking and alcohol history. Previous studies have reported that higher concentration of cytotoxic T cells, activated memory T cells to be associated with longer survival in OSCC [15,16,29,31-37]. Though there was a suggestion of similar associations in our study, we could not exclude chance as a plausible explanation. These earlier studies have also evaluated the prognostic value of regulatory T cells, and have reported conflicting results. We observed a similar conflict: higher concentrations of regulatory T cells were associated with better survival in the TCGA data while no corresponding association was seen in the Oralchip data.

Some studies [28,29] evaluated the prognostic value of TILs exclusively in HPV-positive OPC patients. They found that certain immune populations like cytotoxic T cells (CD8+) and activated memory (CD4+) T cells are predictive of better prognosis after controlling for smoking status and stage in these patients. We confirmed these findings in our analysis of 50 HPV-positive OPC cases.

Additionally, we observed that an increased infiltration of activated mast cells was associated with poor survival in HPV-positive OPC patients. Previous studies reported an increased infiltration of mast cells in oral cancer. [38,40] Mast cells have been implicated in many studies to be a promoter of inflammatory response and angiogenesis in tumors, including OSCC [25]. Angiogenesis promotes neovascularization and metastases and is an important hallmark of cancer. [41-46] Mast cells have also been observed to be an independent marker for poor prognosis in several other cancers. [47-50] Similarly, higher concentration of neutrophils increase tumor inflammatory response. They have also been associated with poor prognosis in the pan cancer analysis. [7]

Our results suggest that HPV-positive OPC patients with a higher concentration of activated dendritic cells are at a higher risk of dying. Tumor-associated dendritic cells have been known to play a role in immune suppression through development of T cell tolerance and reducing T cell infiltrate in tumor. [26] However, Reichert et al, in a study of oral cavity tumors (which presumably were mostly HPV-negative) [17], found a lower concentration of activated dendritic cells to be associated with reduced survival. We saw a similar association in our analysis of 30 HPV negative OPC tumors but it was not statistically significant, perhaps due to the small sample size. To understand why activated dendritic cells may act as a tumor promoter in HPV positive tumors and tumor suppressor in HPV negative tumors, further research is needed.

There are a number of limitations of our study, and these call for a cautious interpretation of our results. Both HPV status and tumor stage information was missing for a substantial number of patients in TCGA. We dealt with the missing data by estimating HPV16 through CDKN2a expression which, though it is an established method, is still prone to error. Also, our estimates could be subject to residual confounding. For example, to the extent that HPV status has not been accurately categorized, we would not have fully taken this variable into account. Then, our model analyzed the effect of each immune cell individually, ignoring the possibility of an interaction between different immune cell populations. Lastly, there is the issue of missing information regarding the cause of death in the TCGA patients, which prevents us from measuring associations with disease-specific survival in OSCC patients, a more appropriate outcome measure.

The results of our study (and earlier ones) support the hypothesis that the immune microenvironment has a bearing on disease progression in patients with HPV-positive oropharyngeal cancer. Potentially, this could

have a bearing on the clinical management of these patients, considering the recent development of immune-modulating therapies.

## REFERENCES:

1. SEER Cancer Stat Facts: Oral Cavity and Pharynx Cancer. National Cancer Institute. Bethesda, MD, <http://seer.cancer.gov/statfacts/html/oralcav.html>
2. Sethi, S., Ali-Fehmi, R., Franceschi, S., Struijk, L., van Doorn, L. J., Quint, W., ... & Kato, I. (2012). Characteristics and survival of head and neck cancer by HPV status: a cancer registry-based study. *International journal of cancer*, 131(5), 1179-1186.
3. Hanahan, & Weinberg. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), 646-674..
4. Coussens, L., Zitvogel, L., & Palucka, A. (2013). Neutralizing Tumor-Promoting Chronic Inflammation: A Magic Bullet? *Science*, 339(6117), 286-291.
5. Bense, R.D., Sotiriou, C., Piccart-Gebhart, M.J., Haanen, J.B., van Vugt, M.A., de Vries, E.G., Schröder, C.P. and Fehrmann, R.S. (2017). Relevance of tumor-infiltrating immune cell composition and functionality for disease outcome in breast cancer. *Journal of the National Cancer Institute*, 109(1), djw192.
6. Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P. and Zinzindohoué, F., (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*, 313(5795), 1960-1964.
7. Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S.V., Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D. and Diehn, M. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nature medicine*, 21(8), 938-945.
8. Conway, C., Graham, J., Chengot, P., Daly, C., Chalkley, R., Ross, L., Stead, L. (2015). Elucidating drivers of oral epithelial dysplasia formation and malignant transformation to cancer using RNAseq. *Oncotarget*, 6(37), 40186-40201.
9. Newman, Aaron M, Chih Long Liu, Michael R Green, Andrew J Gentles, Weiguo Feng, Yue Xu, Chuong D Hoang, Maximilian Diehn, and Ash A Alizadeh. (2015) Robust Enumeration of Cell Subsets from Tissue Expression Profiles." *Nature Methods* 12(5), 453-7.
10. Angelova, Mihaela, Pornpimol Charoentong, Hubert Hackl, Maria L Fischer, Rene Snajder, Anne M Krogsdam, Maximilian J Waldner, Gabriela Bindea, Bernhard Mlecnik, Jerome Galon, and Zlatko Trajanoski. (2015) Characterization of the Immunophenotypes and Antigenomes of Colorectal Cancers Reveals Distinct Tumor Escape Mechanisms and Novel Targets for Immunotherapy. *Genome Biology* 16(1), *Genome Biology*, 2015, Vol.16(1). Web.
11. Fujita, Yohei, Okamoto, Masato, Goda, Hiroyuki, Tano, Tomoyuki, Nakashiro, Koh-Ichi, Sugita, Atsuro, Fujita, Tomonobu, Koido, Shigeo, Homma, Sadamu, Kawakami, Yutaka, and Hamakawa, Hiroyuki. (2014) Prognostic Significance of Interleukin-8 and CD163-Positive Cell-Infiltration in Tumor Tissues in Patients with Oral Squamous Cell Carcinoma. *PLoS One* 9(12), E110378.
12. Watanabe, Katou, Ohtani, Nakayama, Yoshie, & Hashimoto. (2010). Tumor-infiltrating lymphocytes, particularly the balance between CD8 + T cells and CCR4 + regulatory T cells, affect the survival of patients with oral squamous cell carcinoma. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontology*, 109(5), 744-752.
13. Costa, Nadia Lago, Andreia Souza Goncalves, Allisson Filipe Martins, Diego Antonio Arantes, Tarcilia Aparecida Silva, Aline Carvalho Batista, NI Lopes Martins, and As Costa Arantes. (2016) Characterization of Dendritic Cells in Lip and Oral Cavity Squamous Cell Carcinoma." *Journal Of Oral Pathology & Medicine* 45(6), 418-24.
14. Robins, H. S., Ericson, N. G., Guenthoer, J., O'Briant, K. C., Tewari, M., W.Drescher, C., & Bielas, J. H. (2013). Digital Quantification of Tumor Infiltrating Lymphocytes. *Science Translational Medicine*, 5(214), 214ra169. <http://doi.org/10.1126/scitranslmed.3007247>
15. Fridman, W. H., Pages, F., Sautès-Fridman, C., & Galon, J. (2012). The immune contexture in human tumours: impact on clinical outcome. *Nature Reviews Cancer*, 12(4), 298-306.
16. Badoual, C., Hans, S., Rodriguez, J., Peyrard, S., Klein, C., Agueznay, N. E. H., ... & Brasnu, D. F. (2006). Prognostic value of tumor-infiltrating CD4+ T-cell subpopulations in head and neck cancers. *Clinical Cancer Research*, 12(2), 465-472.



17. Reichert, T. E., Scheuer, C., Day, R., Wagner, W. & Whiteside, T. L. The number of intratumoral dendritic cells and zeta-chain expression in T cells as prognostic and survival biomarkers in patients with oral carcinoma. *Cancer* 91, 2136–2147(2001).
18. Shibuya, T. Y. et al. Clinical significance of poor CD3 response in head and neck cancer. *Clin. Cancer Res.* 8, 745–751 (2002)
19. Zhang, Y. L. et al. Different subsets of tumor infiltrating lymphocytes correlate with NPC progression in different ways. *Mol. Cancer* 9, 4 (2010)
20. Quan, H., Fang, L., Pan, H., Deng, Z., Gao, S., Liu, O., ... Tang, Z. (2016). An adaptive immune response driven by mature, antigen-experienced T and B cells within the microenvironment of oral squamous cell carcinoma. *International Journal of Cancer*, 138(12), 2952–2962.
21. The Cancer Genome Atlas, N. (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, 517(7536), 576-582. doi:10.1038/nature14129 <http://www.nature.com/nature/journal/v517/n7536/abs/nature14129.html#supplementary-information>
22. Reimers, N., Kasper, H. U., Weissenborn, S. J., Stützer, H., Preuss, S. F., Hoffmann, T. K., Speel, E. J. M., Dienes, H. P., Pfister, H. J., Guntinas-Lichius, O. and Klussmann, J. P. (2007), Combined analysis of HPV-DNA, p16 and EGFR expression to predict prognosis in oropharyngeal cancer. *Int. J. Cancer*, 120: 1731–1738. doi:10.1002/ijc.22355
23. Klussmann JP, Gultekin E, Weissenborn SJ, Wieland U, Dries V, Dienes HP, Eckel HE, Pfister HJ, Fuchs PG. Expression of p16 protein identifies a distinct entity of tonsillar carcinomas associated with human papillomavirus. *Am J Pathol* 2003; 162: 747–53.
24. Grimm, M., Feyen, O., Hofmann, H., Teriete, P., Biegner, T., Munz, A., & Reinert, S. (2016). Immunophenotyping of patients with oral squamous cell carcinoma in peripheral blood and associated tumor tissue. *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine*, 37(3), 3807–3816. <http://doi.org/10.1007/s13277-015-4224-2>
25. Coussens, L. M. et al. Inflammatory mast cells up-regulate angiogenesis during squamous epithelial carcinogenesis. *Genes Dev.* 13, 1382-1397 (1999).
26. Tran Janco, J. M., Lamichhane, P., Karyampudi, L., & Knutson, K. L. (2015). Tumor-Infiltrating Dendritic Cells in Cancer Pathogenesis. *The Journal of Immunology*, 194(7), 2985-2991. doi:10.4049/jimmunol.1403134
27. Krupar, R., Robold, K., Gaag, D., Spanier, G., Kreutz, M., Renner, K. . . . Bosserhoff, A. K. (2014). Immunologic and metabolic characteristics of HPV-negative and HPV-positive head and neck squamous cell carcinomas are strikingly different. *Virchows Archiv*, 465(3), 299-312. doi:10.1007/s00428-014-1630-6
28. Ward, M. J., Thirdborough, S. M., Mellows, T., Riley, C., Harris, S., Suchak, K. . . . Thomas, G. J. (2014). Tumour-infiltrating lymphocytes predict for outcome in HPV-positive oropharyngeal cancer. *Br J Cancer*, 110(2), 489-500. doi:10.1038/bjc.2013.639
29. Badoual, C., Hans, S., Merillon, N., Van Ryswick, C., Ravel, P., Benhamouda, N., . . . Tartour, E. (2013). PD-1–Expressing Tumor-Infiltrating T Cells Are a Favorable Prognostic Biomarker in HPV-Associated Head and Neck Cancer. *Cancer Research*, 73(1), 128-138. doi:10.1158/0008-5472.can-12-2606
30. Lei, Y., Xie, Y., Tan, Y. S., Prince, M. E., Moyer, J. S., Nör, J., & Wolf, G. T. (2016). Telltale Tumor Infiltrating Lymphocytes (TIL) in Oral, Head & Neck Cancer. *Oral Oncology*, 61, 159–165. <http://doi.org/10.1016/j.oraloncology.2016.08.003>
31. Nguyen, N., Bellile, E., Thomas, D., McHugh, J., Rozek, L., Virani, S., . . . Wolf, G. T. (2016). Tumor infiltrating lymphocytes and survival in patients with head and neck squamous cell carcinoma. *Head Neck*, 38(7), 1074-1084. doi:10.1002/hed.24406
32. Matlung, S. E., Wilhelmina van Kempen, P. M., Bovenschen, N., van Baarle, D., & Willems, S. M. (2016). Differences in T-cell infiltrates and survival between HPV+ and HPV- oropharyngeal squamous cell carcinoma. *Future Sci OA*, 2(1), Fso88. doi:10.4155/fso.15.88
33. Balermipas P, Rodel F, Weiss C, Rodel C, Fokas E. Tumor-infiltrating lymphocytes favor the response to chemoradiotherapy of head and neck cancer. *Oncoimmunology*. 2014;3(1):e27403
34. Lukesova E, Boucek J, Rotnaglova E, et al. High level of Tregs is a positive prognostic marker in patients with HPV-positive oral and oropharyngeal squamous cell carcinomas. *Biomed. Res. Int.* 2014;2014:303929.
35. Wansom, D., Light, E., Thomas, D., Worden, F., Prince, M., Urba, S., ... the UM Head and Neck SPORE Program. (2012). Infiltrating lymphocytes and human papillomavirus-16 associated oropharynx cancer. *The Laryngoscope*, 122(1), 121–127. <http://doi.org/10.1002/lary.22133>

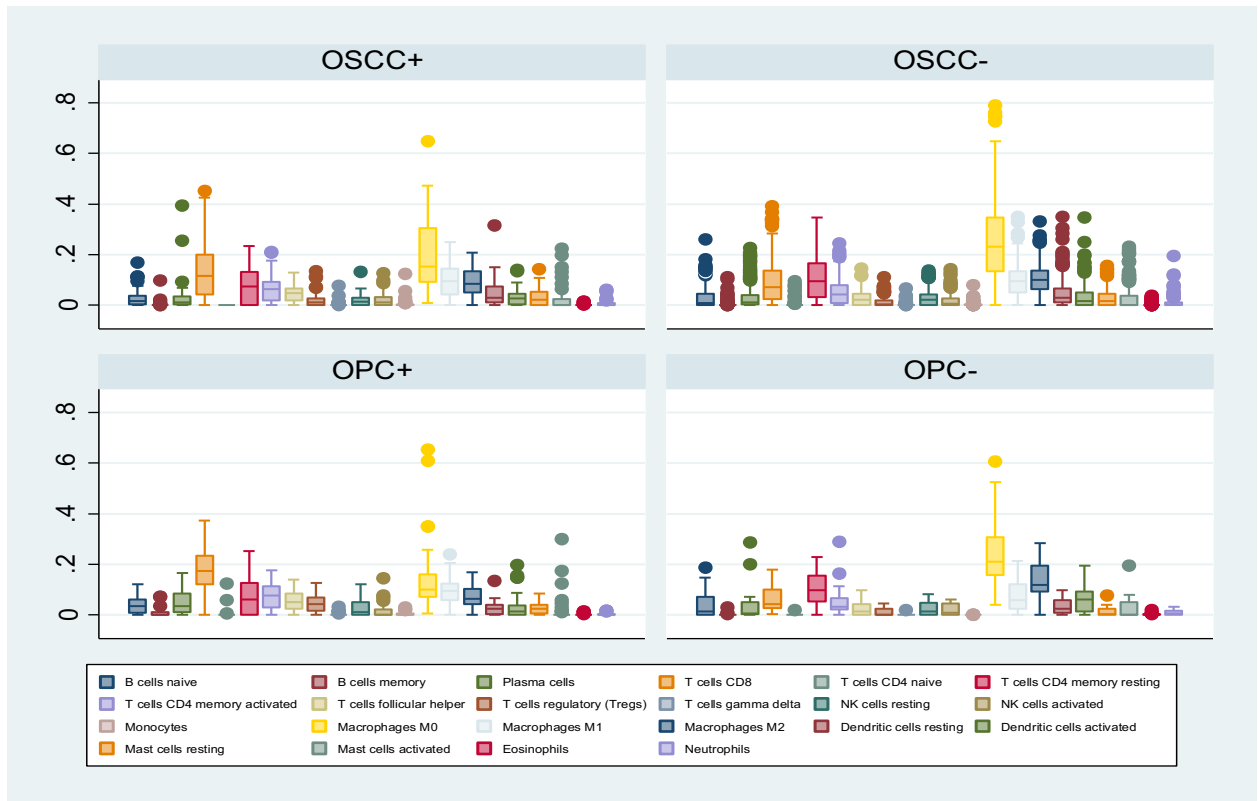
36. Eckert, A. W., Wickenhauser, C., Salins, P. C., Kappler, M., Bukur, J., & Seliger, B. (2016). Clinical relevance of the tumor microenvironment and immune escape of oral squamous cell carcinoma. *Journal of Translational Medicine*, 14(1), 85. doi:10.1186/s12967-016-0828-6
37. Strauss, L., Bergmann, C., Szczepanski, M., Gooding, W., Johnson, J. T., & Whiteside, T. L. (2007). A unique subset of CD4+CD25highFoxp3+ T cells secreting interleukin-10 and transforming growth factor-beta1 mediates suppression in the tumor microenvironment. *Clin Cancer Res*, 13(15 Pt 1), 4345-4354. doi:10.1158/1078-0432.ccr-07-0472
38. Ankle MR, Kale AD, Nayak R. Mast cells are increased in leukoplakia, oral submucous fibrosis, oral lichen planus and oral squamous cell carcinoma. *J Oral Maxillofacial Pathol* 2007; 11: 18-22
39. Gomes APN, Johann JE, Lovato GG, Ferreira AM. Comparative analysis of the mast cell density in normal oral mucosa, actinic cheilitis and lip squamous cell carcinoma. *Braz Den J* 2008; 19: 186-189
40. Rojas IG, Spencer ML, Martinez A, Maurelia MA, Rudolph MI. Characterization of mast cell subpopulations in lip cancer. *J Oral Pathol Med* 2005; 34: 269-273
41. Iamaroon, A., Pongsiriwet, S., Jittidecharaks, S., Pattanaporn, K., Prapayasadok, S. and Wanachantararak, S. (2003), Increase of mast cells and tumor angiogenesis in oral squamous cell carcinoma. *Journal of Oral Pathology & Medicine*, 32: 195–199. doi:10.1034/j.1600-0714.2003.00128.x
42. Elpek G, Gelen T, Aksoy NH, et al. The prognostic relevance of angiogenesis and mast cells in squamous cell carcinoma of the oesophagus. *J Clin Pathol* 2001; 54: 940–4.
43. Tomita M, Matsuzaki Y, Edagawa M, et al. Association of mast cells with tumor angiogenesis in esophageal squamous cell carcinoma. *Dis Esophagus* 2001; 14: 135–8.
44. Tomita M, Matsuzaki Y, Onitsuka T. Effect of mast cells on tumor angiogenesis in lung cancer. *Ann Thorac Surg* 2000; 69: 1686–90.
45. Benitez-Bribiesca L, Wong A, Utrera D, Castellanos E. The role of mast cell tryptase in neoangiogenesis of premalignant and malignant lesions of the uterine cervix. *J Histochem Cytochem* 2001; 49: 1061–2.
46. Williams, J. K., Carlson, G. W., Cohen, C., Derose, P. B., Hunter, S., & Jurkiewicz, M. J. (1994). Tumor angiogenesis as a prognostic factor in oral cavity tumors. *Am J Surg*, 168(5), 373-380.
47. D Ribatti, MG Ennas, A Vacca, F Ferreli, B Nico, S Orru, P Sirigu Tumor vascularity and tryptase-positive mast cells correlate with a poor prognosis in melanoma Eur J Clin Invest, 33 (2003), pp. 420–425
48. Mast cell infiltration around gastric cancer cells correlates with tumor angiogenesis and metastasis
49. Johansson, A., Rudolfsson, S., Hammarsten, P., Halin, S., Pietras, K., Jones, J., . . . Bergh, A. (2010). Mast Cells Are Novel Independent Prognostic Markers in Prostate Cancer and Represent a Target for Therapy. *The American Journal of Pathology*, 177(2), 1031-1041. doi:https://doi.org/10.2353/ajpath.2010.100070
50. E.R. Fisher, S.M. Paik, H. Rockette, J. Jones, R. Caplan, B. Fisher Prognostic significance of eosinophils and mast cells in rectal cancer: findings from the National Surgical Adjuvant Breast and Bowel Project (protocol R-01) *Hum. Pathol.*, 20 (1989), pp. 159–163
51. Chen C, Méndez E, Houck J, Fan W, Lohavanichbutr P, Doody D, Yueh B, Futran ND, Upton M, Farwell DG, Schwartz SM, Zhao LP. Gene expression profiling identifies genes predictive of oral squamous cell carcinoma. GEO Data Repository: GSE30784. *Cancer Epidemiol Biomarkers Prev*. 2008; 17(8):2152–62. PMID: 18669583, PMCID: PMC2575803.
52. Williams, C. B., Yeh, E. S., & Soloff, A. C. (2016). Tumor-associated macrophages: unwitting accomplices in breast cancer malignancy. *NPJ breast cancer*, 2.
53. Chaturvedi, A.K. Epidemiology and Clinical Aspects of HPV in Head and Neck Cancers. *Head and Neck Pathol* (2012) 6(Suppl 1): 16. doi:10.1007/s12105-012-0377-0.

**APPENDIX:**

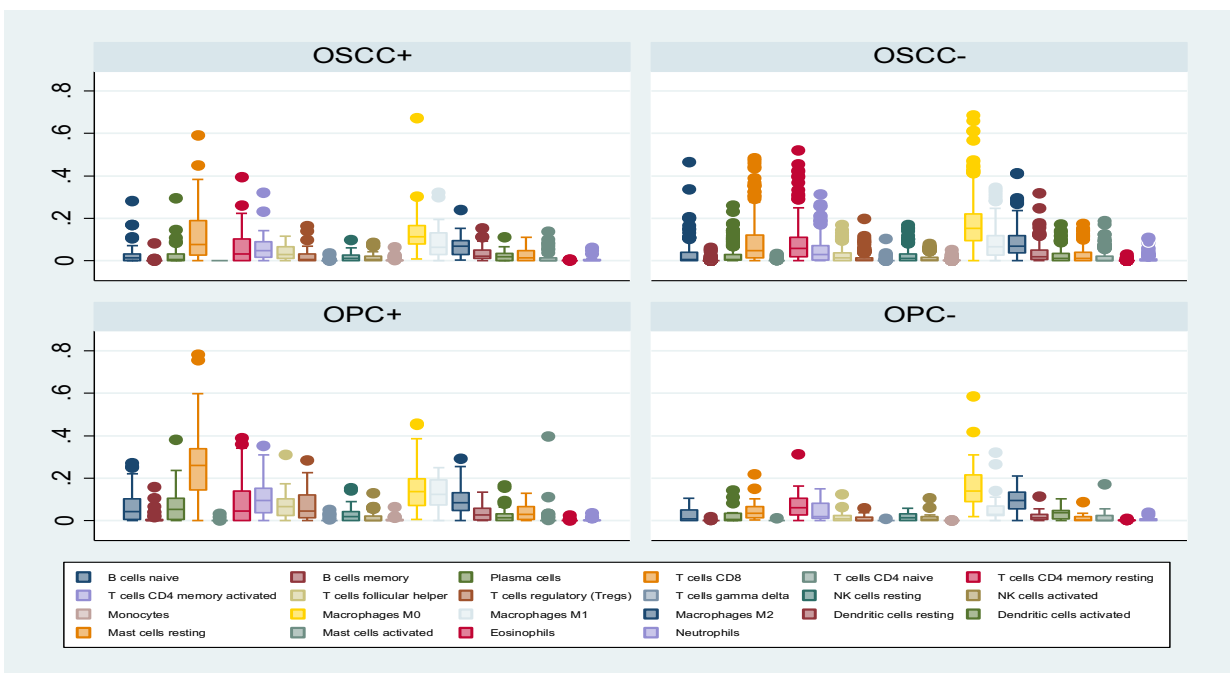
Supplementary Table 1: Prognostic value of each immune cell type, adjusted for age at diagnosis, tumor location, tumor stage, smoking and alcohol history, in HPV-negative OPC samples from TCGA and Oralchip data combined. (Complete case analysis, n = 30)

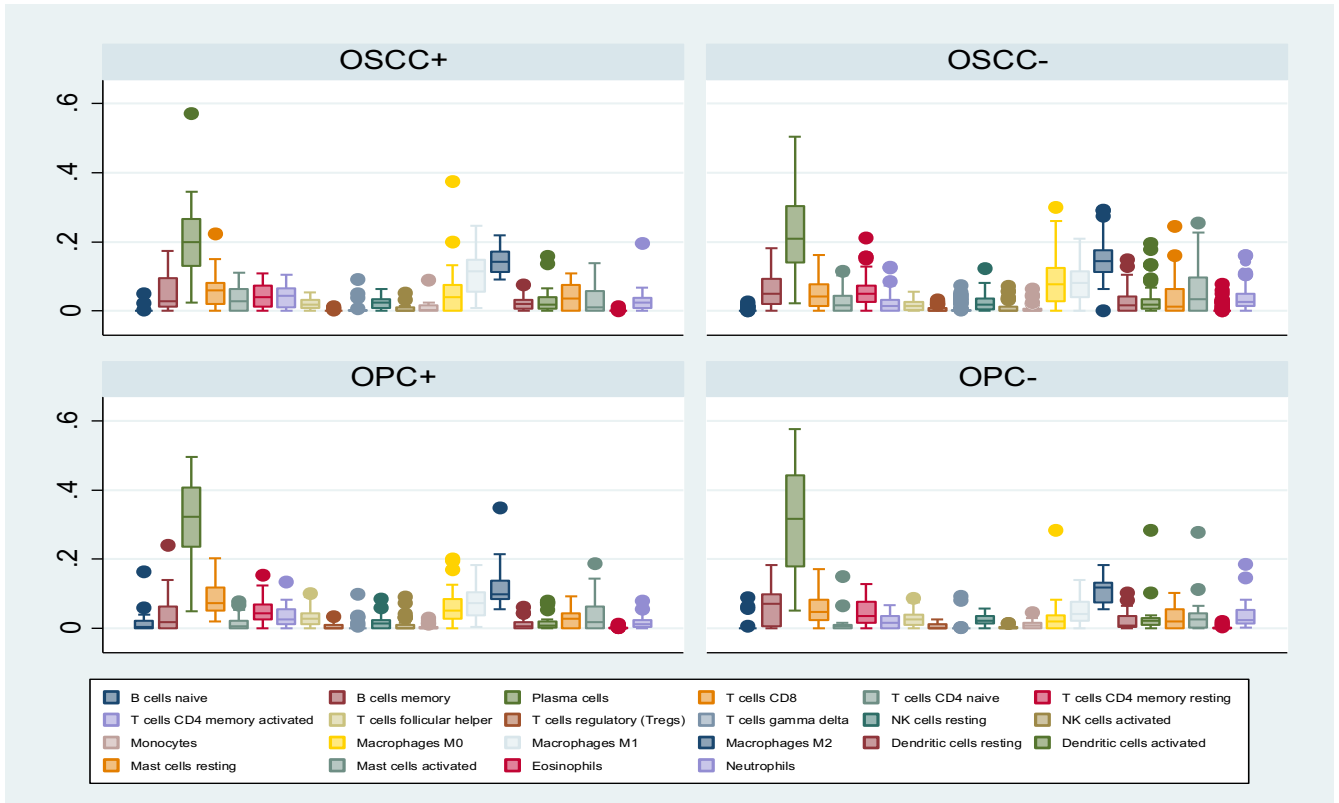
Immune cell type	Absolute scores as exposure		Relative fractions as exposure	
	Hazard ratio(95% CI)	P-value	Hazard ratio (95% CI)	P-value
B cell naïve	0.095 (0.002-4.39)	0.2	0.19 (0.012-2.96)	0.2
B cell memory	0.08 (0.002-3.46)	0.2	0.44 (0.11-1.71)	0.2
Plasma cell	1.34 (0.57 – 3.18)	0.5	1.1 (0.81-1.5)	0.5
T cell CD8+	1.14 ( 0.79-1.65)	0.5	2.07 (0.44-9.72)	0.4
T cell CD4 Naive	5538.9 (1.04-2.93e+07)	0.05	36.02 (0.99-1305.8)	0.05
T cell CD4 Memory (resting)	1.02 (0.26-4.04)	0.9	0.84 (0.28-2.53)	0.8
T cell CD4 Memory (activated)	0.33 (0.1-9.34)	0.5	1.27 (0.15-11.036)	0.8
T follicular helper	0.099(0.002-4.5)	0.2	0.15 (0.01-1.95)	0.1
T regulatory cells	0.010 (5.90e-06 - 18.58)	0.2	0.015 (0.0001-2.17)	0.1
T gamma delta	11077 (31.8 - 3853624)	0.002	85.94 (5.24-1413.34)	0.002
NK resting	0.0000169 (3.51e-10 – 0.81)	0.04	0.013 (0.0001-1.18)	0.06
NK activated	105.21 (1.6-6931.8)	0.029	41.9 (1.07-1640.6)	0.04
Monocytes	0.97 (9.13e-06 -102358.8)	0.9	0.82 (0.004-157.09)	0.9
Macrophage M0	0.90 (0.57-1.41)	0.7	0.95 (0.64-1.42)	0.8
Macrophage M1	0.95 (0.23-3.86)	0.9	0.86 (0.26-2.79)	0.8
Macrophage M2	1.05 (0.27-4.12)	0.9	1.32 (0.42-4.15)	0.6
Dendritic cells (resting)	6.5 (0.63-67.2)	0.1	5.2 (0.77-34.87)	0.08
Dendritic cells (activated)	0.07 (0.002-2.44)	0.1	0.29 (0.04-2.18)	0.2
Mast cells (resting)	13.45 (0.39-469.8)	0.2	1.83 (0.31-10.78)	0.5
Mast cells (activated)	1.62 (0.44-5.96)	0.5	1.15 (0.53-2.46)	0.7
Eosinophils	41.89 (1.37e-16 - 1.28e+19)	0.9	0.32 (1.05e-08 - 9983100)	0.9
Neutrophils	0.42 (0.02-8.1)	0.5	0.64 (0.17-2.34)	0.5

BH corrected critical p-value for both absolute scores and relative fractions is **0.0022**

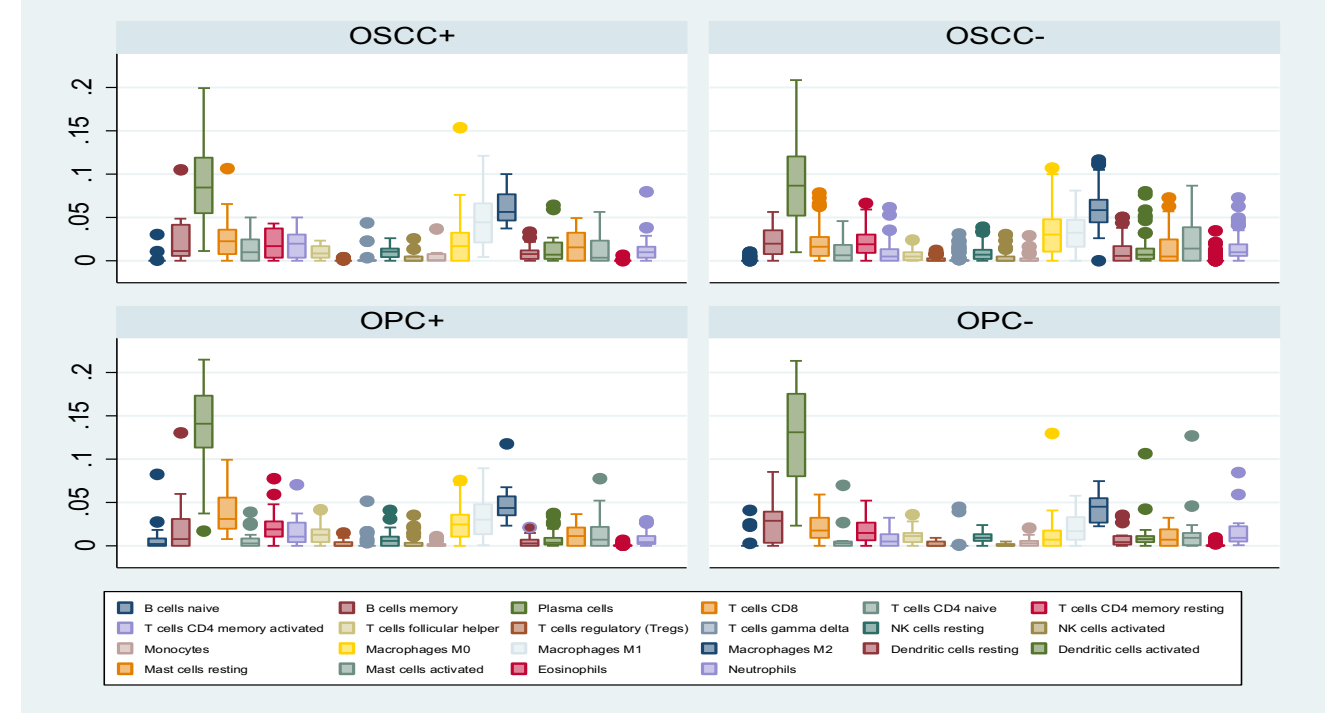


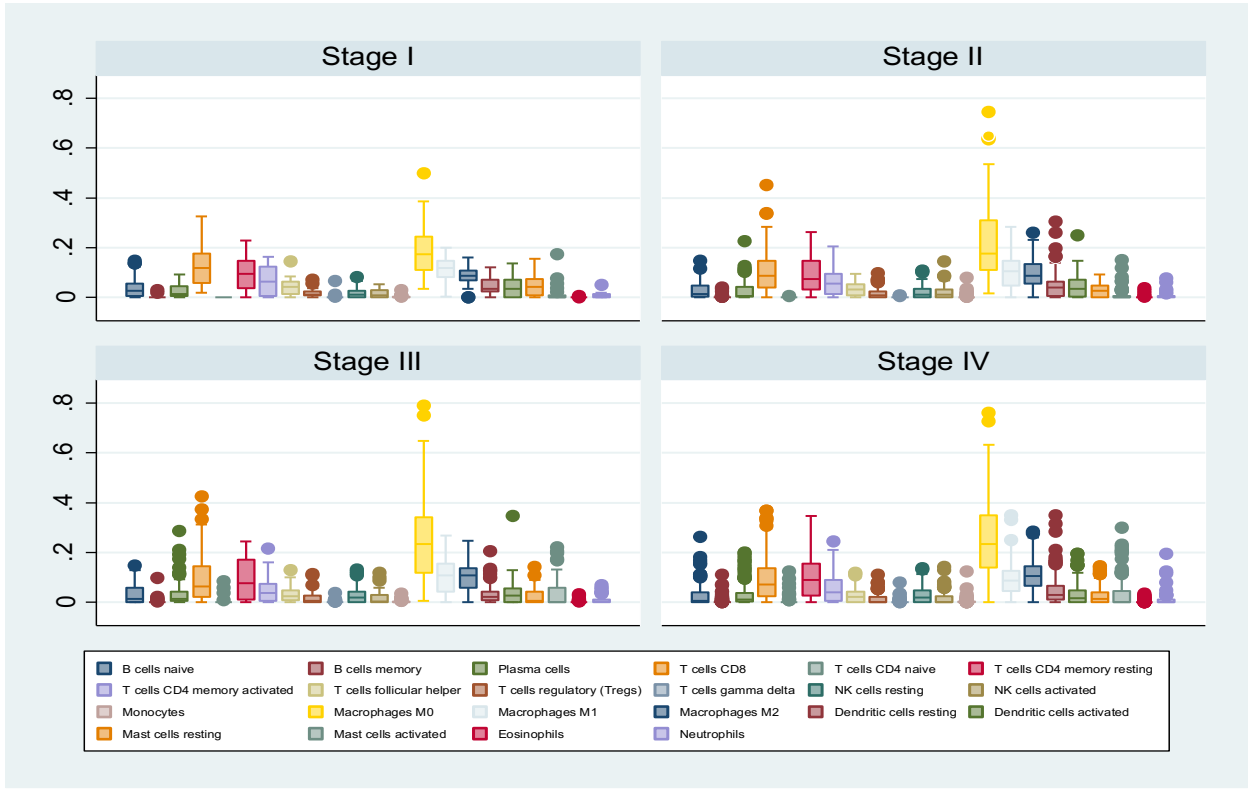
Supplementary Fig 1a (above): Box plot depicting the estimated relative fractions of 22 immune cell types by HPV status and location in **TCGA data**. The Y axis here show the relative proportion which can range from 0 to 1 in theory. Fig 1b (below): Box plot depicting the estimated absolute scores of 22 immune cell types by HPV status and location. The Y axis here show the absolute score which can range from 0 to 100 in theory. It is evident from both the box plots and the bar graph in Fig D that the distribution of immune cell fraction is highly skewed.



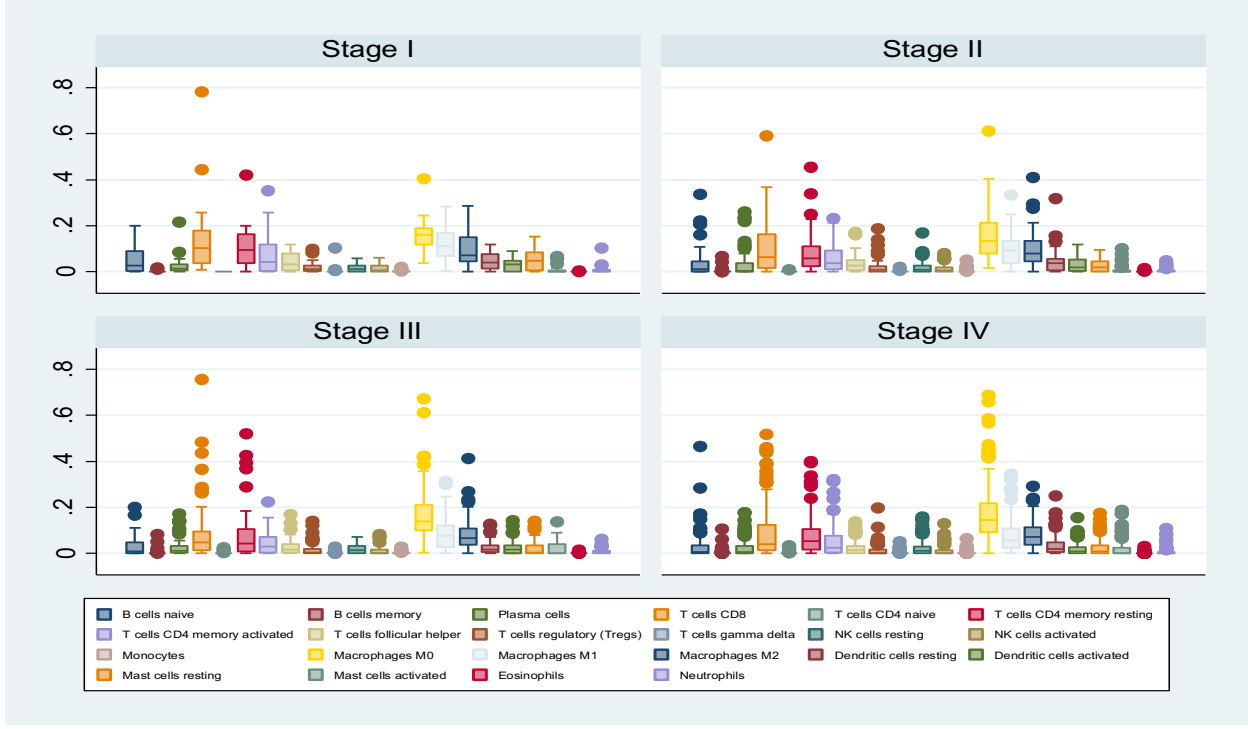


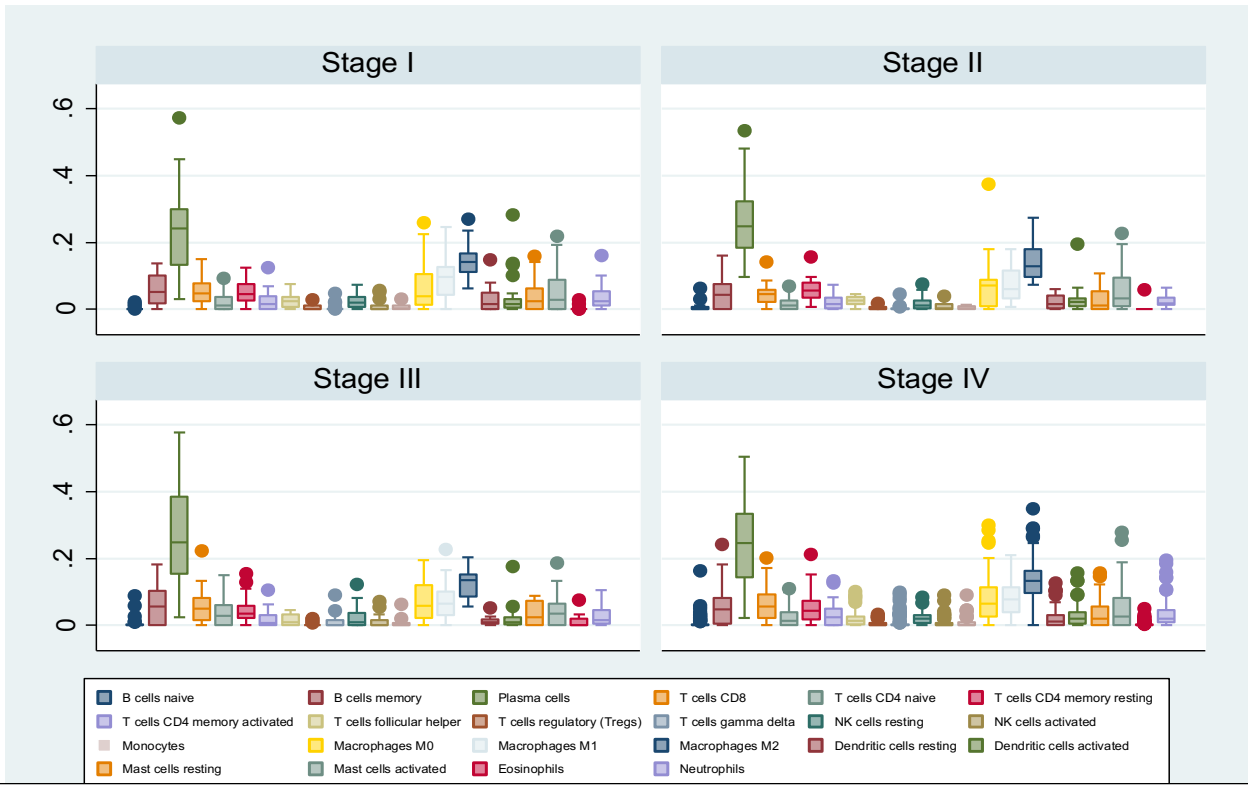
Supplementary Fig 2a (above): Box plot depicting the estimated relative fractions of 22 immune cell types by HPV status and location in **Oralchip data**. The Y axis here show the relative proportion which can range from 0 to 1 in theory. Fig 2b (below): Box plot depicting the estimated absolute scores of 22 immune cell types by HPV status and location. The Y axis here show the absolute score which can range from 0 to 100 in theory.



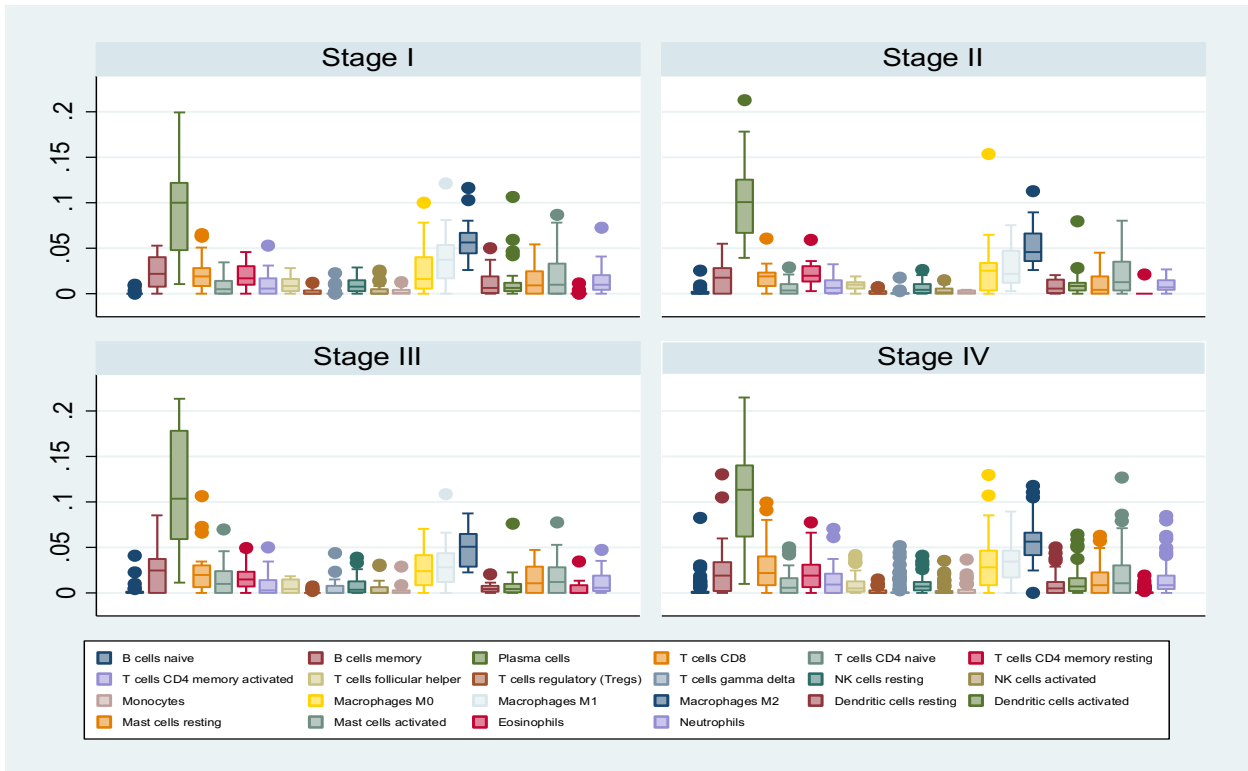


Supplementary Fig 3a (above): Box plot depicting the estimated relative fractions of 22 immune cell types by tumor stage In **TCGA data**. The Y axis here show the relative proportion which can range from 0 to 1 in theory. Fig 3b (below): Box plot depicting the estimated absolute scores of 22 immune cell types by tumor stage. The Y axis here show the absolute score





Supplementary Fig 4a (above): Box plot depicting the estimated relative fractions of 22 immune cell types by tumor stage in **Oralchip data**. The Y axis here show the relative proportion which can range from 0 to 1 in theory. Fig 4b (below): Box plot depicting the estimated absolute scores of 22 immune cell types by tumor stage. The Y axis here show the absolute score



Supplementary Figure 5a (top): Kaplan-Meier survival function graph by HPV status and location of tumor samples from TCGA data and Oralchip data combined (n=546). Supp. Fig. 5(b) (bottom): Kaplan-Meier survival function graph of HPV positive and negative OPC tumor samples from TCGA data and Oralchip data combined (n=546).

