

## Gene expression analysis of TIL rich HPV-driven head and neck tumors reveals a distinct B-cell signature when compared to HPV independent tumors

Oliver Wood<sup>1,\*</sup>, Jeongmin Woo<sup>1,\*</sup>, Gregory Seumois<sup>3,\*</sup>, Natalia Savelyeva<sup>1</sup>, Katy J. McCann<sup>1</sup>, Divya Singh<sup>3</sup>, Terry Jones<sup>2</sup>, Lailah Peel<sup>1</sup>, Michael S. Breen<sup>1</sup>, Matthew Ward<sup>1</sup>, Eva Garrido Martin<sup>1</sup>, Tilman Sanchez-Elsner<sup>1,#</sup>, Gareth Thomas<sup>1,#</sup>, Pandurangan Vijayanand<sup>1,3,#</sup>, Christopher H. Woelk<sup>1,#</sup>, Emma King<sup>1,#</sup>, Christian Ottensmeier<sup>1,#</sup>, for the SPARC Consortium

<sup>1</sup>Faculty of Medicine, University of Southampton & University Hospital Southampton, Southampton, UK

<sup>2</sup>Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, UK

<sup>3</sup>La Jolla Institute for Allergy & Immunology, La Jolla, CA, USA

\*These first authors have contributed equally to this work

#These senior authors have contributed equally to this work

Correspondence to: Emma King, email: e.king@soton.ac.uk

Christopher H. Woelk, email: C.H.Woelk@soton.ac.uk

**Keywords:** head and neck squamous cell carcinoma, human papilloma virus, tumor-infiltrating lymphocyte, RNA-sequencing, transcriptome

Received: February 15, 2016

Accepted: June 30, 2016

Published: July 22, 2016

### ABSTRACT

Human papilloma virus (HPV)-associated head and neck squamous cell carcinoma (HNSCC) has a better prognosis than its HPV negative (HPV(-)) counterpart. This may be due to the higher numbers of tumor-infiltrating lymphocytes (TILs) in HPV positive (HPV(+)) tumors. RNA-Sequencing (RNA-Seq) was used to evaluate whether the differences in clinical behaviour simply reflect a numerical difference in TILs or whether there is a fundamental behavioural difference between TILs in these two settings. Thirty-nine HNSCC tumors were scored for TIL density by immunohistochemistry. After the removal of 16 TIL<sub>low</sub> tumors, RNA-Seq analysis was performed on 23 TIL<sub>high/med</sub> tumors (HPV(+) n=10 and HPV(-) n=13). Using EdgeR, differentially expressed genes (DEG) were identified. Immune subset analysis was performed using Functional Analysis of Individual RNA-Seq/ Microarray Expression (FAIME) and immune gene RNA transcript count analysis. In total, 1,634 DEGs were identified, with a dominant immune signature observed in HPV(+) tumors. After normalizing the expression profiles to account for differences in B- and T-cell number, 437 significantly DEGs remained. A B-cell associated signature distinguished HPV(+) from HPV(-) tumors, and included the DEGs *CD200*, *GGA2*, *ADAM28*, *STAG3*, *SPIB*, *VCAM1*, *BCL2* and *ICOSLG*; the immune signal relative to T-cells was qualitatively similar between TILs of both tumor cohorts. Our findings were validated and confirmed in two independent cohorts using TCGA data and tumor-infiltrating B-cells from additional HPV(+) HNSCC patients. A B-cell associated signal segregated tumors relative to HPV status. Our data suggests that the role of B-cells in the adaptive immune response to HPV(+) HNSCC requires re-assessment.

### INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) accounts for 6% of all cancers, with ~650,000 new cases reported and 350,000 HNSCC-related deaths per year

worldwide [1, 2]. Historically, the risk factors for HNSCC have been smoking and alcohol [3]. However, changes in social behaviour have led to an increase in human papilloma virus (HPV)-associated HNSCC [4]. The incidence of HPV-associated HNSCC is approximately

30% [5–7], of which the majority are caused by HPV16 within the anatomical location oropharynx, which includes the base of tongue and tonsil [6]. HPV positive (HPV(+)) patients have a significantly better prognosis than HPV negative (HPV(-)) patients, with the 3- and 5-year survival at 84% and 62% for HPV(+) patients compared to 57% and 26% for HPV(-) patients, respectively [8].

A high number of tumor-infiltrating lymphocytes (TILs) is linked to good prognosis in many solid tumors, including HPV(+) HNSCC [9]. More recent analyses of The Cancer Genome Atlas (TCGA) data demonstrate that the effect is mediated by CD8<sup>+</sup> GZMA<sup>+</sup> PRF1<sup>+</sup> T-cells [10, 11]. In HPV(+) disease, the persistent viral oncoproteins E6 and E7 cause the malignant phenotype, while the immunological visibility of these two proteins contributes to the infiltration of the tumor by T-cells [12]. By contrast, HPV(-) tumors are considered a separate disease entity and are driven by heterogeneous genetic events [11, 13]. Differential gene expression profiling comparing HPV(+) and HPV(-) tumors using microarray, RNA-Sequencing (RNA-Seq) and RT-PCR have led to an improved understanding of the events associated with cellular transformation and oncogenesis [14–17]. Thurlow *et al.* have used spectral clustering and gene ontology (GO) analysis to identify discrete gene expression patterns that linked to patient outcome, which involved the genes of adaptive and innate immunity [17]. However, the underlying biology of TILs has not been addressed despite their clear link to survival [7, 11, 14–16, 18, 19].

Our aim was to evaluate whether transcriptome analysis would identify characteristics that could differentiate TILs in HPV(+) from those in HPV(-) tumors. We established and optimized sample collection from a cohort of consecutive patients under our care that were undergoing surgery for HNSCC. Tumor samples were processed according to controlled standard operating procedures and evaluated both morphologically and by immunohistochemistry (IHC) on the one hand, and by RNA-Seq to determine whole tumor transcriptomes on the other. We demonstrated that after correction for the number of immune cells infiltrating the tumors, the T-cell signature between HPV(+) and HPV(-) tumors was very similar. By contrast, B-cell associated genes emerged as differentially expressed. High expression of these genes proved to be a distinguishing immunological feature of HPV(+) HNSCC, suggesting a fundamental biological difference in adaptive immune responses against virally driven versus virus-independent tumors. We were able to verify our findings in a large publicly available dataset from TCGA (HNSC) and also by RT-qPCR of the key DEGs.

## RESULTS

### Prognostic effect of TIL density

We had previously demonstrated that in HPV(+) HNSCC TIL density correlates with outcome [7]. From

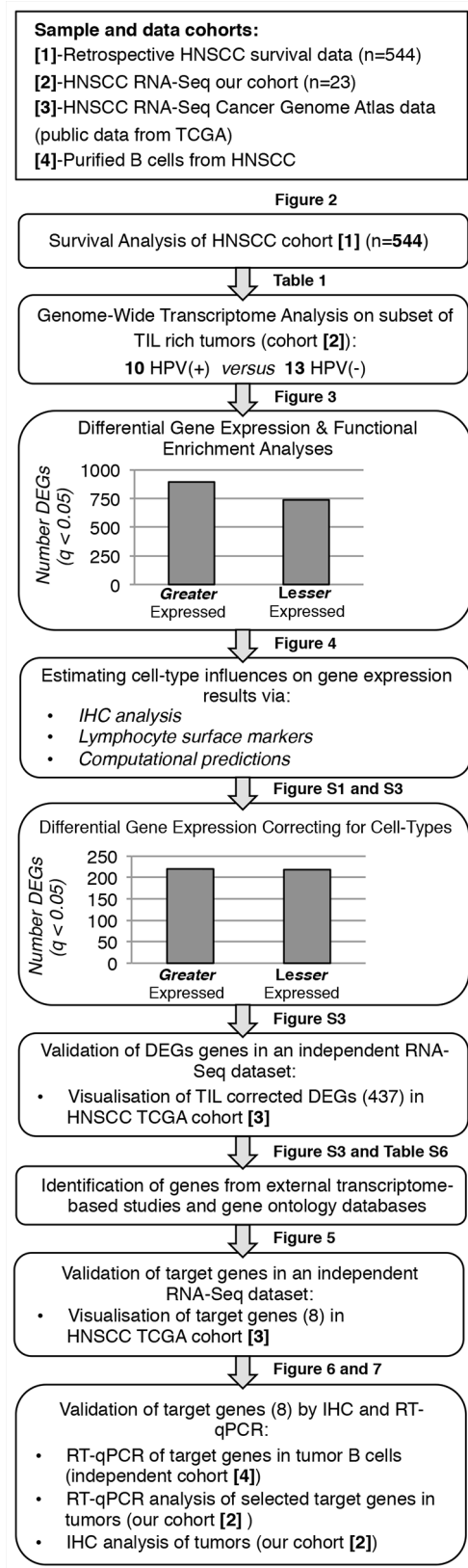
this starting point we undertook a multi-step analysis to understand the features of tumor-infiltrating immune cells in patients with both HPV(+) and HPV(-) HNSCC (Figure 1).

Using the same methodology as previously [7], we extended our cohort (n=544) to include HPV(-) HNSCC. As in HPV(+) tumors, TIL status stratified for outcome in HPV(-) HNSCC (Figure 2). Furthermore, HPV(+) TIL<sub>high/mod</sub> patients had significantly better survival compared to HPV(-) TIL<sub>high/mod</sub> patients (Figure 2, log rank p<0.001). In order to better understand this survival difference, we examined the transcriptome associated with TILs in a prospective cohort (n=39), patient demographics, tumor characteristics and tumor sampling/ processing information for the HPV(+) and HPV(-) patient cohorts are shown in Table 1. As our interest was in understanding immune cells, we focussed on TIL rich (TIL<sub>high/mod</sub>) tumors and excluded the 16 TIL<sub>low</sub> cases from our analyses. Of the 23 TIL<sub>high/mod</sub> cases, 10 were HPV(+) and 13 were HPV(-). The clinical and histological descriptors were distributed as expected (Table 1); HPV(+) tumors were located in the oropharynx of non-smoking patients, HPV(-) tumors were located in the larynx (n=4), oral cavity (n=5) and oropharynx (n=4). The clinical classification of HPV status was determined by routine IHC against p16 and mapped appropriately to the expression of E6 and E7, as detected by RNA-Seq (Table 1).

### Gene expression analysis of HPV(+) and HPV(-) tumors

Following RNA-Seq analysis, 1,634 genes were identified as significantly differentially expressed (*q*-value <0.05) between the HPV(+) and HPV(-) tumors (Figure 3A and Supplementary Table S1). Of these genes, 894 were expressed to a greater extent and 740 to a lesser extent in HPV(+) compared to HPV(-) tumors. These gene expression differences segregated HPV(+) and HPV(-) tumors in all except one HPV(-) subject (patient 21), whose tumor clustered within the HPV(+) cohort (Figure 3A). On histological review, this patient had a basaloid SCC, which is a rare and clinically distinct HNSCC subform. We therefore removed this case from further evaluation.

Gene expression of the HPV associated genes *CDKN2A* (*p16*), *E6* and *E7* were as expected between HPV(+) and HPV(-) tumors (Figure 3B). We also found differences in the expression of genes associated with ‘immune cell markers’ between HPV(+) and HPV(-) tumors (Figure 3B); expression of these genes was greater in HPV(+) tumors. Similarly, the expression of *GZMA*, *IFNG* and *CDNK2A* and key genes that link to T-cell activation and exhaustion, such as *CTLA4*, *PD1* and *HAVCR2* (encoding TIM3), were all increased in HPV(+) compared to HPV(-) tumors.

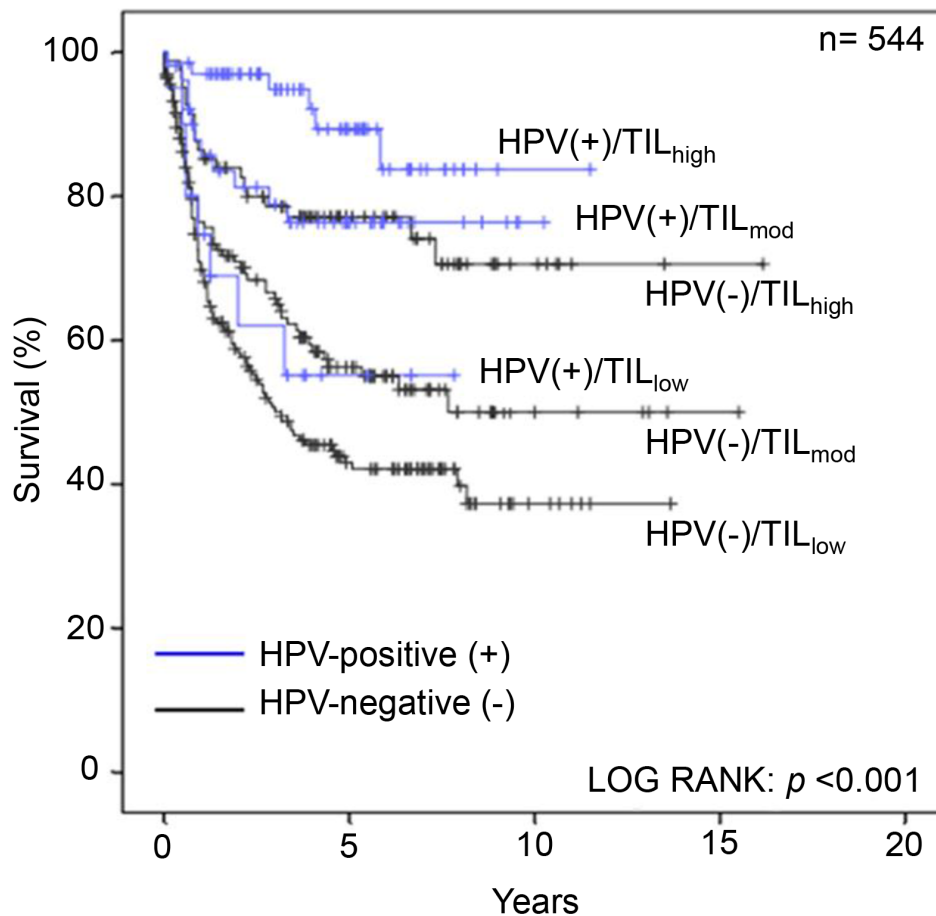


**Figure 1: Overview of experimental procedures and analysis used to evaluate TIL.** Schematic representation of the multi-step analysis performed to understand the features of tumor-infiltrating immune cells in patients with HPV(+) and HPV(-) HNSCC.

### GO and pathway analysis

GO and pathway analysis was performed to understand the biological significance of the 1,634 DEGs. GO terms that were significantly over-represented for DEGs were identified using CPDB [20]; GO analysis was performed independently for those genes expressed to a

greater extent and for genes expressed to a lesser extent in HPV(+) compared to HPV(-) tumors. Full details of the GO terms, including the specific genes, number and percentage of DEGs associated with each GO term, are presented in Supplementary Tables S2 and Supplementary S3 respectively.



HPV(+)/TIL <sub>high</sub>	61	26	1	
HPV(+)/TIL <sub>mod</sub>	51	21	1	
HPV(+)/TIL <sub>low</sub>	20	4		
HPV(-)/TIL <sub>high</sub>	81	33	7	1
HPV(-)/TIL <sub>mod</sub>	134	47	6	1
HPV(-)/TIL <sub>low</sub>	192	49	6	

**Figure 2: Kaplan-Meier curves for HNSCC mortality stratified according to HPV status and TIL density.** Survival of a retrospective cohort of HNSCC patients (n=544) with respect to HPV status and the density of immune cell infiltrate was assessed. TIL density predicts for outcome in both the HPV(+) and HPV(-) patients; log-rank test,  $p < 0.001$ .

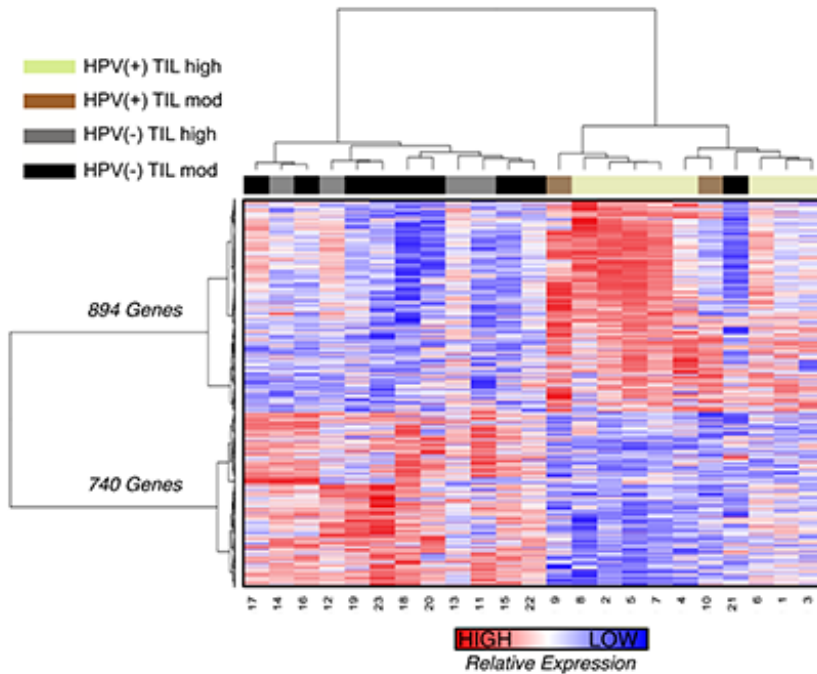
**Table 1: Clinical and histopathological information for HPV(+) and HPV(-) HNSCC patients**

Variable	Description	HPV(+)	HPV(-)	p-value	Test <sup>#</sup>
		(n=10)	(n=13)		
Gender	% Male	90.00	76.92	6.04E-01	<i>a</i>
Age	Mean	56.90 ± 8.66	63.77 ± 15.85	2.32E-01	<i>b</i>
Collection site	Liverpool	1	0	2.81E-01	<i>a</i>
	Poole	2	6		
	Southampton	7	7		
Sequencing Batch	1	3	6	3.92E-01	<i>a</i>
	2	7	5		
	3	0	2		
RIN*	Mean	8.51 ± 0.80	8.51 ± 1.01	7.56E-01	<i>c</i>
Smoking	Non smoker	10	6	2.56E-02	<i>a</i>
	Smoker	0	4		
	Heavy smoker	0	3		
Alcohol	Non drinker	5	3	4.02E-01	<i>a</i>
	Moderate drinker	4	9		
	Heavy drinker	1	1		
Tumor site	Larynx	0	4	1.60E-03	<i>a</i>
	Oral	0	5		
	Oropharynx	10	4		
White blood cells	Mean	9.47 ± 2.23	8.52 ± 1.92	2.86E-01	<i>b</i>
Lymphocytes	Mean	1.82 ± 0.60	1.77 ± 0.81	5.76E-01	<i>c</i>
Staging	I	0	3	3.48E-01	<i>a</i>
	II	1	0		
	III	2	3		
	IV	7	7		
TIL status	High	8	4	3.61E-02	<i>a</i>
	Medium	2	9		
Pattern of invasion	Cohesive	10	6	7.49E-03	<i>a</i>
	Discohesive	0	7		
Differentiation	Poor	9	2	8.85E-04	<i>a</i>
	Moderate	1	9		
	Well	0	2		
Smooth muscle actin	Low	10	9	2.37E-01	<i>a</i>
	Moderate	0	2		
	High	0	2		
Tumor cell	%	68.00 ± 20.58	65.38 ± 21.06	7.69E-01	<i>B</i>
E6 expression	Mean	4.57 ± 1.11	-4.36 ± 0.57	1.75E-06	<i>C</i>
E7 expression	Mean	5.07 ± 1.30	-4.36 ± 0.57	1.75E-06	<i>C</i>

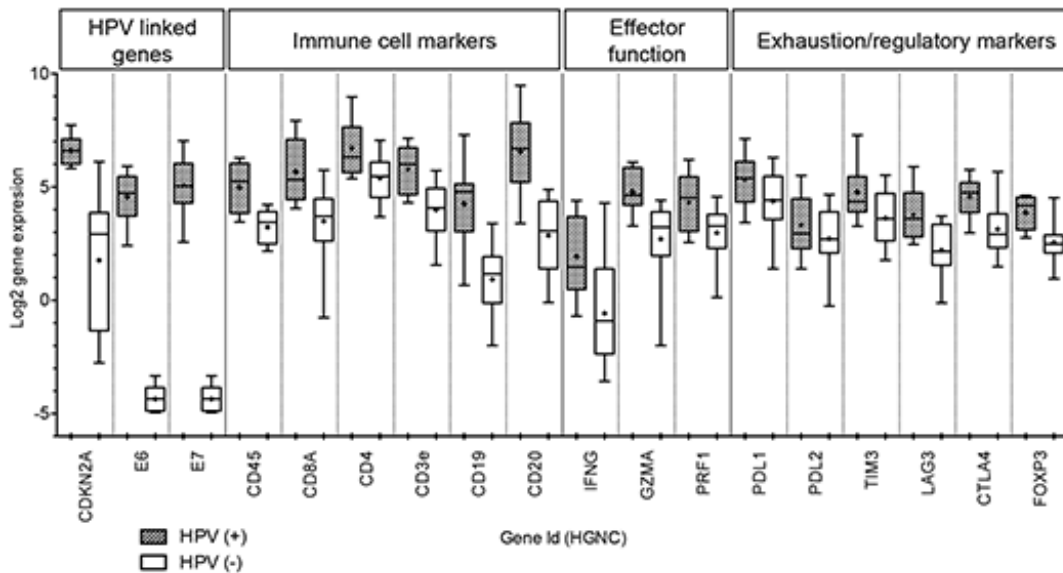
\*RIN; RNA integrity number

<sup>#</sup>P-values were obtained from Fisher's exact test (*a*) for categorical variables. A two sample t-test (*b*) was performed for numerical variables with normal distribution (Shapiro-Wilk test,  $P \geq 0.05$ ). A Wilcoxon rank sum test (*c*) was performed for numerical variables with non-normal distribution (Shapiro-Wilk test,  $P < 0.05$ ). Statistically significant *p*-values ( $< 0.05$ ) are indicated in bold.

A



B



**Figure 3: Differentially expressed genes between HPV(+) and HPV(-) tumors.** **A.** a heatmap to illustrate the DEGs between HPV(+) and HPV(-) tumors; each row represents the z-score of normalized gene expression values for a given gene; each column represents the gene expression for a given tumor. Z-scores are calculated from the average gene expression, plus standard deviation, in all tumors for a given gene, this enables the relative comparison of gene expression between tumors; the scale of z-score is shown: red shading denotes greater gene expression, blue shading denotes lower gene expression. Hierarchical clustering of genes and tumors based on their expression profile is reflected in the dendrograms to the left and the top of the heatmap, respectively, and was performed by calculating distance using the Pearson's correlation metric and then clustering distance using the ward linkage method\*. **B.** the expression of key genes associated with HPV, immune cell markers, immune effector function and immune exhaustion/regulation are displayed for HPV(+) and HPV(-) tumors as box plots (min/max) with the + representing the mean. A greater expression of immune associated genes is observed in HPV(+) tumors\*\*. \*Unsupervised clustering of gene expression data was normalized using the TMM method followed by variance stabilizing transformation of the TMM normalized data. \*\*Gene expression data from the normalized transcript counts; data was normalized using the TMM method followed by variance stabilizing transformation of the TMM normalized data.

Our data revealed that those genes with greater expression in the HPV(+) cohort were predominantly associated with an immune reaction (e.g., adaptive immune response, *GO:0002250*; lymphocyte activation, *GO:0046649*; positive regulation of immune system process, *GO:0002684*; B-cell activation, *GO:0042113*), whereas those expressed to a lesser extent were associated with cellular processes involved in tissue development, (*GO:0009888*), keratinization (*GO:0031424*) and cell differentiation (*GO:0030154*). Specifically, there was greater expression of genes associated with the adaptive immune system, including T-cells (CD4<sup>+</sup> and CD8<sup>+</sup>) and B-cell receptor signalling pathways, as well as NK-cell-mediated cytotoxicity (Supplementary Table S2).

Those genes expressed to a lesser extent represented different biological processes, including extracellular matrix organisation, collagen formation, beta1 integrin cell surface interactions and alpha6 beta4 integrin-ligand interactions (Supplementary Table S3). The biological pathways (e.g., *KEGG*) over-represented for DEGs mirrored the results of GO analysis displaying an enriched number of genes linked to immunological signalling pathways in greater expressed genes (Supplementary Tables S4) and pathways linked to extracellular matrix organisation and collagen formation in genes expressed to a lesser extent (Supplementary Tables S5).

### Quantification of TILs

GO and pathway analysis indicated that the biological processes related to the immune system and specifically to B and T-cells were over-represented for genes expressed to a greater extent in HPV(+) tumors (Figure 3B and Supplementary Tables S2 and S4). However, it was not clear whether these B- and T-cell immune-related terms identified by GO and pathway analysis were simply the result of numerical differences in lymphocytes between HPV(+) and HPV(-) tumors or the result of differences on a per cell basis in the transcriptional signature.

To be able to address this question we compared three approaches to quantifying tumor-infiltrating lymphocytes: IHC for CD4, CD8, CD20 and CD3 followed by manual counting of 10 high-power fields (Figure 4A), gene RNA transcript levels (log<sub>2</sub> normalised) for lymphocyte cell surface markers (Figure 4B) and computational evaluation using FAIME to measure the size of lymphocyte subsets in each sample (Figure 4C). A significant difference in cell number was observed between HPV(+) and HPV(-) tumors for the cell markers CD4, CD20 and CD3 ( $P \leq 0.05$ ) but not CD8 (ns) by IHC. Gene RNA transcript levels were significantly different for all cell markers (*CD4*, *CD20*, *CD3e* and *CD8A*) (Figure 4B). Spearman correlation analysis between IHC and gene expression for the cell markers CD8 ( $R^2=0.76$ ), CD3 ( $R^2=0.52$ ) and CD19 ( $R^2=0.47$ ) is shown in Supplementary Figure S1A, B and C respectively. FAIME analysis uses

gene expression biomarkers (Supplementary Table S6) to score the size of each lymphocyte subset; this score was significantly different between the HPV(+) and HPV(-) tumors for B-cells ( $q$ -value =  $1.35E-03$ ), CD4<sup>+</sup> T-cells ( $q$ -value =  $2.66E-02$ ) and CD8<sup>+</sup> T-cells ( $q$ -value =  $9.64E-03$ ). FAIME scores indicated that the number of cells belonging to these lymphocyte subsets was higher in HPV(+) tumors compared to HPV(-) tumors.

Both molecular analyses demonstrated that the numbers of B-cells and CD4<sup>+</sup> and CD8<sup>+</sup> T-cells were higher in HPV(+) tumors compared to HPV(-) tumors, this was confirmed by the 'gold standard' assessment by IHC for B-cells, CD3<sup>+</sup> T-cells and CD4<sup>+</sup> T-cells ( $P \leq 0.05$ ). A significant difference was not observed by IHC for CD8<sup>+</sup> T-cells, however a trend towards higher numbers in HPV(+) was observed (Figure 4A).

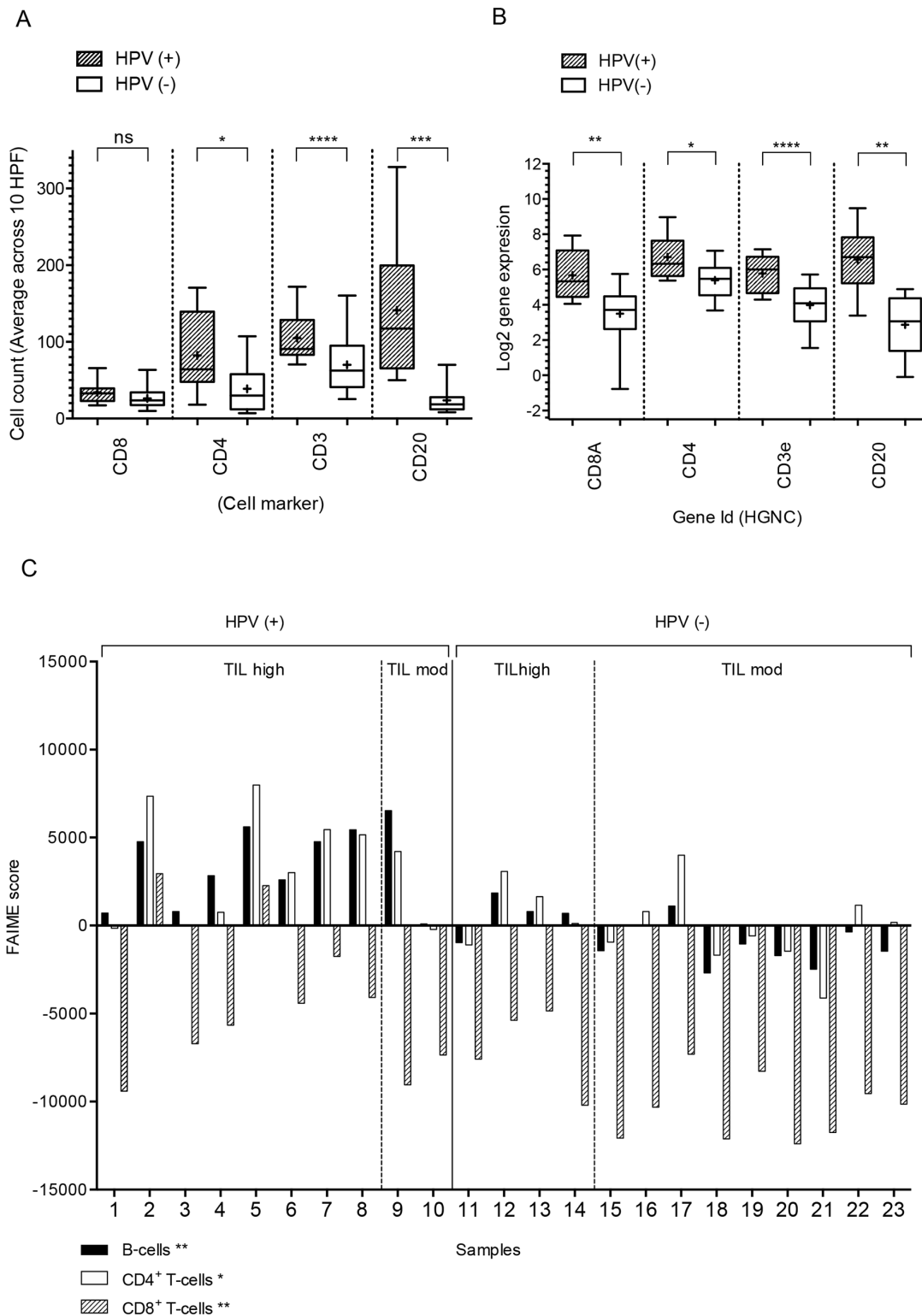
### Analysis of gene expression data following correction for numerical differences in TILs

The global gene expression data were next corrected for TIL number using the gene expression of *CD19* (pan B-cell marker, CD20 was used only for IHC comparison), *CD4* and *CD8A* in each sample as a covariate. When correcting both HPV(+) and HPV(-) cohorts in this way, genes co-ordinately expressed in lymphocyte subsets were no longer differentially expressed; as a result the number of DEGs dropped from 1,634 to 437 (Supplementary Table S7). As expected, there was a large overlap in DEGs between the initial uncorrected and the TIL corrected data sets (Supplementary Figure S2). The TIL corrected dataset was next subject to GO and pathway analysis.

### GO and pathway analysis of TIL corrected data

GO and pathway analysis was again performed independently for genes expressed to a greater extent ( $n=219$ ; Supplementary Table S8) and a lesser extent ( $n=218$ ; Supplementary Table S9) in HPV(+) compared to HPV(-) tumors. Consistent with the FAIME analysis, the vast majority of immune and lymphocyte-related terms were no longer over-represented in HPV(+) tumors. This included markers of T-cell effector function (e.g., *IFNG*, *GZMB* and *PRF1*; Figure 3B), which prior to TIL correction were all over-represented in the HPV(+) tumors. Pathway analysis also confirmed the loss of immune-related signalling pathways (Supplementary Table S10).

Immune GO terms that remained following correction for numerical differences in TIL were B-cell activation (*GO:0042113*), which included *BCL2*, *VCAM1* and *ICOSLG*, with a greater expression in HPV(+) compared to HPV(-) tumors (Supplementary Table S8). The surviving non-immune GO terms and biological pathways (Supplementary Table S8 and S10) over-represented in HPV(+) tumors were associated with cell cycle (*GO:0007049*), cell phase transition (*GO:0044770*) and chromosome organisation (*GO:0051276*).



**Figure 4: Immune cell subset analysis of HPV(+) and HPV(-) tumors.** **A.** The distribution of CD4, CD8, CD3 and CD20-expressing cells in HPV(+) and HPV(-) tumors as detected by IHC; cell counts are given as a mean of 10 high-power fields. **B.** Gene expression of *CD4*, *CD8A*, *CD3E* and *CD20* of HPV(+) and HPV(-) tumors displayed as box plots (min/max) with the + representing the mean\*. Differences in TIL density between HPV(+) and HPV(-) tumors are observed both by gene expression profiling and IHC analysis. **C.** the FAIME score of lymphocytes in HPV(+) and HPV(-) tumors; the difference in distribution of specific cell subsets based on ranked gene expression is shown. Asterisks in column labels indicate a significance level of two-sample *t*-test comparisons of FAIME scores between HPV(+) and HPV(-) tumors: \**P* < 0.05 and \*\**P* < 0.01). \*Gene expression data from the normalized transcript counts; data was normalized using the TMM method followed by variance stabilizing transformation of the TMM normalized data.



The loss of T-cell and the majority of B-cell-related GO terms following TIL correction of gene expression data indicated that gene expression differences between HPV(+) and HPV(-) tumors largely resulted from numerical differences in these cells types. To determine if any differences in lymphocyte gene expression between the HPV(+) and HPV(-) tumor cohorts were retained following TIL correction, the DEGs identified after correction were overlapped with the lymphocyte-specific marker genes used to determine the cell proportions in the FAIME analysis (CD19<sup>+</sup> B-cell genes and CD8<sup>+</sup> and CD4<sup>+</sup> T-cell genes, Supplementary Table S6 [21–26]). The majority of the surviving signals were associated with the following B-cell associated genes: *GGA2*, *SPIB*, *CD200*, *ADAM28* as well as *STAG3*, which was not previously known to be a B-cell gene. A single CD8-associated gene (*CD8B*) also survived correction (Supplementary Figure S3).

### Validation of findings using TCGA RNA-Seq data

We next assessed the expression of the identified DEGs (n=437) including the 8 B-cell-related genes *GGA2*, *SPIB*, *CD200*, *ADAM28*, *BCL2*, *VCAMI*, *ICOSLG* and *STAG3*, in an independent HNSCC dataset from TCGA (TCGA HNSC HiSeqV2 2015-02-24 data source outlined in methods) [11].

Analysis of TCGA data allowed us to address whether the DEGs might be the result of anatomical bias: since HPV(+) tumors predominantly arise in specific anatomical locations (tonsil, base of tongue and oropharynx) these might per se contribute to DEGs. Hence we identified 72 cases (46 HPV(+) and 26 HPV(-)) from anatomically matched locations (tonsil, base of tongue and oropharynx). Visualisation of the 8-gene signature (*GGA2*, *SPIB*, *CD200*, *ADAM28*, *BCL2*, *VCAMI*, *ICOSLG* and *STAG3*) identified from our cases in a hierarchically clustered heatmap (Figure 5A) together with the 72 TCGA cases (Figure 5B) demonstrated the clustering of tumors according to HPV status, with a greater gene expression in HPV(+) compared to HPV(-) tumors. These data confirmed that anatomical bias was not the reason for the B-cell-associated differences in gene expression. Unsupervised hierarchical clustering of all 437 TIL corrected DEGs for our own dataset and TCGA data is shown in Supplementary Figure S4.

### Validation of RNA-seq data by RT-qPCR and IHC

We confirmed our findings with RT-qPCR, identifying the expression of the genes *GGA2*, *SPIB*, *CD200*, *STAG3*, *ADAM28*, *BCL2*, *VCAMI* and *ICOSLG* in B-cells isolated from an independent HPV(+) HNSCC tumor cohort (n=6) (Figure 6). In addition to this, RT-qPCR

of *CD200* and *STAG3* was carried out on the whole tumor RNA samples used for the RNA-Seq (n=8 HPV(+) and n=8 HPV(-)). This showed the same trend with HPV(+) tumors compared to HPV(-) tumors having increased expression of *STAG3* and *CD200* (Supplementary Figure S5; *STAG3*, \*\*\*p<0.001 and *CD200* nsd, p=0.116). We have exhausted the material meaning additional genes and cases could not be assessed, limiting the statistical power for *CD200*. *STAG3*, a component of the meiosis specific cohesin complex [27] was expressed at a low level in B-cells, RT-qPCR of whole tumor tissue confirmed its differential expression between HPV(+) and HPV(-) tumors (Supplementary Figure S5).

IHC assessment of HPV(+) and HPV(-) tumors identified dense clusters of tumor-infiltrating B-cells in the former; Figure 7 shows representative histology for one HPV(+) and one HPV(-) tumor. The follicular morphology within HPV(+) tumors was apparent following staining for CD23, a marker of follicular B-cells. Furthermore, IHC confirmed the presence of CD200<sup>+</sup> cells within and around tertiary lymphoid follicles, together with a diffuse CD8<sup>+</sup> T-cell infiltrate.

### Retained non-immune gene differences between HPV(+) and HPV(-) tumor after TIL correction

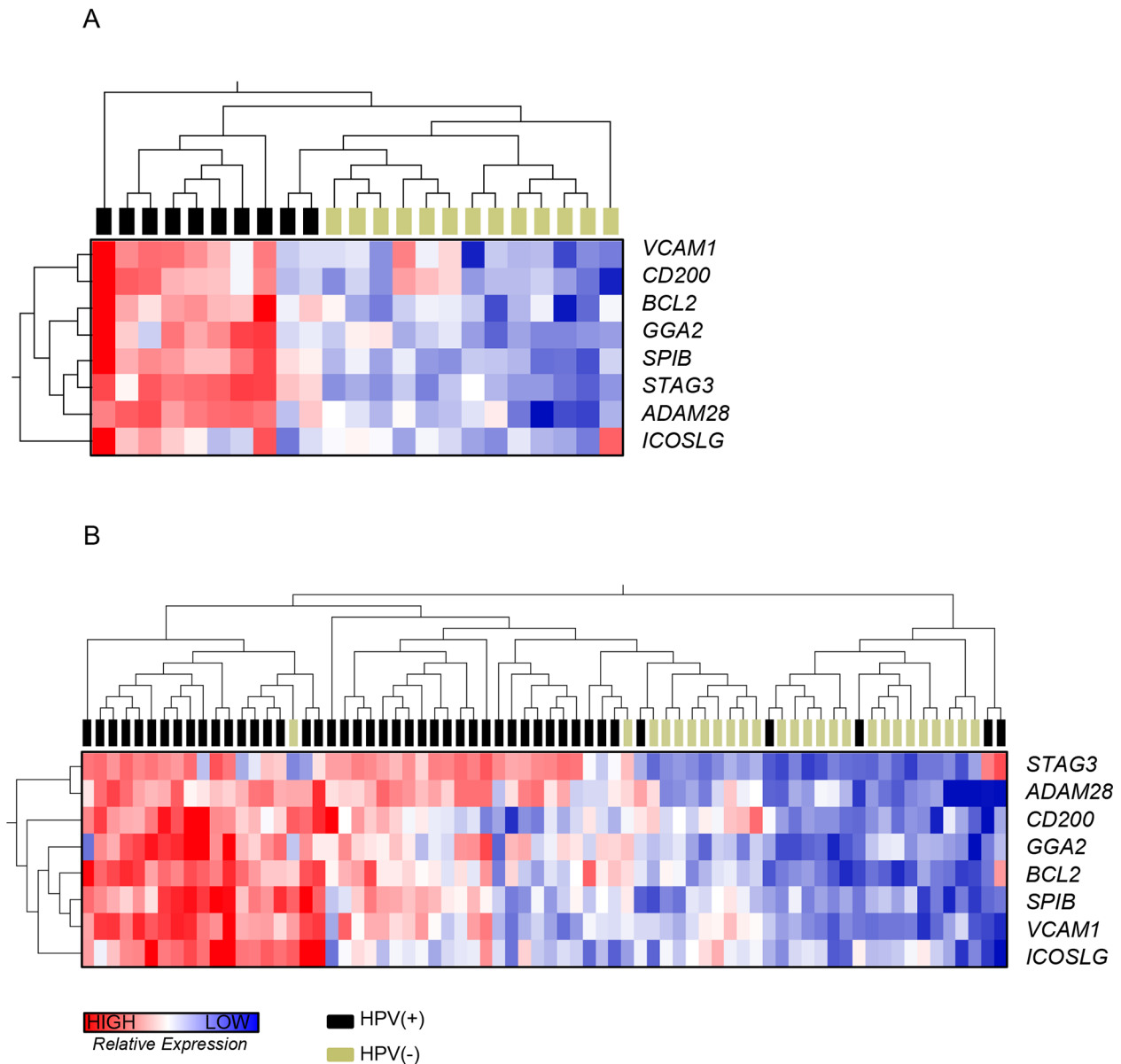
GO and pathway analysis of DEGs expressed to a lesser extent in HPV(+) compared to HPV(-) tumors were largely unchanged following TIL correction of the data. There were 218 lesser expressed genes following lymphocyte correction, of which many were associated with development (skin, epidermis, epithelium, tissue, organ) and keratinization (Supplementary Figure S2 and Supplementary Table S9). In addition, lesser expressed genes associated with IL-12 and IL-6 production, the inflammatory response and the response to oxidative stress. Individual DEG's included keratin's (*KRT-10*, *14*, *16* and *17*), kallikrein-related peptidase 5, 7 and 14 (*KLK5*, *7* and *14*), caspase 14 (*CASP14*), tumor necrosis factor (ligand) superfamily member 9 (*TNFSF9* or *CD137L*), thrombospondin receptor (*CD36*) and chemokine (C-C motif) ligand 19 (*CCL19*); pathway analysis of lesser expressed genes shown in Supplementary Table S11 returned no significantly over-represented pathways (q-value <0.05).

## DISCUSSION

We have previously shown that HPV(+) HNSCC patients with a dense immune cell infiltrate within the tumor have a better outcome than those with a sparse infiltrate [7] and show here that this is also the case for HPV(-) tumors. We wished to evaluate whether fine-resolution transcriptomic analysis could give insight into the biological difference in the immune infiltrate between patients with a known viral (HPV)

driver, compared with those patients where virus is absent. Although differential gene expression studies comparing HPV(+) and HPV(-) tumors have been reported [11, 14–17], the question as to whether the pathogenesis of the cancer is reflected in differences in immune cells themselves remains open.

In this study we focused on TIL<sub>high/mod</sub> cases from our consecutive cohort and following RNA-Seq analysis we examined DEGs. Initially, the most striking difference was the immune signature, which was significantly greater in HPV(+) tumors and hence likely reflected an immune responses to viral antigens. In contrast HPV (-) tumors



**Figure 5: Expression of B-cell-associated genes by RNA-Seq.** Heatmaps\* to illustrate gene expression of the identified B-cell-associated genes between HPV(+) and (-) tumors: *GGA2*, *SPIB*, *CD200*, *STAG3*, *ADAM28*, *BCL2*, *VCAM1* and *ICOSLG*. **A.**, a heatmap of our HNSCC dataset (HPV(+) n=10 and HPV(-) n=13). **B.**, a heatmap of the TCGA HNSCC dataset (HPV(+) n=46 and HPV(-) n=26); publicly available data from anatomically matched tumors arising in the oropharynx, tonsil and base of tongue. In both datasets, tumors cluster according to HPV status, with a greater expression of the identified B-cell-associated genes in HPV(+) tumors. \*Unsupervised clustering of gene expression data was normalized using the TMM method followed by variance stabilizing transformation of the TMM normalized data. Each row represents normalized gene expression values for a given gene; each column represents the gene expression for a given tumor: red shading denotes greater gene expression, blue shading denotes lower gene expression. Hierarchical clustering of genes and tumors based on their expression profile is reflected in the dendrograms to the left and the top of the heatmap, respectively, and was performed by calculating distance using the Pearson's correlation metric and then clustering distance using the ward linkage method.

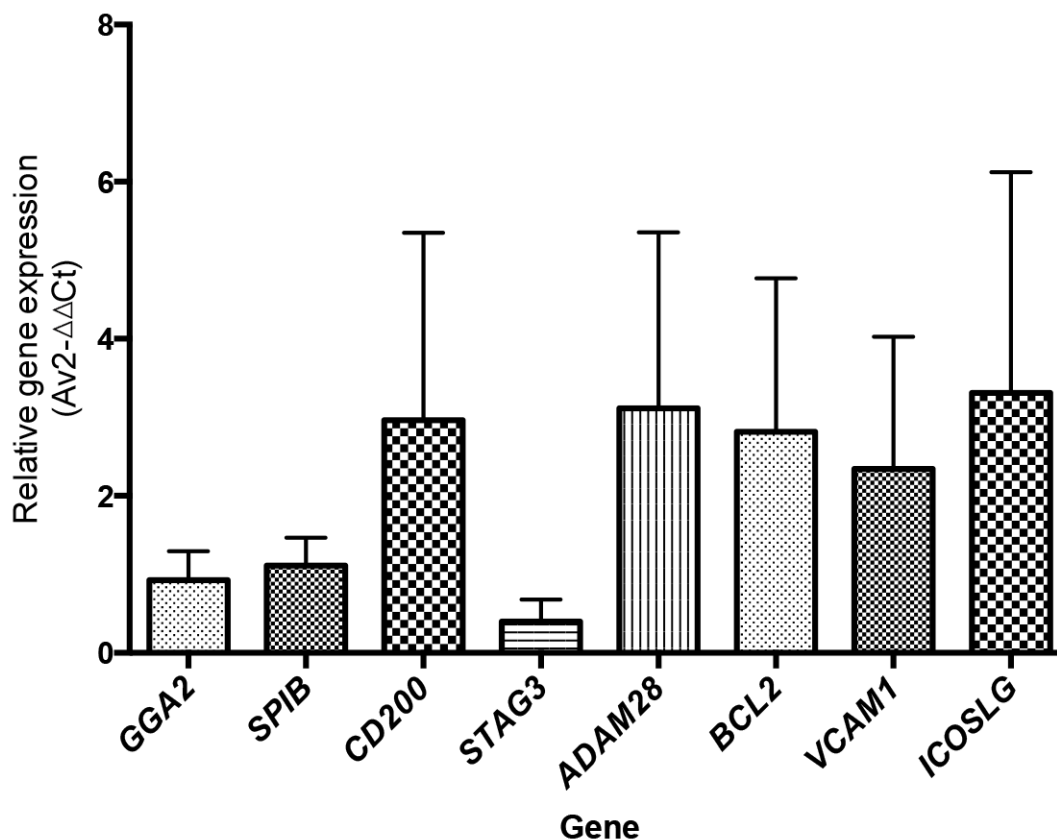
revealed a prominent tissue development/ re-organisation gene signature. The molecular data mapped well onto the TIL characterization afforded by IHC, although the variability was higher when TILs were counted manually. It is likely that this is a reflection of the fact that RNA-Seq analysis uses homogenized tumor, which averages the geographical differences within the tissue.

A significant difference in TIL density determined by RNA-Seq gene transcript levels of *CD4* and *CD8A* (T-cells) and *CD19* (B-cells) remained between HPV(+) and HPV(-) patient cohorts. This was clearly demonstrated in the IHC assessment and by FAIME analysis, where in ranked order, HPV(-) TIL<sub>high/mod</sub> patients had a significantly lower expression of B- and T-cell-related genes compared with HPV(+) TIL<sub>high/mod</sub> patients. In order to assess the difference between HPV(+) and HPV(-) TIL enriched tumors, the data were corrected according to the number of immune cells present in the tissue as determined by RNA-Seq gene transcript levels (*CD8A*, *CD4* and *CD19*), with the aim to account for the numerical difference in immune cells.

RNA-Seq data that had been corrected for TIL number showed that the vast majority of immune-related

DEGs were 'lost', suggesting that in both patient cohorts the lymphocytes were qualitatively similar. The T-cell immune signature, present in both HPV(+) and HPV(-) tumors, no longer showed differentially expressed genes. Thurlow *et al.* have previously demonstrated that both HPV(+) and HPV(-) cohorts can mount adaptive immune responses [19], while the association of cytolytic activity of effector cells and immunoediting of the tumor has also been reported [10]. These studies however did not correct for lymphocyte numbers; the HPV(+) tumors in which no adaptive response was detected likely represented TIL<sub>low</sub> patients. We had expected to find that a viral driver would promote a distinct T-cell-driven TIL signature, but this was not observed in our data. Instead, our data demonstrate that T-cells in TIL<sub>high/mod</sub> tumors of different pathogeneses are transcriptomically similar, at the level of bulk population analysis.

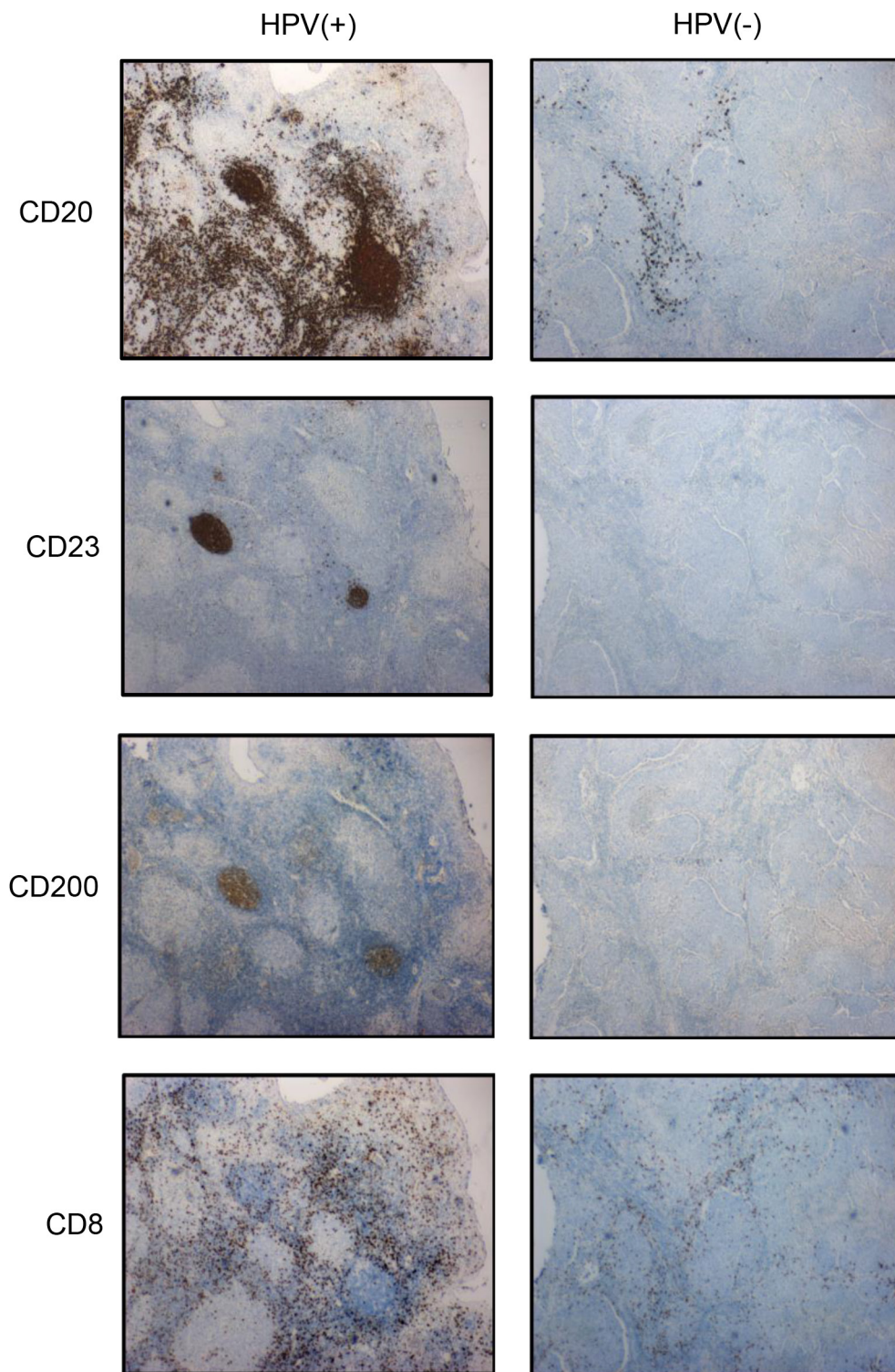
In contrast, there was a distinct B-cell signature between HPV(+) and HPV(-) tumors after correction for TIL numbers, there was a small subset of B-cell associated genes that continued to have greater expression in HPV(+) tumors, including *GGA2*, *SPIB*, *CD200*, *ADAM28*, *BCL2*, *VCAM1* and *ICOSLG*, as determined by



**Figure 6: Relative expression of B-cell-associated genes by RT-qPCR.** The average relative gene expression of B-cell-associated genes was measured by RT-qPCR\* in HPV(+) tumors. The expression of the B-cell-associated genes *GGA2*, *SPIB*, *CD200*, *STAG3*, *ADAM28*, *BCL2*, *VCAM1* and *ICOSLG* was confirmed in B-cells sorted from an independent cohort of HPV(+) tumors (n=6). \*Relative gene expression by RT-qPCR, calculated using the comparative Ct method with *Actin* as the control gene (2-ΔΔCt method) (23).

previously published datasets for B-cell gene signatures and the GO term (B-cell activation, *GO:0042113*) [21–26]. The expression of the differentially expressed B

cell-associated genes is not unique to B cells as they can be identified in other cell types. This includes CD200, which has been found to be expressed on 1-2% of basal



**Figure 7: Expression of B-cell markers by IHC.** Cell subset analysis by IHC for the B-cell markers CD20, CD23 and CD200, as well as the T-cell marker CD8, was performed on sequential sections for HPV(+) (n=9) and HPV(-) (n=13) tumors; representative data is shown for one tumor from each cohort. Pseudo-follicle formation is apparent in HPV(+) tumors along with dense infiltrate of CD20<sup>+</sup> B-cells. TIL density was greater in HPV(+) tumors.

cell carcinoma cells [28]. However, our data on purified B-cell populations from HPV(+) HNSCC confirms the expression of the *BCL2*, *ADAM28*, *CD200*, *ICOSLG* and *SPIB* genes in B cells isolated from the tumors. We were able to confirm our DEGs in a larger, independent cohort derived from HNSCC TCGA data. We could show that the DEGs are not the result of anatomical location bias of HPV(+) tumors, as the differences are maintained when anatomically matched HPV(-) and HPV(+) only are compared. Additionally, applying the TIL corrected DEG list to the TCGA confirmed the validity of our observations.

Greater expression of *BCL2* by HNSCC has been proposed as a predictor of good response to chemotherapy; this is consistent with its expression in our HPV(+) cohort, a group of patients that generally respond well to treatment, including chemotherapy [29, 30]. During a normal humoral response, *ICOSLG* is expressed on activated B-cells within germinal centers, which are formed in follicles and are central for an antigen-specific humoral response. Histologically, all of our HPV(+) tumors had very well developed follicles/tertiary lymphoid structure [31]. Follicles within solid human tumors have previously been described in breast, cervical and non-small-cell lung carcinoma [32, 33]. In melanoma, *ICOSLG* is linked with increasing numbers of regulatory T-cells ( $T_{reg}$ ) [34], but has not previously been described in HPV(+) HNSCC. Interestingly, when a stimulatory ICOS antibody was used in combination with anti-CTLA-4, there was a significant improvement in tumor rejection in both melanoma and colon cancer mouse models, suggesting that HPV(+) tumors expressing high levels of *ICOSLG* may respond better to anti-CTLA-4 therapy [35]. *SPIB* is a transcriptional activator that is specific for lymphoid cells and has previously been identified in germinal centers where it was associated with an activated B-cell phenotype [36].

Using IHC to probe protein expression of CD200, we identified CD200<sup>+</sup> B-cells within and outwith of follicular structures. Given that the viral antigens E6 and E7 generate a strong B-cell response in patients with HPV(+) HNSCC [37], it is likely that the expression of the B-cell activation marker CD200 [38] is linked to a persistence of HPV-driven, tumor-derived antigens, which in turn stimulate specific B-cells in the germinal center and the tumor itself. It is also possible that the CD200<sup>+</sup> B-cells resident at specific locations in the tumor vs. the follicles have a distinct phenotype and properties. These cells could be part of an inhibitory pathway: CD200R activation stimulates the differentiation of T-cells to  $T_{reg}$  [39], the numerical increase of  $T_{reg}$  in parallel with TIL number has been reported in HPV(+) HNSCC [40]. These CD200<sup>+</sup> B cells can act through CD200R and effector mechanisms may include immunomodulatory cytokines, such as IL-10, [41] release of granzyme B [42] or other yet to be identified modes of action. The receptor is expressed

on both lymphoid and myeloid cells, hence direct and indirect inhibition of T-cell immunity could be possible [43]. It must however be noted that expression of CD200 is thought to have links to anti-tumor effects by inhibiting activity of tumor-associated myeloid cells via IL-10 [44], arguing caution in targeting a molecule with potentially pleiotropic effects.

It is apparent that phenotypic differences exist between the B-cells of HPV(+) and HPV(-) tumors. Data from an HPV-driven mouse tumor model supports a reduction in tumor growth as a result of the depletion of B-cells via anti-CD20 [45]. It is tempting to speculate that anti-CD200 could provide a more targeted approach to B-cell manipulation within these tumors rather than the use of anti-CD20, which would globally remove multiple B-cell subsets. In the clinic, an anti-CD200 blocking antibody was safe and well tolerated in Phase II testing (NCT00648739), but as yet the clinical effects have not been reported. The B-cell receptor targeting therapies (ibrutinib) have efficacy on CD200<sup>+</sup> cells within the B-CLL setting and potentially could be exploited [46].

This paper is the first to compare RNA-Seq data from HPV(+) and HPV(-) HNSCC, when controlled for TIL number. We have identified genes expressed both to a greater and lesser extent between tumor types. We are the first investigators to propose a correction calculation of transcriptomic signals in immune cells, to overcome the numerical imbalance between virally driven and virus-independent HNSCC. This analysis reveals that differences between the transcriptome in T-cells between HPV(+) and HPV(-) HNSCC appears predominantly quantitative, but that a distinct B-cell profile exists in HPV(+) cancers. However, to truly assess qualitative differences in B- and T-cells between HPV(+) and HPV(-) tumors, these subsets must be isolated and analysed separately. Such disaggregation of tumor tissue will be the focus of future work.

## MATERIALS AND METHODS

### Study subjects and tumor processing

Following LREC approval and written informed consent, 39 consecutive HNSCC samples were obtained from patients at three centers (Southampton, n=22; Poole, n=15; Liverpool, n=2) from 2010-2012. Tumor samples were collected, following general anaesthesia but before surgical resection, and were snap frozen immediately. Cryosections (10 $\mu$ m) were cut and used for RNA isolation with the RNeasy Mini Kit (Qiagen Ltd., Manchester, UK). Table 1 shows the patient demographics, tumor characteristics and tumor sampling/processing information for the HPV(+) and HPV(-) patient cohorts. A retrospective cohort of 544 HNSCC patients were also used for the generation of survival data.

## Histology and immunohistochemistry

Frozen tumor sections taken immediately adjacent to the tissue analysed by RNA-Seq were stained with hematoxylin and eosin (H&E); tumors were assessed as TIL high (TIL<sub>high</sub>), moderate (TIL<sub>mod</sub>) and low (TIL<sub>low</sub>) by an accredited pathologist [G.J.T] as previously described [7] (Supplementary Methods S1). Formalin-fixed, paraffin-embedded (FFPE) tumor tissue blocks from the same patients were also collected and used to evaluate the cell surface marker expression of CD3, CD4, CD8, CD20, CD23 and CD200 by IHC; enumeration of CD3, CD4, CD8 and CD20 was expressed as an average of ten high-power fields [7] (Supplementary Methods S1).

## Survival data

Two HNSCC patient cohorts consisting of 137 HPV(+) and 407 HPV(-) tumors were analysed retrospectively for survival relative to TIL density and HPV status. The primary endpoint was death from HNSCC, i.e., disease-specific survival, as previously described [7].

## RNA-Sequencing and data analysis

RNA-Seq (Single end, 35 bp) was performed using the HiSeq 2000 platform (Illumina Inc., San Diego, USA) (Supplementary Methods S2). RNA-Seq data have been deposited in the Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (NCBI) under accession number GSE72536. RNA-Seq data was mapped using human genome (hg19) and TopHat (version 2.0.9), counted with HTSeq-count (version 0.5.4) [47] and differentially expressed genes (DEGs) identified with EdgeR (version 3.4.2) [48, 49]. EdgeR was also used to identify DEGs while adjusting for the covariates associated with varying proportions of lymphocyte subsets in each tumor sample, gene expression of *CD19* (B-cells) and *CD4* and *CD8A* (T-cells) were used as the covariates. Unsupervised clustering of tumors was performed following variance stabilizing transformation of trimmed mean of M-values (TMM) normalized data. DEGs between HPV(+) and HPV(-) tumors were identified with a false discovery rate (FDR) corrected *p*-value of <0.05 (i.e., *q*-value <0.05) and a fold change of >2 or <-2. A detailed description of the RNA-Seq data analysis performed can be found in the Supplementary Methods S3.

## Real-time quantitative reverse transcription PCR

Real-time quantitative reverse transcription PCR (RT-qPCR) assays were performed using Taqman® probes for golgi-associated, gamma adaptin ear containing, ARF binding protein 2 (*GGA2*), ADAM metalloproteinase

domain 28 (*ADAM28*), *CD200*, Spi-B transcriptional factor (*SPIB*), stromal antigen 3 (*STAG3*), Vascular Cell Adhesion Molecule 1 (*VCAMI*), Inducible T-Cell Co-Stimulator Ligand (*ICOSLG*) and B-Cell CLL/Lymphoma 2 (*BCL2*) and adhere to the MIQE guidelines for RT-qPCR (Supplementary Methods S4) [50]. Gene expression (RT-qPCR) of *STAG3* and *CD200* was validated in the original tumor RNA, HPV(+) n=8 and HPV(-) n=8 patient tumor samples. B-cells (CD19<sup>+</sup>) isolated from an independent cohort of HPV(+) tumors (n=6) using a BD FACSAria™ sorter (BD Biosciences, Oxford, UK, Supplementary Methods S4) were assessed for gene expression of *ADAM28*, *BCL2*, *CD200*, *GGA2*, *ICOSLG*, *SPIB*, *STAG3*, and *VCAMI*. RT-qPCR analysis was performed using the comparative Ct (cycle threshold) method ( $2^{-\Delta\Delta Ct}$ ) using *Actin* as the control gene and is defined as a normalized relative gene expression compared to a the control gene [51].

## Gene ontology and pathway analysis

GO terms associated with biological processes and biological pathways that were significantly over-represented for DEGs (*q*-value <0.05) were identified with ConsensusPathDB [20] (CPDB, release 30) using the hypergeometric test. ConsensusPathDB represents a first generation tool for functional genomics and was sufficient for the purpose of showing the change in GO terms and pathways before and after correction for the number of B- and T-cells between HPV(+) and HPV(-) tumors. GO and pathway analyses were performed for genes that were expressed (i) to a greater extent and (ii) to a lesser extent in HPV(+) compared to HPV(-) tumors.

## Molecular quantification of TILs in HPV(+) and HPV(-) tumors

The distribution and proportions of TILs were assessed at both the tumor sample and group level. At the sample level, the gene expression (RNA-Seq) and surface protein expression (IHC) of *CD3*, *CD20*, *CD4* and *CD8* were evaluated. In addition, “Functional Analysis of Individual RNA-Seq or Microarray Expression” (FAIME) [52] was adapted to generate a score for a large number of tissues and cell types, including B-cells and CD4<sup>+</sup> and CD8<sup>+</sup> T-cells (additional details in Supplementary Methods S5 and Supplementary Table S6). The FAIME score was calculated for each cell type, for each tumor and was followed by a student’s t-test to assess whether the FAIME scores for a particular cell subset were significantly different (*q*-value <0.05) between the HPV(+) and HPV(-) cohorts.

## Validation of findings in TCGA data set

HNSCC RNA-Seq data (TCGA HNSC HiSeqV2 2015-02-24) was obtained from The Cancer Genome Atlas (TCGA) Genome Data Analysis Center (GDAC)

Firehose website ([http://gdac.broadinstitute.org/runs/stddata\\_2015\\_11\\_01/data/HNSC/20151101/](http://gdac.broadinstitute.org/runs/stddata_2015_11_01/data/HNSC/20151101/)), the RNA-Seq methodology and processing have been described by TCGA [11]. As HPV-driven cancers typically arise in the oropharynx, tonsil and base of tongue, we identified and evaluated tumors matched for these anatomical sites from TCGA. Unsupervised clustering of 46 HPV16(+) and 26 HPV(-) anatomically matched tumors from the oropharynx, tonsil and base of tongue was performed using the differential gene lists generated from our own analysis.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

## FINANCIAL SUPPORT

Cancer Research UK programme Grant C491/A15951 and NIHR and CR UK Experimental Cancer Medicine Centre C491/A20613

## REFERENCES

1. Argiris A, Karamouzis MV, Raben D and Ferris RL. Head and neck cancer. *Lancet*. 2008; 371:1695-1709.
2. Parkin DM, Bray F, Ferlay J and Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin*. 2005; 55:74-108.
3. Blot WJ, McLaughlin JK, Winn DM, Austin DF, Greenberg RS, Preston-Martin S, Bernstein L, Schoenberg JB, Stemhagen A and Fraumeni JF, Jr. Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer research*. 1988; 48:3282-3287.
4. Chaturvedi AK, Engels EA, Anderson WF and Gillison ML. Incidence trends for human papillomavirus-related and -unrelated oral squamous cell carcinomas in the United States. *Journal of clinical oncology*. 2008; 26:612-619.
5. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, Jiang B, Goodman MT, Sibug-Saber M, Cozen W, Liu L, Lynch CF, Wentzensen N, Jordan RC, Altekruze S, Anderson WF, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *Journal of clinical oncology*. 2011; 29:4294-4301.
6. Attner P, Du J, Nasman A, Hammarstedt L, Ramqvist T, Lindholm J, Marklund L, Dalianis T and Munck-Wikland E. The role of human papillomavirus in the increased incidence of base of tongue cancer. *International journal of cancer*. 2010; 126:2879-2884.
7. Ward MJ, Thirdborough SM, Mellows T, Riley C, Harris S, Suchak K, Webb A, Hampton C, Patel NN, Randall CJ, Cox HJ, Jogai S, Primrose J, Piper K, Ottensmeier CH, King EV, et al. Tumour-infiltrating lymphocytes predict for outcome in HPV-positive oropharyngeal cancer. *British journal of cancer*. 2014; 110:489-500.
8. Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tan PF, Westra WH, Chung CH, Jordan RC, Lu C, Kim H, Axelrod R, Silverman CC, Redmond KP and Gillison ML. Human papillomavirus and survival of patients with oropharyngeal cancer. *The New England journal of medicine*. 2010; 363:24-35.
9. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, Tosolini M, Camus M, Berger A, Wind P, Zinzindohoue F, Bruneval P, Cugnenc PH, Trajanoski Z, Fridman WH and Pages F. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*. 2006; 313:1960-1964.
10. Rooney MS, Shukla SA, Wu CJ, Getz G and Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*. 2015; 160:48-61.
11. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. 2015; 517:576-582.
12. Heusinkveld M, Goedemans R, Briet RJ, Gelderblom H, Nortier JW, Gorter A, Smit VT, Langeveld AP, Jansen JC and van der Burg SH. Systemic and local human papillomavirus 16-specific T-cell immunity in patients with head and neck cancer. *International journal of cancer*. 2012; 131:E74-85.
13. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A, Shefler E, Ramos AH, Stojanov P, Carter SL, Voet D, Cortes ML, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science*. 2011; 333:1157-1160.
14. Jung AC, Briolat J, Millon R, de Reynies A, Rickman D, Thomas E, Abecassis J, Clavel C and Wasyluk B. Biological and clinical relevance of transcriptionally active human papillomavirus (HPV) infection in oropharynx squamous cell carcinoma. *International journal of cancer*. 2010; 126:1882-1894.
15. Pyeon D, Newton MA, Lambert PF, den Boon JA, Sengupta S, Marsit CJ, Woodworth CD, Connor JP, Haugen TH, Smith EM, Kelsey KT, Turek LP and Ahlquist P. Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. *Cancer research*. 2007; 67:4605-4619.
16. Slebos RJ, Yi Y, Ely K, Carter J, Evjen A, Zhang X, Shyr Y, Murphy BM, Cmelak AJ, Burkey BB, Netterville JL, Levy S, Yarbrough WG and Chung CH. Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. *Clinical cancer research*. 2006; 12:701-709.
17. Russell S, Angell T, Lechner M, Liebertz D, Correa A, Sinha U, Kokot N and Epstein A. Immune cell infiltration patterns and survival in head and neck squamous cell carcinoma. *Head & neck oncology*. 2013; 5:24.
18. Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, Butterfoss D, Xiang D, Zanation A, Yin X,

- Shockley WW, Weissler MC, Dressler LG, Shores CG, Yarbrough WG and Perou CM. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer cell*. 2004; 5:489-500.
19. Thurlow JK, Pena Murillo CL, Hunter KD, Buffa FM, Patiar S, Betts G, West CM, Harris AL, Parkinson EK, Harrison PR, Ozanne BW, Partridge M and Kalna G. Spectral clustering of microarray data elucidates the roles of microenvironment remodeling and immune responses in survival of head and neck squamous cell carcinoma. *Journal of clinical oncology*. 2010; 28:2881-2888.
  20. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H and Herwig R. ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic acids research*. 2011; 39:D712-717.
  21. Shoemaker JE, Lopes TJ, Ghosh S, Matsuoka Y, Kawaoka Y and Kitano H. CTen: a web-based platform for identifying enriched cell types from heterogeneous microarray data. *BMC genomics*. 2012; 13:460.
  22. Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, Fong S, van Lookeren Campagne M, Godowski P, Williams PM, Chan AC and Clark HF. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun*. 2005; 6:319-331.
  23. Watkins NA, Gusnanto A, de Bono B, De S, Miranda-Saavedra D, Hardie DL, Angenent WG, Attwood AP, Ellis PD, Erber W, Foad NS, Garner SF, Isacke CM, Jolley J, Koch K, Macaulay IC, et al. A HaemAtlas: characterizing gene expression in differentiated human blood cells. *Blood*. 2009; 113:e1-9.
  24. Palmer C, Diehn M, Alizadeh AA and Brown PO. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC genomics*. 2006; 7:115.
  25. Grigoryev YA, Kurian SM, Avnur Z, Borie D, Deng J, Campbell D, Sung J, Nikolcheva T, Quinn A, Schulman H, Peng SL, Schaffer R, Fisher J, Mondala T, Head S, Flechner SM, et al. Deconvoluting post-transplant immunity: cell subset-specific mapping reveals pathways for activation and expansion of memory T, monocytes and B cells. *PloS one*. 2010; 5:e13358.
  26. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA and Brown PO. Individuality and variation in gene expression patterns in human blood. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:1896-1901.
  27. Prieto I, Suja JA, Pezzi N, Kremer L, Martinez AC, Rufas JS and Barbero JL. Mammalian STAG3 is a cohesin specific to sister chromatid arms in meiosis I. *Nature cell biology*. 2001; 3:761-766.
  28. Colmont CS, Benketaf A, Reed SH, Hawk NV, Telford WG, Ohyama M, Udey MC, Yee CL, Vogel JC and Patel GK. CD200-expressing human basal cell carcinoma cells initiate tumor growth. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:1434-1439.
  29. Moreno-Galindo C, Hermsen M, Garcia-Pedrero JM, Fresno MF, Suarez C and Rodrigo JP. p27 and BCL2 expression predicts response to chemotherapy in head and neck squamous cell carcinomas. *Oral oncology*. 2014; 50:128-134.
  30. Ward MJ, Mellows T, Harris S, Webb A, Patel NN, Cox HJ, Piper K, Ottensmeier CH, Thomas GJ and King EV. Staging and treatment of oropharyngeal cancer in the human papillomavirus era. *Head & neck*. 2014.
  31. Liu D, Xu H, Shih C, Wan Z, Ma X, Ma W, Luo D and Qi H. T-B-cell entanglement and ICOSL-driven feed-forward regulation of germinal centre reaction. *Nature*. 2014.
  32. Kobayashi A, Darragh T, Herndier B, Anastos K, Minkoff H, Cohen M, Young M, Levine A, Grant LA, Hyun W, Weinberg V, Greenblatt R and Smith-McCune K. Lymphoid follicles are generated in high-grade cervical dysplasia and have differing characteristics depending on HIV status. *The American journal of pathology*. 2002; 160:151-164.
  33. Nelson BH. CD20+ B cells: the other tumor-infiltrating lymphocytes. *Journal of immunology*. 2010; 185:4977-4982.
  34. Martin-Orozco N, Li Y, Wang Y, Liu S, Hwu P, Liu YJ, Dong C and Radvanyi L. Melanoma cells express ICOS ligand to promote the activation and expansion of T-regulatory cells. *Cancer research*. 2010; 70:9581-9590.
  35. Fan X, Quezada SA, Sepulveda MA, Sharma P and Allison JP. Engagement of the ICOS pathway markedly enhances efficacy of CTLA-4 blockade in cancer immunotherapy. *The Journal of experimental medicine*. 2014; 211:715-725.
  36. Su GH, Ip HS, Cobb BS, Lu MM, Chen HM and Simon MC. The Ets protein Spi-B is expressed exclusively in B cells and T cells during development. *The Journal of experimental medicine*. 1996; 184:203-214.
  37. Liang C, Marsit CJ, McClean MD, Nelson HH, Christensen BC, Haddad RI, Clark JR, Wein RO, Grillone GA, Houseman EA, Halec G, Waterboer T, Pawlita M, Krane JF and Kelsey KT. Biomarkers of HPV in head and neck squamous cell carcinoma. *Cancer research*. 2012; 72:5004-5013.
  38. Barclay AN, Wright GJ, Brooke G and Brown MH. CD200 and membrane protein interactions in the control of myeloid cells. *Trends in immunology*. 2002; 23:285-290.
  39. Holmannova D, Kolackova M, Kondelkova K, Kunes P, Krejsek J and Andrys C. CD200/CD200R paired potent inhibitory molecules regulating immune and inflammatory responses; Part I: CD200/CD200R structure, activation, and function. *Acta Medica (Hradec Kralove)*. 2012; 55:12-17.
  40. Badoual C, Hans S, Rodriguez J, Peyrard S, Klein C, Agueznay Nel H, Mosseri V, Laccourreye O, Bruneval P, Fridman WH, Brasnu DF and Tartour E. Prognostic value of tumor-infiltrating CD4+ T-cell subpopulations in head and



- neck cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2006; 12:465-472.
41. Lindner S, Dahlke K, Sontheimer K, Hagn M, Kaltenmeier C, Barth TF, Beyer T, Reister F, Fabricius D, Lotfi R, Lunov O, Nienhaus GU, Simmet T, Kreienberg R, Moller P, Schrezenmeier H, et al. Interleukin 21-induced granzyme B-expressing B cells infiltrate tumors and regulate T cells. *Cancer research*. 2013; 73:2468-2479.
  42. Iwata Y, Matsushita T, Horikawa M, Dilillo DJ, Yanaba K, Venturi GM, Szabolcs PM, Bernstein SH, Magro CM, Williams AD, Hall RP, St Clair EW and Tedder TF. Characterization of a rare IL-10-competent B-cell subset in humans that parallels mouse regulatory B10 cells. *Blood*. 2011; 117:530-541.
  43. Rygiel TP and Meyaard L. CD200R signaling in tumor tolerance and inflammation: A tricky balance. *Current opinion in immunology*. 2012; 24:233-238.
  44. Wang L, Liu JQ, Talebian F, El-Omrani HY, Khattabi M, Yu L and Bai XF. Tumor expression of CD200 inhibits IL-10 production by tumor-associated myeloid cells and prevents tumor immune evasion of CTL therapy. *European journal of immunology*. 2010; 40:2569-2579.
  45. Affara NI, Ruffell B, Medler TR, Gunderson AJ, Johansson M, Bornstein S, Bergsland E, Steinhoff M, Li Y, Gong Q, Ma Y, Wiesen JF, Wong MH, Kulesz-Martin M, Irving B and Coussens LM. B cells regulate macrophage phenotype and response to chemotherapy in squamous carcinomas. *Cancer cell*. 2014; 25:809-821.
  46. Woyach JA, Johnson AJ and Byrd JC. The B-cell receptor signaling pathway as a therapeutic target in CLL. *Blood*. 2012; 120:1175-1184.
  47. Anders S, Pyl PT and Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2014.
  48. Nikolayeva O and Robinson MD. edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. *Methods in molecular biology*. 2014; 1150:45-79.
  49. Robinson MD, McCarthy DJ and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26:139-140.
  50. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J and Wittwer CT. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry*. 2009; 55:611-622.
  51. Livak KJ and Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  Method. *Methods*. 2001; 25:402-408.
  52. Yang X, Regan K, Huang Y, Zhang Q, Li J, Seiwert TY, Cohen EE, Xing HR and Lussier YA. Single sample expression-anchored mechanisms predict survival in head and neck cancer. *PLoS computational biology*. 2012; 8:e1002350.