The Mutational Landscape of Recurrent vs Non-Recurrent Human Papillomavirus-Associated Oropharyngeal Cancer

Richard Alexander Harbison

A thesis submitted

in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2017

Committee:

Stephen M. Schwartz (Chair)

Eduardo Mendez

Chu Chen

Christopher S. Carlson

Program Authorized to Offer Degree:

Epidemiology

University of Washington

Abstract

The Mutational Landscape of Recurrent vs Non-Recurrent Human Papillomavirus-Associated Oropharyngeal Cancer

Richard Alexander Harbison

Chair of the Supervisory Committee:

Stephen M. Schwartz, PhD

Professor and Interim Chair

Department of Epidemiology

Current therapies are effective for a subset of human papillomavirus (HPV)-associated head and neck cancers (HNC), although up to a quarter of these cases fail treatment. Identification of genomic alterations associated with therapeutic response or lack thereof from the outset of treatment would have substantial clinical implications. We hypothesized that HPV-related oropharyngeal squamous cell carcinoma (OPSCC) that recur share similar genomic features to HPV-unrelated HNC, a clinically aggressive phenotype. Data on 57 cases of HPV-related primary OPSCC tumors from three sources, of which 38 did not recur and 19 recurred, were included in the analyses of recurrence. *KMT2D*, a histone H3 lysine 4 methyltransferase, was the most frequently mutated gene across tumors (16%) and among the recurrent cases (21%). DNA damage response and MAPK genes had the strongest association with recurrence (Multivariable Prevalence Ratio (PR) (95% CI): 1.22 (0.88 – 1.68) and 1.23 (0.89 – 1.70)). Gene mutations and/or copy number aberrations with the strongest association with recurrence in multivariable analyses included *CASP8* (PR (95% CI): 3.83 (1.06 – 13.8)), *NSD1* (PR (95% CI): 2.34 (0.98 – 5.60)), and *PEG3* (PR (95% CI): 3.10 (1.04 – 9.16)). These data suggest that

frequently altered genes involved in HPV-unrelated HNC tumorigenesis may play a key role in recurrence in HPV-related OPSCC.

Statement of Significance

Surgery and/or platinum-based chemoradiotherapy is effective for a subset of HPV-related head and neck squamous cell carcinomas, yet up to 25% fail treatment. Understanding the genomic aberrations associated with recurrence will build a foundation for developing patient-specific, targeted treatment from the outset of therapy to minimize the risk of recurrence.

Introduction

Human papillomavirus (HPV)-related cancers are a burgeoning public health concern. HPV infects more than half of U.S. adults [1], and HPV-related oropharyngeal squamous cell carcinoma (OPSCC) incidence is rising. In fact, HPV-related OPSCC is expected to surpass the incidence of cervical cancer by 2020 while HPV-unrelated smoking and alcohol-related head and neck cancers (HNC) are decreasing in incidence [2]. Although a subset of HPV-related OPSCC tumors respond well to concurrent cisplatin chemoradiation [3], disease progression occurs in up to 25% of cases within three years of treatment completion [4,5]. Furthermore, patients with recurrence fare poorly with 40% mortality at two-years post-disease progression [5]. Given ongoing interest in de-intensifying chemoradiation protocols [6], it is paramount to identify appropriate candidates for de-intensification from the start of therapy so that alternative therapeutic approaches can be considered for patients with genomic profiles portending a poor prognosis. Furthermore, understanding genomic characteristics of HPV-related tumors that confer a treatment-resistant phenotype may generate insight into molecular pathways that are relevant to treatment-resistance across cancer types.

Compared to HPV-unrelated head and neck cancer, HPV-related tumors were characterized by a paucity of mutations in one study [7] whereas they had a similar mutation burden in another [8]. While HPV-unrelated HNC is mediated by carcinogens such as alcohol and smoking,

HPV16/18 E6 and E7 oncoproteins play a key role in HPV-related tumorigenesis. E6 and E7 mediate inactivation [9-12] of the p53 and pRb tumor suppressors contributing to genomic instability and a resulting propensity for genomic alterations and tumorigenesis. Moreover, HPV oncoprotein-mediated genomic instability [13] and viral integration into cancer-associated genes further accelerate tumorigenesis [14,15]. Despite our current understanding of this pathogen-mediated cancer, the non-HPV genomic drivers associated with recurrence in this population have yet to be evaluated due to the relatively rarity of this cancer and favorable treatment response.

Prior assessments of the genomic landscape of head and neck cancer were not limited to HPV-related OPSCC but instead focused on analyzing a broad variety of head and neck tumor subsites and/or metachronous or recurrent tumors [8,16]. Morris et al. [16] studied HPV-related and -unrelated recurrent and metachronous HNC to test the feasibility of the MSK-IMPACT precision oncology assay in this population. The Cancer Genome Atlas (TCGA) study [8] characterized the genomic landscape of a large sample of HNC to uncover the overall molecular profile of head and neck cancer. These studies did not to evaluate the association between genomic characteristics and recurrence.

Given the stark survival differences between HPV-related and HPV-unrelated HNC, we hypothesized that primary HPV-related OPSCC tumors that recur share features with HPV-unrelated head and neck cancer. This includes a propensity for tumor suppressor inactivation, a greater mutational burden, and genomic instability. Furthermore, we hypothesized that the mutational landscape of primary tumors that recurred would harbor a preponderance of mutations in mitogenic signaling (i.e., PI(3)K, JAK/STAT, MAPK, receptor tyrosine kinase (RTK)), cell cycle, cell death, survival and differentiation pathways as frequently observed in HPV-unrelated tumors [8].

In this study, we sought to compare the genomic profiles of primary HPV-related OPSCC tumors that did or did not recur to test the hypothesis that tumors that recur share genomic features in common with HPV-unrelated HNC including alterations in signaling pathways critical to tumorigenesis.. Data for the samples included in this study were derived from the TCGA [8] study as well as from samples from the Universities of Washington (UW) and Pittsburgh (UPitt).

Results

Patient Baseline Characteristics

Data from sixty-six patients were included in this study following application of inclusion and exclusion criteria (**Supplementary Methods**, Figure S1). Patients ranged from 35 to 77 years old. Patients with tumors that did or did not recur tended to be over 55 years of age and were predominantly male (Table 1). Using a contemporary staging system for HPV-related OPSCC [17], 75% of patients (n = 29) that did recur or did not recur (n = 20) were stage IA or IB (i.e., encompassing T1/T2 tumors with N0 through N2 nodal disease). Fifteen-percent (n = 6) of patients without recurrence had stage III disease—defined by the presence of a T4 tumor—compared to 7% (n = 2) of patients with recurrence. Most cases are early stage reflecting the favorable prognosis even with regional disease. Nearly one-half (46%, n = 18) of patients without recurrence had a history of tobacco use greater than 10 years compared to 30% (n = 8) of patients whose tumors recurred.

Somatic Mutations

Mean fold tumor target coverage for UW and UPitt whole exome sequencing (WES) data was consistent with TCGA [8] coverage at 93X with 95% of target bases above 20X coverage (**Supplemental Table S1**). Mean fold normal tissue target coverage was 88X with 96% of target bases above 20X coverage. A total of 57 samples were included in genomic analyses following

whole exome sequencing quality control. Among 57 samples, we observed 8,755 nonsynonymous and 21,709 synonymous somatic nucleotide variants (SNVs). We evaluated the mutational landscape across HPV-related primary OPSCC tumors that remained following exclusion criteria in our variant calling pipeline (**Supplemental Figure S2**). We implemented MutSigCV to identify genes mutated more often than expected by chance considering patient-specific somatic mutation frequency, mutation spectrum, and genomic position-based factors (e.g., DNA replication timing and chromatin state estimation). MutSigCV analysis identified significant mutations in *ZNF750* and *RB1* across all HPV-related OPSCC tumors (*FDR q-value* < 0.1; **Supplemental Table S2a**, *left column*).

Next, we analyzed somatic mutations within each group of HPV-related OPSCC tumors that did or did not recur. The total number of mutations was greater among the primary HPV-related OPSCC tumors that recurrence compared to those that did not recur. Tumors that recurred had 5,066 nonsynonymous and 18,529 synonymous variants from a total of 19 tumors. A substantial proportion of this mutational burden was associated with six hypermutated samples (> 1,000 total mutations per tumor) more consistent with that seen in lung squamous cell carcinoma [18]. Thirty-eight primary HPV-related OPSCC tumors that did not recur encompassed 3,689 nonsynonymous variants and 3,180 synonymous variants (four tumors did not contain somatic variants after filtering). Tumors that did not recur did not have greater than 1,000 mutations. No genes were significantly mutated in either group alone based on MutSigCV analysis. While not significant, the top mutated genes from cases with recurrence were *MACC1*, *CPS1*, and *FAM193A* (*FDR q-value* > 0.1, **Supplemental Table S2a**, *right column*). *MACC1* is a transcription activator for MET which activates multiple signaling pathways including RAS/MAPK [19], PI(3)K/Akt [20], STAT [21,22], Wnt/β-catenin [23], and Notch [24] signaling pathways. In comparison, the top mutated genes from cases without recurrence were *B2M*, *RB1*, and *ZNF750* (*FDR q-value* > 0.1, **Supplemental Table S2a**, *center column*).

In order to focus the analysis on genes and pathways critical to tumorigenesis and/or that are potentially targetable, we limited the dataset to mutations affecting a set of 467 "driver genes" based on prior work from Vogelstein et al. [25], Hedberg et al. [26], and Pritchard et al. [27] (**Supplemental Table S3**). Thirty-eight primary HPV-related OPSCC tumors that did not recur had 165 nonsynonymous variants and 108 synonymous variants affecting any of the 467 driver genes. Five tumors that did not recur did not have a mutated driver gene. Nineteen primary HPV-related OPSCC tumors that recurred had 169 nonsynonymous and 485 synonymous variants in driver genes, again reflecting the higher overall mutational burden of critical driver genes in the pool of tumors that recurred. When restricting the MutSigCV analysis to the 467 driver genes, *FGFR3*, *CYLD*, *RB1*, and *PTEN* were significantly mutated across all tumors (*FDR q-value* < 0.1, **Supplemental Table S2b**, *left column*). *EP300*, *KMT2D*, *GSTP1*, and *TRAF3* were also highly mutated in MutSigCV analysis, although they did not reach significance per se.

Mutations involving driver genes that affected at least 5% of samples are illustrated in Figure 1a. *KMT2D*, a lysine methyltransferase that regulates the chromatin state at Notch target and other genes [28], was the most frequently mutated gene across all primary HPV-related OPSCC tumors with six truncating mutations and nine missense mutations affecting multiple regions of the resulting amino acid sequence (Figure 1b). It was also co-mutated with *NOTCH1* and/or *NOTCH3* in two samples from the recurrences. While *KMT2D* was not statistically significantly mutated, it trended towards significance in MutSigCV analysis of all primary HPV-related OPSCC tumors (*FDR q-value* = 0.35) and was one of the top genes in MutSigCV analysis of the primary tumors that recurred (**Supplemental Table S2b**, *right column*). Not only was *KMT2D* the most frequently mutated gene across all tumors, it was also the most frequently mutated gene among the tumors that recurred (Figure 2). It was also more frequently altered in primary tumors that recurred (21%) than those that did not recur (13%).

KMT2D truncating mutations including *Q56**, *S1632**, and *Q4687** were observed in the TCGA head and neck data [8]. *KMT2D* truncating mutation, *Q4347**, was found in the TCGA head and neck data as well as two samples from the UW data. *Q2458** was observed in the TCGA head and neck data as well as in a bladder cancer sample from the MSK-IMPACT study [29]. It was also observed in one UPitt sample from this study. Several *KMT2D* missense mutations were found in this study including *R83Q* which is seen in renal carcinoma and hemangioblastoma [30]. We noted other missense mutations that have not been previously described (*D1114Y*, *E820Q*, *L5326L*, *R125Q*, *S4010P*, *T1983N*).

We also compared the mutational landscape of primary HPV-related OPSCC tumors that did or did not recur limited to the driver gene set (Figure 2). There were no significantly mutated genes within either group alone based on MutSigCV analysis (**Supplemental Table S2b**, *center and right columns*). The top genes identified by MutSigCV analysis among tumors that did not recur involved PI(3)K signaling (i.e., *FGFR3*, *PTEN*, and *PIK3CA*), cell cycle (i.e., *RB1*), and immune signaling (i.e., *B2M*, *CYLD*, *TRAF3*, and *IL6*) (**Supplemental Table S2**, *center column*). MutSigCV identified several candidate genes among primary HPV-related OPSCC tumors that recurred which involved RTK signaling (i.e., *FLT1*, *IGF1*, *FGFR3*), MAPK signaling (i.e., *MAP2K5*), PI(3)K signaling (i.e., *PIK3R4*, *PIK3CA*, *PIK3CB*), DNA damage response (i.e., *PMS2*, *NBN*), Wnt signaling (i.e., *MED12* and *SOX17*), epigenetic regulation (i.e., *KMT2D* [*MLL2*], *HIST1H3B*, *HIST1H1C*), and oxidative stress response (e.g., *NFE2L2*) (**Supplemental Table S2b**, *right column*).

Genes involved in mitogenic signaling, DNA damage repair, epigenetic signaling, differentiation, and oxidative stress response were substantially mutated based on MutSigCV analysis in tumors that recurred. Moreover, frequency-based analysis revealed several genes that were mutated at a higher rate or exclusively in tumors that recurred (Figure 2). Among the driver gene set, genes involved in mitogenic signaling (i.e., *TSC2*, *KIT*, *MTOR*, *FLT1*), cell death

signaling (i.e., *PEG3*, *BCL2L11*), DNA damage response (i.e., *BRIP1*, *MLH3*, *NBN*, *POLQ*), Notch signaling (i.e., *NOTCH1*), oxidative stress response (i.e., *NFE2L2*), epigenetic (i.e., *BRD4*), and Wnt signaling (i.e., *MED12*) were more frequently mutated among the recurrent cases (Figure 2; *FDR q-value* > 0.1, Fisher exact test). The genes most strongly associated with recurrence based on Fisher exact tests comparing between-group frequencies were *NFE2L2* and *TSC2* (Figure 2).

Copy Number Variants

Copy number variants (CNVs) were inferred using whole exome sequencing coverage results from TCGA, UW, and UPitt data. Tumor coverage was compared to matched normal coverage to obtain raw copy number estimates at each base and segmented to estimate copy number altered regions. Figure 3a illustrates raw copy number data for the HPV-related primary OPSCC tumors that did or did not recur across 53 samples for which copy number data were available. The fraction of exome harboring copy number amplifications did not differ between tumors that did or did not recur (**Supplemental Table S4**, median (IQR) 0.087 (0.16) versus 0.081 (0.11), n = 18 and n = 35, p > 0.05, Wilcoxon rank sum test). Copy number deletions also did not differ between cases that did or did not recur (**Supplemental Table S4**, median (IQR) 0.12 (0.093) versus 0.095 (0.10), n = 18 and n = 35, p > 0.05, Wilcoxon rank sum test). Most tumors demonstrated gains of 1p, 3q, 8q, 11q, 14q, and 20q with copy number losses of 2q, 3p, 11q, 13q, and 14q (Figure 3a, **Supplemental Figure S3**, **Supplemental Tables S5 and S6**).

HPV-related primary OPSCC tumors that did not recur featured amplifications of 3q, 11q, and 20q harboring *PIK3CA*, *SOX2*, *CCND1*, and *E2F1* among other genes (Figures. 3b and 3c, *top left panel*, **Supplemental Table S7**). Copy number deletions in the tumors that did not recur were prevalent on 1p, 2q, 10q, 11q, 13q, 14q, and 20p which contain key tumor suppressors (*PTEN*, *TP73*, and *RB1*), DNA damage repair (*ATM*), epigenetic regulatory (*MLL*), and genes

involved in cell death regulation (*CASP1/4/5/12*; Figures. 3b and 3c, *top right panel*, **Supplemental Table S8**).

In contrast, primary HPV-related OPSCC tumors that recurred featured amplification on 20q where *E2F1* resides (Figures 3b and 3c, *bottom left panel*, **Supplemental Table S9**). Focal deletions were noted on several chromosomes including 14q which encodes *JAG2*, a ligand for Notch signaling (Figs, 3b and 3c, *bottom right panel*, **Supplemental Table S10**). Moreover, *KMT2A* [*MLL*], or lysine methyltransferase 2A, on 11q which is involved in chromosomal translocations in the development of acute leukemia was frequently deleted in both tumors that did and did not recur. Deletions on 11q (*CASP1/4/5/12*, *ATM*, *BIRC2/3*) and 14q (*TRAF3*) were also featured among the tumors that recurred implicating DNA damage repair response and pro-survival constituents.

In summary, we observed a greater overall quantity of mutations among the primary tumors that recurred compared to those that did not recur. The non-recurrent cases had a predilection for PI(3)K, cell cycle, and immune signaling pathway aberrations whereas the recurrent cases featured mitogenic signaling, epigenetic, DNA damage repair, oxidative stress, and Wnt signaling mutations more in line with HPV-negative head and neck cancer. Primary tumors that recurred hosted a higher frequency of *KMT2D* mutations. *KMT2D* was also the most frequently mutated gene across tumors. The copy number landscape was similar between groups with both groups harboring *E2F1* amplification and 11q and 14q deletions.

Comparison to HPV-Negative HNSCC

We compared the genomic profiles of HPV-related OPSCC tumors that did or did not recur to 59 TCGA HPV-negative oral cavity (OCSCC) and OPSCC tumor data to test the hypothesis that tumors that recur have a higher frequency of HPV-negative-like genomic aberrations than tumors that do not recur. HPV-negative-like genomic aberrations were based on TCGA findings

coupled with prior investigations [7,8,31]. We tested the mutational burden among primary HPV-related OPSCC tumors that did not recur versus that did recur compared to TCGA HPV-unrelated OCSCC/OPSCC (Figure 4a). We did not observe a statistically significant difference in the median mutation rates between HPV-related OPSCC tumors that did not recur versus those that did recur or HPV-unrelated tumors (*p-value* = 0.20, Kruskall-Wallis rank sum test). Hypermutated tumors in the group of tumors that recurred account for the wide variation in mutation rate among this group. The median (IQR) mutations per sample were 82 (136.5) and 71 (464.75) for HPV-related primary OPSCC tumors that did not and did recur, respectively, compared to 101 (120) for HPV-unrelated tumors.

Following mutational and copy number analyses, we merged mutation and CNV data to assess overall differences in genomic aberrations among a set of 41 significantly mutated and/or copy number altered genes found in in HPV-unrelated tumors as described in the TCGA study [8] (**Supplemental Table S11**). Among these genes were *CDKN2A*, *TP53*, *NOTCH1*, *FAT1*, *AJUBA*, *CASP8*, *PIK3CA*, *KMT2D*, *NSD1*, *HLA-A* and others. Copy number variants were classified using GISTIC2.0 analysis categorizations of high-level amplification, copy loss, or copy neutral. The following analyses were limited to samples for which mutational and copy number data were available.: 35 HPV-related OPSCC primary tumors that did not recur, 18 HPV-related OPSCC primary tumors that did recur, and 59 HPV-unrelated OCSCC/OPSCC tumors . Multiple correspondence analysis (MCA) illustrated greater overlap between the HPV-related OPSCC primary tumors that did or did not recur than with the HPV-unrelated HNSCC tumors (Figure 4b). *PSIP1*, *JAK2*, *CD274* [*PDL1*], *PTPRD*, *PDE4D*, and other genes critical to tumorigenesis such as *CDKN2A*, *TP53*, *NOTCH1*, *FAT1*, *KDM6A*, and *NFE2L2* had the greatest contribution to the variance encompassed by the first dimension of the MCA (Supplemental Figure S4a). *FADD*, *CTTN*, *CCND1*, *BIRC2*, and *EGFR* had the greatest

contribution to the second dimension of the MCA which begins to separate HPV-related OPSCC tumors that did or did not recur (Supplemental Figure 4b).

Limiting the analysis to tumors with both somatic mutation and copy number data among the 41-gene HPV-negative-like set, *NOTCH1* was deleted and/or mutated in 33% and 11% of HPV-related tumors with recurrence compared to 11% and 3% in tumors without recurrence (**Supplemental Tables S12a and S12c**). *TP53* was deleted and/or mutated in 39% and 0% of tumors that recurred vs 26% and 3% of tumors that did not recur. *AJUBA* was mutated in 11% and amplified in one tumor that recurred compared to no mutations and deletions in 17% of tumors that did not recur. *KMT2D* was mutated in 22% of tumors that recurred compared to 14% of tumors that did not recur. *NSD1* was mutated in 11% and deleted in 39% of tumors that recurred compared to 6% mutations and 17% deletions among the non-recurrent tumors. *HLA-A* and *NFE2L2* were mutated and/or deleted mostly in tumors that recurred.

Moreover, we were interested in investigating genomic aberrations among biological pathways central to HNC tumorigenesis. With the goal of identifying genomically altered, targetable pathways conferring a propensity for tumor recurrence, we calculated the frequency of SNVs and/or CNVs in four major mitogenic pathways, DNA damage repair pathway genes (DDR), cell cycle and death, and differentiation genes. Pathway components were defined as follows: **PI3K** (*PIK3CA/D/G*, *PIK3AP1*, *PIK3C2A/B/G*, *PIK3IP1*, *PIK3R1/2/3/4/5/6*, *AKT1/2/3*, *MTOR*, *PTEN*, *PDK1*, *TSC1/2*, *RICTOR*, *RPTOR*), **JAK/STAT** (*JAK1/2/3*, *STAT1/2/3/4/5A/5B/6*, *SOCS3*, *SHP2*, *IL6*, *IL6R*, *IL6ST*), **MAPK** (*ERK1/2*, *MEK1/2*, *RAF1*, *ARAF*, *BRAF*, *HRAS*, *KRAS*, *NRAS*, *SHC1/2/3*, *GRB2*), **RTK** (*EGFR*, *FGFR1/2/3*, *ERBB2*, *EGFR1*, *EPHA2*, *DDR2*, *MET*). **DDR**, **cell cycle**, **cell death**, and **differentiation** pathway components are outlined in **Supplemental Table S13**. Figure 5 compares HPV-related primary OPSCC tumors that did not or did recur versus HPV-negative OCSCC and OPSCC tumors. JAK/STAT, MAPK, DNA damage repair, and cell death pathway alterations were more frequent among the HPV-related

OPSCC tumors that recurred compared to those that did not recur (Figure 5, **Supplemental Table S14a and S14b**).

To test the strength of the association between pathway alteration frequency and recurrence among the primary HPV-related OPSCC tumors, we estimated univariable and multivariable prevalence ratios and their associated 95% confidence intervals (Table 2). MAPK, DNA damage response, and differentiation pathway hits had the strongest association with recurrence in univariable analyses. In multivariable analyses MAPK, DNA damage response, and JAK/STAT pathways had the strongest association with recurrence when adjusting for all pathways, although not significant. The adjusted prevalence ratio for recurrence for the MAPK pathway was 1.23 (95% CI: 0.89 – 1.70) for each additional pathway hit. For each additional DNA damage response pathway aberration, patients had a 1.22 (95% CI: 0.88 – 1.68) times greater likelihood of recurrence. For JAK/STAT pathway hits, the prevalence ratio was 1.17 (95% CI: 0.85 – 1.61).

Combined Somatic Mutation and Copy Number Genomic Analyses

We also sought to identify the strength of association between specific mutational and/or copy number gene alterations with recurrence. We evaluated genes mutated in greater than 15% of primary OPSCC tumors that recurred (Figure 2a; *NFE2L2*, *TSC2*, *PEG3*, *FBXW7*, *PIK3CA*, *KMT2D*) and significantly mutated genes from HPV-unrelated head and neck cancer as described by the TCGA study [8] (*CDKN2A*, *TP53*, *FAT1*, *AJUBA*, *CASP8*, *NSD1*, *HLA-A*). These 14 genes were used as covariates in a generalized linear model to estimate the prevalence ratios for the association between genomic alterations (i.e., gene mutations and/or copy number variants) with tumor recurrence. Potential confounders including age, gender, stage, and smoking history were not included in the final estimates as they were well-balanced between groups except for smoking which had a higher frequency in the tumors that did not recur.

In the univariable analysis, *HLA-A* and *PEG3* were associated with recurrence (Table 3). In the multivariable analysis, *PEG3* (PR 3.10 [95% CI 1.04 – 9.16]) and *CASP8* (HR 3.83 [95% CI 1.06 – 13.8]) were associated with recurrence. *HLA-A*, *NSD1*, *KMT2D*, and *FAT1* tended towards an association with recurrence, although not significant.

Discussion

With the epidemic of HPV infection in the United States [1] coupled with a rapidly rising incidence of HPV-related head and neck cancer [2], there is an urgent public health need to optimize the management of these patients from the outset of treatment. This is critical given the younger age group affected by HPV-related tumors and the potential long-lasting impact on quality-of-life, longevity, and lost productivity. Moreover, cumulative mortality among HPV-related head and neck cancers with disease progression is upwards of 40% warranting deeper investigation into patient-specific, targeted-therapies for mitigating recurrence from the outset of treatment.

In this report, we combined bioinformatic and epidemiologic approaches to identify mutationally- and/or copy number-altered, targetable genes and pathways associated with HPV-related OPSCC tumor recurrence. Analysis of all currently available OPSCC data from clinically annotated tumors with known recurrence status from 57 tumors from TCGA, UW, and UPitt illustrated several key findings. The most frequently recurring genomic aberrations in this study recapitulate the results of prior work [8] evaluating HPV-related OPSCC in that we observed a propensity for *PIK3CA* mutations, 3q amplification, *E2F1* amplification, and intact 9p21 which contains *CDKN2A* plus a paucity of *TP53* mutations. Overall, we observed a predilection for genomic aberration in mitogenic signaling pathways as illustrated by frequent mutations in *FGFR3*, *PIK3CA*, *STAT3*, *FLT1*, *MTOR*, and *TSC2* (Figure 1). These findings are consistent

with work by Lui et al. illustrating the association between *PIK3CA* pathway mutations and advanced stage head and neck cancer [32].

Additionally, we observed frequent mutations and copy number aberrations affecting Notch family members. Differentiation pathways, predominantly involving Notch mutations, are implicated in head and neck carcinogenesis [7,8,31], cutaneous malignancies, and lung cancers [33]. Interestingly, our analysis considering all mutations across study samples revealed that *ZNF750* was significantly mutated [8]. *ZNF750* plays a role in terminal epidermal differentiation as a target of *TP63* [34].

A surprising finding was the prevalence of *KMT2D* mutations. *KMT2D*, also known as *MLL2*, is infrequently involved in other pathogen-mediated cancers and functions as a tumor suppressor in follicular lymphoma and diffuse large B cell lymphoma [35]. Target genes of *KMT2D* include Notch family members [28], CD40, JAK/STAT, Toll-like receptors, B-cell receptor signaling pathways and other tumor suppressor genes such as *TNFAIP3*, *SOCS3*, and *TNFRSF14* [35]. Furthermore, *KMT2D* converges on MAPK signaling pathways in Kabuki syndrome, a disorder with developmental delay and congenital anomalies, resulting in attenuated MEK/ERK activation [36]. In Kabuki syndrome, *KMT2D* is often inactivated [36]. In the current study, MAPK pathway aberrations had the strongest association with recurrence (Table 2). Another mitogenic signaling pathway implicated in KMT2D chromatin regulatory function is PI3K signaling. Toska et al. elegantly illustrated the effect of PI(3)Kα inhibition on enhancing *KMT2D* activity in estrogen receptor-positive breast cancer as a mechanism of resistance to PI(3)Kα inhibition [37]. We observed a trend towards mutual exclusivity between *PIK3CA* and *KMT2D* mutations. Of three *PIK3CA* mutations in each of the groups that did or did not recur, four of six mutations were in non-*KMT2D*-altered samples.

In our study, we observed a *KMT2D* truncating mutation resulting in the amino acid change, Q4337* in the primary HPV-related OPSCC tumor of a patient initially treated with surgery and

radiotherapy. Another primary tumor in our study treated upfront with surgery and concurrent chemoradiation had a mutation leading to the R1252Q amino acid change. Further investigation into the interaction between downstream *KMT2D* targets with mitogenic signaling pathways as they relate to chemoradiotherapy resistance is warranted.

Preliminary work by Morris et al. [16] found that recurrent and metastatic HPV-related head and neck cancers had genomic profiles closer to HPV-unrelated tumors. Our data suggest that the mutational burden does not vary by HPV-status. In contrast, Stransky et al. [7] found that the mutation rate of HPV-related tumors was approximately half that of HPV-unrelated HNC. One plausible reason for this discrepancy is the prevalence of smoking among patients in this study population. Alternatively, APOBEC-mediated cytosine deamination, a substantial mutagenic factor in cancer, may contribute the genomic instability of HPV-related OPSCC as APOBEC activity is induced in virally-infected cells [38]. It is possible that APOBEC-mediated genomic instability could increase the likelihood of random mutations permitting treatment-resistance which may not be strongly associated with an increased overall mutation burden. Prior work also linked APOBEC activity to generation of helical hot spot mutations in *PIK3CA* mutations across multiple cancers [38] which could promote resistance independent of a high mutational burden.

Analysis of the combined somatic mutation and copy number landscape between HPV-unrelated OCSCC/OPSCC and HPV-related OPSCC primary tumors that did or did not recur suggests greater overlap between the primary HPV-related OPSCC tumors than with the HPV-unrelated tumors (Figure 4b). We hypothesized that HPV-related OPSCC tumors that recurred would share similar features with HPV-unrelated head and neck tumors given the marked differences in disease-free survival observed when comparing HPV-related to -unrelated HNC. However, there were notable similarities between the HPV-related primary OPSCC tumors that recurred and HPV-negative OCSCC/OPSCC tumors. For example, *NOTCH1* and *TP53*

deletions occurred more frequently among the HPV-related tumors that recurred than the HPV-related OPSCC tumors that did not recur. However, *TP53* mutations were rare among the HPV-related tumors consistent with prior work. Also, *AJUBA* was mutated in 11% of HPV-related tumors that recurred and no tumors that did not recur. Moreover, *KMT2D* was mutated in 21% of HPV-related OPSCC tumors that recurred compared to 13% of HPV-related tumors that did not recur. *HLA-A* and *NFE2L2* were mutated and/or deleted mainly in HPV-related tumors that recurred and HPV-negative tumors but infrequently in HPV-related tumors that did not recur. Lastly, MAPK, JAK/STAT, DNA damage repair, and cell death pathways had a similar burden of genomic derangements among HPV-related tumors that recurred and HPV-negative tumors, but a greater frequency of genomic aberrations than HPV-related tumors that did not recur (Figure 5). These data suggest that while HPV-related primary OPSCC tumors that recur share an overall genomic landscape similar to primary OPSCC tumors that do not recur, the tumors that recur acquire specific HPV-negative-like genomic features including *NOTCH1*, *TP53*, *AJUBA*, *KMT2D*, *HLA-A*, and *NFE2L2* gene or copy number alterations as well as mitogenic signaling, DNA damage repair, and cell death pathway alterations.

In addition to our hypothesis regarding similarity of genomic landscapes between HPV-related primary OPSCC tumors that recur and HPV-unrelated HNC, we hypothesized that the primary tumors from recurrent cases would be enriched in genomic alterations involving mitogenic, DNA damage repair, and tumor suppressor signaling pathways. When comparing mutations between the HPV-related tumors that did or did not recur, we observed several genes that were only mutated among the tumors that did not recur such as *B2M*, a gene encoding β2-microglobulun which plays a role in antigen presentation. A recent study evaluating the effectiveness of anti-PD1-antibody therapy across 12 tumor types noted that *B2M* was not mutated in any of the primary tumors from treatment-resistant cases [39]. With respect to immune therapy, we observed a strong contribution of *CD274* [PDL1] to the primary dimension of

variance in a multiple correspondence analysis of genomic aberrations. Interestingly, *CD274* was deleted in 56% of HPV-related OPSCC tumors with *KMT2D* mutations. Three of the five tumors with co-occurring *CD274* deletions and *KMT2D* mutations were tumors that recurred. Lastly, there were several genes mutated only among the HPV-related tumors that recurred. These involved diverse signaling pathways from oxidative stress response (*NFE2L2*) to mitogenic signaling (*TSC2*, *KIT*) to DNA damage repair genes (*BRIP1*, *MLH3*, *NBN*, *POLQ*). PI(3)K [40-42], MAPK [43], Src tyrosine kinase [44,45], DNA damage response [46], TGFβ, Wnt, and Notch pathways [47] have previously been implicated in chemoradiation resistance.

NFE2L2, encoding Nrf2, a critical regulator of oxidative stress, was mutated exclusively among HPV-related tumors with recurrence. *NFE2L2* has been implicated as both a tumor suppressor [48,49] and oncogene [50,51] and may play a role in treatment resistance [52,53]. In the TCGA HNC study, it was frequently mutated among HPV-unrelated tumors but not HPV-related tumors [8]. *NFE2L2* also plays a role in transcription of cytoprotective genes and drug transporter proteins in response to chemical and radiation stress. Previous work illustrated decreased apoptosis and increased cell survival in tumor cells with siRNA-mediated down-regulation of Nrf2 exposed to etoposide and UV/γ radiation [54]. Nrf2 is also implicated in Notch1 signaling. One group found that disruption of Nrf2 impedes hepatocyte regeneration after partial hepatectomy but is restored by reestablishment of Notch1 signaling [55]. Of the *NFE2L2* mutations observed in our study, two of three (E79K and E79Q) occurred in oncogenic hotspots as annotated in cBioPortal. The *NFE2L2* E79K mutation has been observed in multiple tumor types including non-small cell lung cancer and glioma [29]. E79Q has been observed in non-small cell lung cancer and bladder cancer [29] in addition to head and neck cancers [8]. Two of the mutations were in patients who never smoked or had less than or equal to ten years smoking history and one mutation was in a patient who had a greater than a ten year smoking history. These findings point to *NFE2L2* as a potential biomarker and/or therapeutic target.

The current study has limitations to consider. One limitation is our limited sample size, which limits our ability to detect weak or moderate differences between recurrent and non-recurrent HPV-related OPSCC. However, this study adds to the existing literature with a several-fold-greater sample size and is the first to explicitly address the association between mutational and copy number variants with recurrence in HPV-related head and neck OPSCC tumors. These data also contribute to the utility of genomic profiling for informing the clinical outcome of disease progression as well as informing potential actionable molecular targets. Additionally, we used a mixture of FFPE and frozen samples. Investigators have noted high concordance (>90%) in mutational data comparing clinically-derived FFPE and fresh frozen samples [56]. Lastly, there is potential for misclassification of HPV-status given that we used p16 status as a marker. Our study nonetheless reflects clinical practice as p16 status is generally used to infer HPV status in HNC treatment.

Methods

Study Design and Subjects

A cross-sectional study design was used to evaluate the association between genomic aberrations among patients with HPV-related OPSCC and disease progression. Whole exome data from the TCGA head and neck genomic characterization study [8] were downloaded (Figure S1). Whole exome data from UW and UPitt were derived from archival and fresh frozen tissue samples. Patients from UW and UPitt who consented to provide tissue for genomic analyses were collected between 2010 and 2015. Subjects included were over 18 years of age. P16 immunohistochemical staining was used as a surrogate for HPV status in the TCGA, UW, and UPitt data [3]. Subjects with HPV-related OPSCC primary tumors were included if their recurrence status was known. TCGA HPV-unrelated OCSCC/OPSCC tumor whole exome data were downloaded for comparison to HPV-related OPSCC tumors that did or did not recur to test our

primary study hypothesis. Additional details regarding study-specific variables used for subject inclusion are found in the **Supplementary Methods**.

Data Collection

Clinical Data

Clinical data from TCGA were downloaded from cBioPortal (http://www.cbioportal.org/index.do) [57,58]. Study investigators for the current study abstracted clinical data from medical records for the UW and UPitt patients. OPSCC includes subsites as defined in the AJCC Cancer Staging Manual, 7th Edition [59].

Molecular Data

Somatic mutation data in the form of mutation annotation format (MAF) files from the TCGA study were obtained through FireBrowse (**Supplementary Methods**). Copy number data for TCGA subjects were inferred through secondary analyses of available controlled-access tumor and matched normal normal binary alignment/map (BAM) files obtained from the National Cancer Institute Genomic Data Commons legacy portal (https://portal.gdc.cancer.gov/legacy-archive/search/f) (**Supplementary Methods**). The following pertain to laboratory processing, sequencing, and analyses of the UW and UPitt tumor and normal (tissue or blood) samples. UW and UPitt frozen tissue specimens were stored in the Seattle Translational Tumor Research head and neck cancer biorepository at the FHCRC. H&E slides were prepared for each specimen and reviewed by a pathologist (E.Q.K.) at UW. Samples with at least 40% tumor content were used for this analysis. Up to 5 mg pieces of frozen tissue were used for DNA extraction. Tissue specimens from UW and UPitt were fixed in 10% neutral buffered formalin, dehydrated in ethanol and embedded with paraffin wax (FFPE). H&E slides were prepared, reviewed by a pathologist, and areas with high tumor density marked. Two-mm punch biopsies corresponding with the areas of high tumor density were taken from the FFPE tissue samples. FFPE tissue was de-paraffinized with xylenes, washed in consecutive ethanol rinses (100% and

70%), and heated to remove formalin cross-linking using a protocol adapted from Shi et al. 2004 [60]. Tumor DNA was extracted using an AllPrep DNA/RNA Micro Kit (Qiagen, Valencia, CA). Double-stranded DNA was quantified using an Invitrogen Quant-iT™ PicoGreen® dsDNA assay kit (Thermo Fisher Scientific, Rockford, IL). Sample integrity was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) to assess the distribution of nucleotide fragment lengths.

Primary Exposures

The primary exposures evaluated were somatic mutation and copy number variants. We investigated specific pathways that are commonly implicated in HPV-unrelated disease (i.e., presumably associated with the relatively poor prognosis) including tumor suppressor, mitogenic signaling, cell cycle control, cell death, DNA damage response, differentiation pathways. Pathway assignments were performed using previously described assignments by Vogelstein and others as well as knowledge of the functional classification of major signaling molecules in the pathway (**Supplemental Table S13**) [8,25,61]. The number of mutational and/or copy number hits were summed for each tumor. To evaluate the association between genomic alterations with recurrence, we categorized somatic mutation and/or copy number variant status for each gene and each sample.

Key Covariates

Key covariates evaluated included age, stage, sex, tobacco use history, and treatment modality. Age was categorized by median split. Stage was defined as a binary variable using a contemporary staging system for HPV-related OPSCC [17]. Tobacco use history was defined as either positive or negative due to the absence of information on more detailed measures (e.g., pack-years) in the medical records.

Sequencing and Alignment

Sequencing and alignment for TCGA head and neck tumor data were previously described [8,16]. UW and UPitt OPSCC tumor sample data were prepared for sequencing as follows. Libraries were prepared via standard protocols using the Agilent SureSelect Human All Exon v6 kit on a PerkinElmer Sciclone G3 NGSx workstation and sequenced on an Illumina HiSeq 2500 sequencing system in high output mode (100 bp, paired-end). Quality control was performed on the raw reads using FastQC (https://goo.gl/cyJgXF) and adapters were removed. Pre-processed reads were aligned to the human genome reference sequence National Center for Biotechnology Information (NCBI, Bethesda, MD) Build 37/hg19 with the Burrows-Wheeler Aligner [62]. Picard (http://broadinstitute.github.io/picard/) software was used to mark and remove PCR duplicates. Base quality score recalibration was performed with the GATK base quality score recalibration pipeline [63]. Variant discovery analysis was performed as suggested by the Broad Institute's Genome Analysis Toolkit (GATK) "Best Practices" [63].

SNV/Indel Analysis

SNV and small insertions or deletions (indel) detection was performed with GATK MuTect2 using default filters (https://goo.gl/nYyrhz) [63,64]. All detected variants were functionally annotated with ANNOVAR [65]. Variant call format (VCF) files were converted to mutation annotation format (MAF) files for downstream analyses using vcf2maf (https://github.com/mskcc/vcf2maf). Further molecular filtering parameters as described in the **Supplementary Methods** and Figure S2 were applied to all TCGA, UW, and UPitt somatic variants to further minimize potential false positives, germline variants, and problematic variants (e.g., *TTN*). Manual review of sequencing alignments using Integrative Genome Viewer [66] was used to confirm mutations in five random nonsynonymous variants and one or more of *PIK3CA*, *FGFR3*, *TP53*, *CDKN2A*, *RB1*, *TSC2*, *ZNF750*, and/or *DDR2* variants for five patients (UPHN11A, UPHN2A, UPHN8A, UWHN14A,

and UWHN 8) with hypermutated samples (>500 SNVs). Filtered MAF files including the TCGA, UW, and UPitt data were merged into one MAF file for downstream analyses. An additional MAF file was generated from the aforementioned dataset by limiting variants to those affecting a list of 467 putative tumorigenesis driver genes as described by Vogelstein et al. [25], Hedberg et al. [26], and the UW OncoPlex assay [27] (**Supplemental Table S3**).

To detect likely driver mutations across the entire study population, we generated MAF files for all samples using the datasets with all filtered variants and variants among the 467 driver genes. These MAF files were processed with MutSigCV v1.2 [67] using default settings. MAF files, limited to the driver gene set, were used to analyze the mutational profile across samples and between the groups with or without recurrence using functionalities of the "maftools" [68] R/bioconductoR package.

CNV Analysis

Somatic copy number variant (CNV) detection was performed using Aberration Detection in Tumor Exome (ADTEx) [69]. ADTEx incorporates a functionality to normalize sequencing coverage data for tumor and normal samples. Coverage files for input to ADTEx were prepared from BAM files using bedtools 2.21.0 [70]. TCGA hg19-aligned BAM files were imported from the NCI Legacy Genomic Data Portal (https://portal.gdc.cancer.gov/legacy-archive/search/f). Of the 37 TCGA mutation data files analyzed, 32 paired tumor and normal BAM files were available for copy number analysis. BAM files from 21 paired primary and normal tissue samples were available from UW and UPitt for a total dataset of 53 primary tumors. Analyses were performed using in-house bash scripts. ADTEx also implements a segmentation feature using the R/bioconductoR package, "DNAcopy" [71]. Segmented copy number data were log2 transformed using R statistical programming software [72]. Segmentation and marker files were prepared and processed with GISTIC2.0 v6.10 [73] to assess for significant CNVs across tumors. Input parameters are described in the **Supplementary Methods**. TCGA CNV data were compared

between platforms to assess for variance introduced by the method of CNV analysis. The TCGA publication [8] utilized a SNP array to infer copy number. We utilized WES data to infer copy number for the TCGA, UW, and UPitt data as described above. We did not observe a difference in copy number alterations based on the analytical technique (Supplemental Fig. S5). We also assessed for batch effect by study site among the HPV-related samples given the different sequencing platforms used in whole exome sequencing between the TCGA data and UW/UPitt samples. We did not observe a difference in copy number data by study site although certain samples contributed more than others to the variance (Supplemental Fig. S6).

Combined Somatic Mutation and Copy Number Genomic Analyses

SNV and CNV were combined to evaluate the frequency of genomic aberrations between HPV-related primary OPSCC tumors that did or did not recur to TCGA HPV-unrelated OCSCC/OPSCC tumor samples. Using GISTIC-thresholded copy number variant calls by gene, we categorized genomic aberrations labeled by GISTIC as "2" as an amplification and GISTIC category "-1" or "-2" as a copy number loss. We merged these data with the mutation data. Thus, for each sample and gene combination, genomic aberrations were categorized as "0" or "1" (absent/present). An aberration was defined as the presence of either a nonsynonymous mutation, high-level amplification (as categorized by GISTIC), or copy number loss (as categorized by GISTIC). The merged SNV and CNV data were used for multiple correspondence analysis which was performed using the R package, "FactoMineR" [74]. The genomic aberrations included in the analysis were those occurring in a set of 41 HPV-negative-like genes (**Supplemental Table S11**) that were significantly affected by SNVs or CNVs in the TCGA head and neck data [8]. The integrated SNV and CNV data were also categorized for pathway analyses limited to variants affecting genes within pathways of interest including mitogenic pathways (PI3K, JAK/STAT, MAPK, RTK), DNA damage response, cell cycle, cell death, and differentiation pathways (**Supplemental Table S13**).

Statistical Analyses

R programming software [72] was used to perform statistical analyses. Prevalence ratios were estimated with generalized linear models using a Poisson distribution with log-link to test the association between pathway hits and gene mutations and/or copy number variants with recurrence. P-values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg correction and adjusted p-values (q-value) of 0.1 and 0.25 were used to determine significance for the mutational analyses and copy number analyses, respectively, as previously described [8].

References

1. McQuillan G, Kruszon-Moran D, Markowitz LE, Unger ER, Paulose-Ram R. Prevalence of HPV in Adults Aged 18-69: United States, 2011-2014. *NCHS Data Brief.* 2017(280):1-8.
2. Chaturvedi AK, Engels EA, Pfeiffer RM, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol.* 2011;29(32):4294-4301.
3. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med.* 2010;363(1):24-35.
4. Lin BM, Wang H, D'Souza G, et al. Long-term prognosis and risk factors among patients with HPV-associated oropharyngeal squamous cell carcinoma. *Cancer.* 2013;119(19):3462-3471.
5. Fakhry C, Zhang Q, Nguyen-Tan PF, et al. Human papillomavirus and overall survival after progression of oropharyngeal squamous cell carcinoma. *J Clin Oncol.* 2014;32(30):3365-3373.
6. Kelly JR, Husain ZA, Burtness B. Treatment de-intensification strategies for head and neck cancer. *Eur J Cancer.* 2016;68:125-133.
7. Stransky N, Egloff AM, Tward AD, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science.* 2011;333(6046):1157-1160.
8. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature.* 2015;517(7536):576-582.
9. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell.* 1990;63(6):1129-1136.
10. Werness BA, Levine AJ, Howley PM. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science.* 1990;248(4951):76-79.
11. Boyer SN, Wazer DE, Band V. E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. *Cancer Res.* 1996;56(20):4620-4624.
12. Wazer DE, Liu XL, Chu Q, Gao Q, Band V. Immortalization of distinct human mammary epithelial cell types by human papilloma virus 16 E6 or E7. *Proc Natl Acad Sci U S A.* 1995;92(9):3687-3691.
13. Chen JJ. Genomic instability induced by human papillomavirus oncogenes. *N Am J Med Sci.* 2010;3(2):43-47.

14. Mesri EA, Feitelson MA, Munger K. Human viral oncogenesis: a cancer hallmarks analysis. *Cell Host Microbe.* 2014;15(3):266-282.
15. Walline HM, Komarck CM, McHugh JB, et al. Genomic integration of high-risk HPV alters gene expression in oropharyngeal squamous cell carcinoma. *Mol Cancer Res.* 2016;14(10):941-952.
16. Morris LG, Chandramohan R, West L, et al. The molecular landscape of recurrent and metastatic head and neck cancers: insights from a precision oncology sequencing platform. *JAMA Oncol.* 2016.
17. Dahlstrom KR, Garden AS, William WN, Lim MY, Sturgis EM. Proposed staging system for patients with HPV-related oropharyngeal cancer based on nasopharyngeal cancer N categories. *J Clin Oncol.* 2016;34(16):1848-1854.
18. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature.* 2013;502(7471):333-339.
19. Graziani A, Gramaglia D, dalla Zonca P, Comoglio PM. Hepatocyte growth factor/scatter factor stimulates the Ras-guanine nucleotide exchanger. *J Biol Chem.* 1993;268(13):9165-9168.
20. Xiao GH, Jeffers M, Bellacosa A, Mitsuuchi Y, Vande Woude GF, Testa JR. Anti-apoptotic signaling by hepatocyte growth factor/Met via the phosphatidylinositol 3-kinase/Akt and mitogen-activated protein kinase pathways. *Proc Natl Acad Sci U S A.* 2001;98(1):247-252.
21. Zhang YW, Wang LM, Jove R, Vande Woude GF. Requirement of Stat3 signaling for HGF/SF-Met mediated tumorigenesis. *Oncogene.* 2002;21(2):217-226.
22. Boccaccio C, Andò M, Tamagnone L, et al. Induction of epithelial tubules by growth factor HGF depends on the STAT pathway. *Nature.* 1998;391(6664):285-288.
23. Purcell R, Childs M, Maibach R, et al. HGF/c-Met related activation of β-catenin in hepatoblastoma. *J Exp Clin Cancer Res.* 2011;30:96.
24. Gude NA, Emmanuel G, Wu W, et al. Activation of Notch-mediated protective signaling in the myocardium. *Circ Res.* 2008;102(9):1025-1035.
25. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science.* 2013;339(6127):1546-1558.
26. Hedberg ML, Goh G, Chiosea SI, et al. Genetic landscape of metastatic and recurrent head and neck squamous cell carcinoma. *J Clin Invest.* 2016;126(1):169-180.
27. Pritchard CC, Salipante SJ, Koehler K, et al. Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens. *J Mol Diagn.* 2014;16(1):56-67.
28. Oswald F, Rodriguez P, Giaimo BD, et al. A phospho-dependent mechanism involving NCoR and KMT2D controls a permissive chromatin state at Notch target genes. *Nucleic Acids Res.* 2016;44(10):4703-4720.
29. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med.* 2017;23(6):703-713.
30. Forbes SA, Beare D, Boutselakis H, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 2017;45(D1):D777-D783.
31. Agrawal N, Frederick MJ, Pickering CR, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science.* 2011;333(6046):1154-1157.
32. Lui VW, Hedberg ML, Li H, et al. Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. *Cancer Discov.* 2013;3(7):761-769.
33. Wang NJ, Sanborn Z, Arnett KL, et al. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. *Proc Natl Acad Sci U S A.* 2011;108(43):17761-17766.
34. Sen GL, Boxer LD, Webster DE, et al. ZNF750 is a p63 target gene that induces KLF4 to drive terminal epidermal differentiation. *Dev Cell.* 2012;22(3):669-677.

35. Ortega-Molina A, Boss IW, Canela A, et al. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat Med.* 2015;21(10):1199-1208.
36. Bögershausen N, Tsai IC, Pohl E, et al. RAP1-mediated MEK/ERK pathway defects in Kabuki syndrome. *J Clin Invest.* 2015;125(9):3585-3599.
37. Toska E, Osmanbeyoglu HU, Castel P, et al. PI3K pathway regulates ER-dependent transcription in breast cancer through the epigenetic regulator KMT2D. *Science.* 2017;355(6331):1324-1330.
38. Henderson S, Chakravarthy A, Su X, Boshoff C, Fenton TR. APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell Rep.* 2014;7(6):1833-1841.
39. Le DT, Durham JN, Smith KN, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science.* 2017;357(6349):409-413.
40. Xia S, Zhao Y, Yu S, Zhang M. Activated PI3K/Akt/COX-2 pathway induces resistance to radiation in human cervical cancer HeLa cells. *Cancer Biother Radiopharm.* 2010;25(3):317-323.
41. Yu HG, Ai YW, Yu LL, et al. Phosphoinositide 3-kinase/Akt pathway plays an important role in chemoresistance of gastric cancer cells against etoposide and doxorubicin induced cell death. *Int J Cancer.* 2008;122(2):433-443.
42. Jin W, Wu L, Liang K, Liu B, Lu Y, Fan Z. Roles of the PI-3K and MEK pathways in Ras-mediated chemoresistance in breast cancer cells. *Br J Cancer.* 2003;89(1):185-191.
43. Menendez JA, Vellon L, Mehmi I, Teng PK, Griggs DW, Lupu R. A novel CYR61-triggered 'CYR61-alphavbeta3 integrin loop' regulates breast cancer cell survival and chemosensitivity through activation of ERK1/ERK2 MAPK signaling pathway. *Oncogene.* 2005;24(5):761-779.
44. Ischenko I, Camaj P, Seeliger H, et al. Inhibition of Src tyrosine kinase reverts chemoresistance toward 5-fluorouracil in human pancreatic carcinoma cells: an involvement of epidermal growth factor receptor signaling. *Oncogene.* 2008;27(57):7212-7222.
45. Hawthorne VS, Huang WC, Neal CL, Tseng LM, Hung MC, Yu D. ErbB2-mediated Src and signal transducer and activator of transcription 3 activation leads to transcriptional up-regulation of p21Cip1 and chemoresistance in breast cancer cells. *Mol Cancer Res.* 2009;7(4):592-600.
46. Bouwman P, Jonkers J. The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance. *Nat Rev Cancer.* 2012;12(9):587-598.
47. Steg AD, Bevis KS, Katre AA, et al. Stem cell pathways contribute to clinical chemoresistance in ovarian cancer. *Clin Cancer Res.* 2012;18(3):869-881.
48. Funes JM, Henderson S, Kaufman R, et al. Oncogenic transformation of mesenchymal stem cells decreases Nrf2 expression favoring in vivo tumor growth and poorer survival. *Mol Cancer.* 2014;13:20.
49. Wolf B, Goebel G, Hackl H, Fiegl H. Reduced mRNA expression levels of NFE2L2 are associated with poor outcome in breast cancer patients. *BMC Cancer.* 2016;16(1):821.
50. Stacy DR, Ely K, Massion PP, et al. Increased expression of nuclear factor E2 p45-related factor 2 (NRF2) in head and neck squamous cell carcinomas. *Head Neck.* 2006;28(9):813-818.
51. DeNicola GM, Karreth FA, Humpton TJ, et al. Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature.* 2011;475(7354):106-109.
52. Homma S, Ishii Y, Morishima Y, et al. Nrf2 enhances cell proliferation and resistance to anticancer drugs in human lung cancer. *Clin Cancer Res.* 2009;15(10):3423-3432.
53. Jiang T, Chen N, Zhao F, et al. High levels of Nrf2 determine chemoresistance in type II endometrial cancer. *Cancer Res.* 2010;70(13):5486-5496.
54. Niture SK, Jaiswal AK. Nrf2 protein up-regulates antiapoptotic protein Bcl-2 and prevents cellular apoptosis. *J Biol Chem.* 2012;287(13):9873-9886.

55. Wakabayashi N, Shin S, Slocum SL, et al. Regulation of notch1 signaling by nrf2: implications for tissue regeneration. *Sci Signal.* 2010;3(130):ra52.
56. Van Allen EM, Wagle N, Stojanov P, et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med.* 2014;20(6):682-688.
57. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401-404.
58. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013;6(269):pl1.
59. Edge S, Byrd D, Compton C, Fritz A, Greene F, Trotti A. *AJCC Cancer Staging Manual, 7th Edition.* France: Springer; 2010.
60. Shi SR, Datar R, Liu C, et al. DNA extraction from archival formalin-fixed, paraffin-embedded tissues: heat-induced retrieval in alkaline solution. *Histochem Cell Biol.* 2004;122(3):211-218.
61. Chung CH, Guthrie VB, Masica DL, et al. Genomic alterations in head and neck squamous cell carcinoma determined by cancer gene-targeted sequencing. *Ann Oncol.* 2015;26(6):1216-1223.
62. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England).* 2009;25(14):1754-1760.
63. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.
64. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* 2013;31(3):213-219.
65. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
66. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29(1):24-26.
67. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499(7457):214-218.
68. Mayakonda A, Koeffler HP. Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. *bioRxiv.* 2016.
69. Amarasinghe KC, Li J, Hunter SM, et al. Inferring copy number and genotype in tumour exome data. *BMC Genomics.* 2014;15:732.
70. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics.* 2014;47:11.12.11-34.
71. Seshan V, Olshen A. DNAcopy: DNA copy number data analysis. *R package version 1.50.1.* 2017.
72. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
73. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;12(4):R41.
74. Le S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. *J Stat Softw.* 2008;25(1):1-18.

Tables and Table Legends

Table 1

Table 1. Baseline characteristics of patients with primary oropharyngeal squamous cell carcinoma, by recurrence status.

	No Recurrence N=39		With Recurrence N=27	
Characteristic	**n**	**(%)**	**n**	**(%)**
Age (years)				
35 - 55	15	(38)	8	(30)
55 – 77	24	(62)	19	(70)
Sex				
Female	4	(10)	3	(11)
Male	35	(90)	24	(89)
Stage[1]				
IA	12	(31)	5	(19)
IB	17	(44)	15	(56)
II	4	(10)	5	(19)
III	6	(15)	2	(7)
Treatment				
CRT[2]	14	(36)	15	(56)
Surgery alone	7	(18)	6	(22)
Surgery with adjuvant radiation	6	(15)	2	(7)
Unknown	12	(31)	4	(15)
History of Tobacco Use				
Never or ≤10 years use	21	(54)	19	(70)
>10 years	18	(46)	8	(30)
Sample Source				
UW	10	(26)	4	(15)
UPitt	0	(0)	16	(59)
TCGA	29	(74)	7	(26)
Median Disease-Free Follow-Up (Months)	20.9		16.0	

[1]Clinical staging at diagnosis using a contemporary system for HPV-associated oropharyngeal squamous cell carcinoma (Dahlstrom et al. 2016). In this system, stage IA and IB encompass N0 through N2 disease with the main difference being tumor size categorized by T1 versus T2 tumors, respectively. Stage II captures T3 and N3 tumors and stage IV is reserved for metastatic disease.

[2]CRT, chemoradiotherapy. This includes primary radiotherapy with concurrent chemotherapy as well as concurrent chemoradiotherapy in the adjuvant setting.

Table 2

Table 2. Univariable and multivariable prevalence ratios for the association between critical tumor biological pathways and recurrence among HPV-related OPSCC tumors.

Biological Pathway	Univariable PR[a] (95% CI)	p-value	Adjusted p-value	Multivariable Variable PR[b] (95% CI)	p-value	Adjusted p-value
PI3K	1.08 (1.01 – 1.16)	0.028	0.045	0.92 (0.76 – 1.12)	0.42	0.48
JAK/STAT	1.12 (1.02 – 1.24)	0.017	0.045	1.17 (0.85 – 1.61)	0.35	0.46
MAPK	1.24 (1.03 – 1.50)	0.021	0.045	1.23 (0.89 – 1.70)	0.21	0.46
DNA Damage Response	1.27 (1.01 – 1.59)	0.039	0.045	1.22 (0.88 – 1.68)	0.22	0.46
Cell Cycle	1.07 (1.00 – 1.15)	0.039	0.045	0.90 (0.75 – 1.09)	0.28	0.46
Cell Death	1.08 (1.01 – 1.14)	0.019	0.045	1.10 (0.98 – 1.22)	0.10	0.46
RTK	1.13 (0.89 – 1.45)	0.31	0.31	0.84 (0.61 – 1.16)	0.29	0.46
Differentiation	1.25 (1.02 – 1.54)	0.032	0.045	1.0 (0.70 – 1.43)	0.99	0.99

[a]Ratio per each additional mutation in a pathway.

[b]Model including all pathways as explanatory continuous variables. Other factors (age, smoking, and gender) were not included in the multi~~ple~~ variable estimates.

Table 3

Table 3. Univariable and multivariable prevalence ratios for the association between gene mutations and/or copy number variants with recurrence among HPV-related OPSCC tumors.

Gene	Univariable **PR[a] (95% CI for PR)**	**p-value**	**Adjusted p-value**	Multivariable **PR[b] (95% CI for PR)**	**p-value**	**Adjusted p-value**
HLA-A	3.01 (1.68 – 5.41)	<0.001	**0.003**	4.28 (0.78 – 23.5)	0.094	0.28
PEG3	2.31 (1.13 – 4.72)	0.021	0.13	3.10 (1.04 – 9.16)	0.042	0.21
TSC2	2.52 (0.98 – 4.35)	0.057	0.13	1.26 (0.55 – 2.92)	0.59	0.77
NOTCH1	2.44 (0.99 – 4.13)	0.052	0.13	0.92 (0.40 – 2.10)	0.84	0.84
NSD1	2.41 (0.94 – 4.03)	0.074	0.15	2.34 (0.98 – 5.60)	0.057	0.21
NFE2L2	2.03 (0.99 – 4.13)	0.052	0.13	0.77 (0.19 – 3.19)	0.72	0.77
TP53	1.47 (0.70 – 3.10)	0.31	0.43	0.76 (0.32 – 1.85)	0.55	0.77
KMT2D	1.40 (0.60 - 3.26)	0.44	0.54	1.25 (0.38 – 4.06)	0.72	0.77
CASP8	2.03 (0.99 – 4.13)	0.052	0.13	3.83 (1.06 – 13.8)	0.041	0.21
FAT1	1.69 (0.82 – 3.51)	0.16	0.27	2.23 (0.78 – 6.43)	0.14	0.34
FBXW7	1.32 (0.63 – 2.78)	0.47	0.54	0.60 (0.19 – 1.90)	0.38	0.64
PIK3CA	1.47 (0.70 – 3.10)	0.31	0.43	0.56 (0.16 – 1.97)	0.37	0.64
CDKN2A	0.56 (0.09 – 3.40)	0.53	0.57	0.38 (0.07 – 2.07)	0.26	0.57
AJUBA	1.23 (0.51 – 2.94)	0.64	0.64	1.25 (0.37 – 4.23)	0.72	0.77

[a]Ratio per each additional mutation and/or copy number variant for each gene.

[b]Model including all pathways as explanatory continuous variables. Other factors (age, smoking, and gender) were not included in the multiple variable estimates.

Figures and Figure Legends

Figure 1

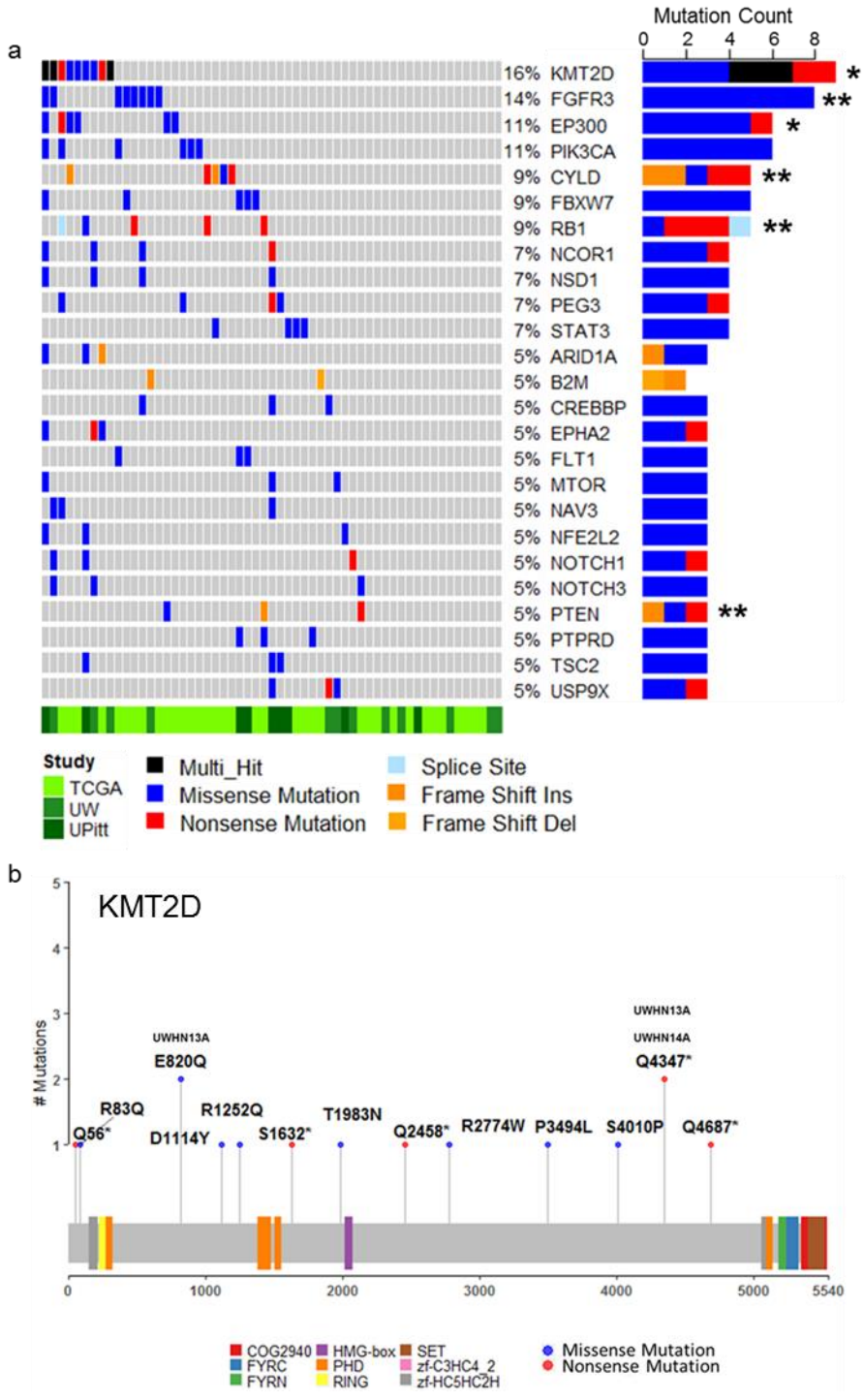


Figure 1 Legend

Figure 1. Frequently mutated genes in OPSCC (n = 57). *a*, Genes mutated in at least 5% of primary OPSCC tumors are shown here. Percentages represent the proportion of samples with a mutation. Each column represents a patient. Rows with significantly mutated genes are annotated with a double asterisk (******, *q-value* <0.1) as identified using the MutSigCV algorithm. Rows with gene mutations trending towards significance are identified using a single asterisk (*****, *q-value* <1). Colored bars represent mutations as described in the legend. Gray bars represent a gene by patient comparison without a mutation. *b*, Amino acid changes associated with respective *KMT2D* mutations in this study population are illustrated. Colored bars represent functional domains of the amino acid sequence. Grey areas in between colored bars represent non-functional domains of the amino acid sequence. Amino acid changes for samples UWHN13 and UWHN14 are highlighted for illustrative purposes.

Figure 2

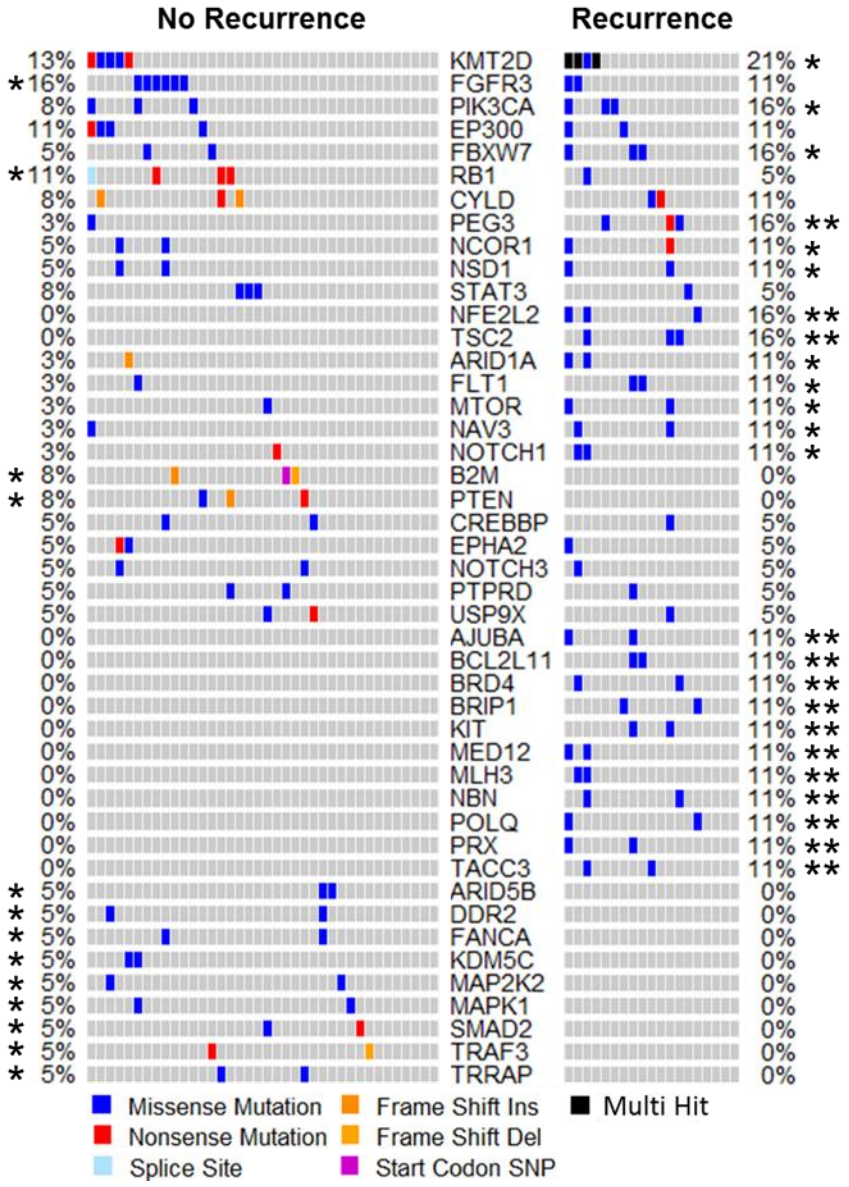
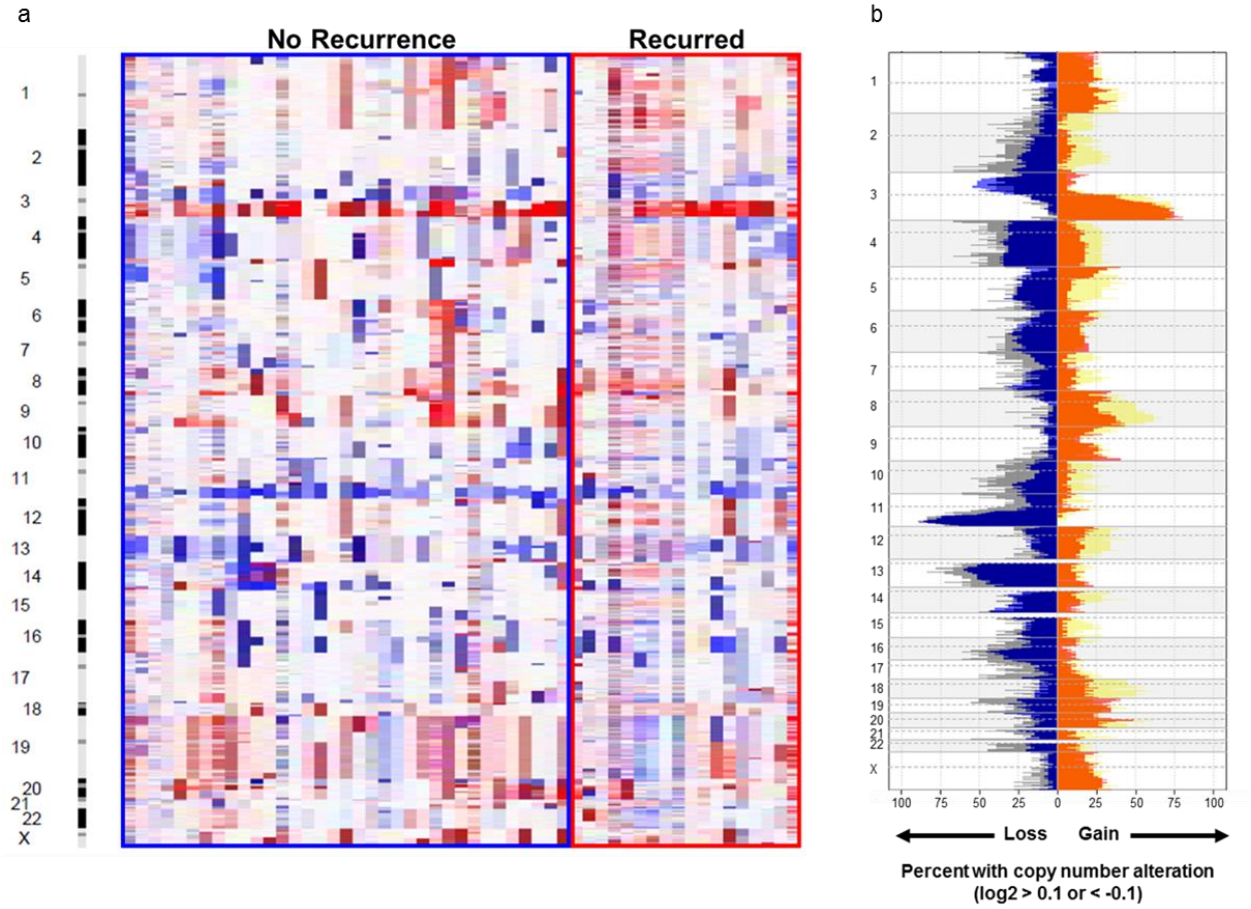

No Recurrence
Recurrence
13% KMT2D 21% *
*16% FGFR3 11%
8% PIK3CA 16% *
11% EP300 11%
5% FBXW7 16% *
*11% RB1 5%
8% CYLD 11%
3% PEG3 16% **
5% NCOR1 11% *
5% NSD1 11% *
8% STAT3 5%
0% NFE2L2 16% **
0% TSC2 16% **
3% ARID1A 11% *
3% FLT1 11% *
3% MTOR 11% *
3% NAV3 11% *
3% NOTCH1 11% *
* 8% B2M 0%
* 8% PTEN 0%
5% CREBBP 5%
5% EPHA2 5%
5% NOTCH3 5%
5% PTPRD 5%
5% USP9X 5%
0% AJUBA 11% **
0% BCL2L11 11% **
0% BRD4 11% **
0% BRIP1 11% **
0% KIT 11% **
0% MED12 11% **
0% MLH3 11% **
0% NBN 11% **
0% POLQ 11% **
0% PRX 11% **
0% TACC3 11% **
* 5% ARID5B 0%
* 5% DDR2 0%
* 5% FANCA 0%
* 5% KDM5C 0%
* 5% MAP2K2 0%
* 5% MAPK1 0%
* 5% SMAD2 0%
* 5% TRAF3 0%
* 5% TRRAP 0%
Missense Mutation
Nonsense Mutation
Splice Site
Frame Shift Ins
Frame Shift Del
Start Codon SNP
Multi Hit

Figure 2 Legend

Figure 2. Mutational differences between OPSCC primary tumors without (n = 38) versus with recurrence (n = 21). *a*, Genes mutated in at least two samples were analyzed via Fisher exact test. Columns represent individual patients. Genes (rows) are sorted by descending frequency across all samples. Rows with frequently mutated genes trending towards a significant difference in frequency between groups are annotated with a double asterisk (******, *q-value* ≤ 0.4) or a single asterisk (*****, *q-value* < 1). Colored bars represent mutations as annotated in the legend, and gray bars represent gene by patient comparisons without a mutation in the given gene.

Figure 3


a
No Recurrence
Recurred
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
b
100
75
50
25
0
25
50
75
100
Loss
Gain
Percent with copy number alteration
(log2 > 0.1 or < -0.1)

c

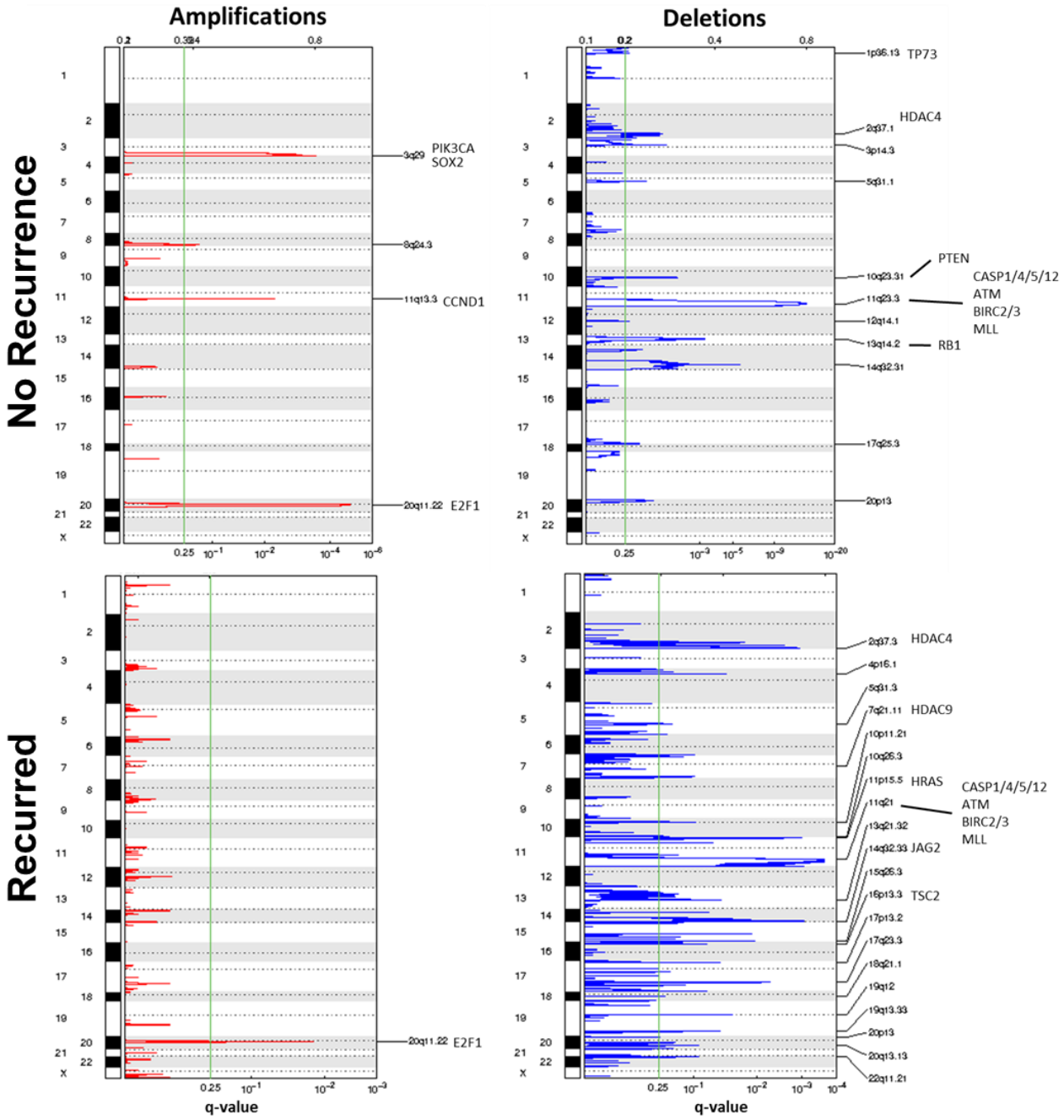

Amplifications
Deletions
No Recurrence
Recurred
PIK3CA
SOX2
CCND1
E2F1
TP73
HDAC4
PTEN
CASP1/4/5/12
ATM
BIRC2/3
MLL
RB1
HDAC9
HRAS
JAG2
TSC2
q-value
q-value

Figure 3 Legend

Figure 3. DNA copy number alterations. *a*, Log2 copy number alterations analyzed using whole exome sequencing data for primary OPSCC tumors without (blue box) vs with recurrence (red box) across samples (n = 53). Rows represent cytobands. Columns represent samples. Gains are represented in shades of red. Log2 copy number ratio equal to one is represented in red with smaller gains in lighter shades of red. Losses are represented by shades of blue. Log2 copy number ratio equal to negative one is represented in blue with smaller losses in lighter shades of blue. *b*, Percent of samples (n = 53) with a copy gain or loss greater than log2 of 0.1 or less than log2 of -0.1 are illustrated to assess low-level deletions and amplification a.. Rows represent cytobands. Copy loss or gain is on the x-axis. Primary tumors without recurrence: *red*, copy number gains; *blue*, copy number losses. Primary tumors from patients who recurred: *yellow*, copy number gains; *black*, copy number losses. *c*, GISTIC2.0 significant amplifications or deletions across tumors that did not recur (*top panel*) or recurred (*bottom panel*). Selected genes are annotated on respective cytobands. *Green line*, significance threshold (*FDR q-value* = 0.25).

Figure 4

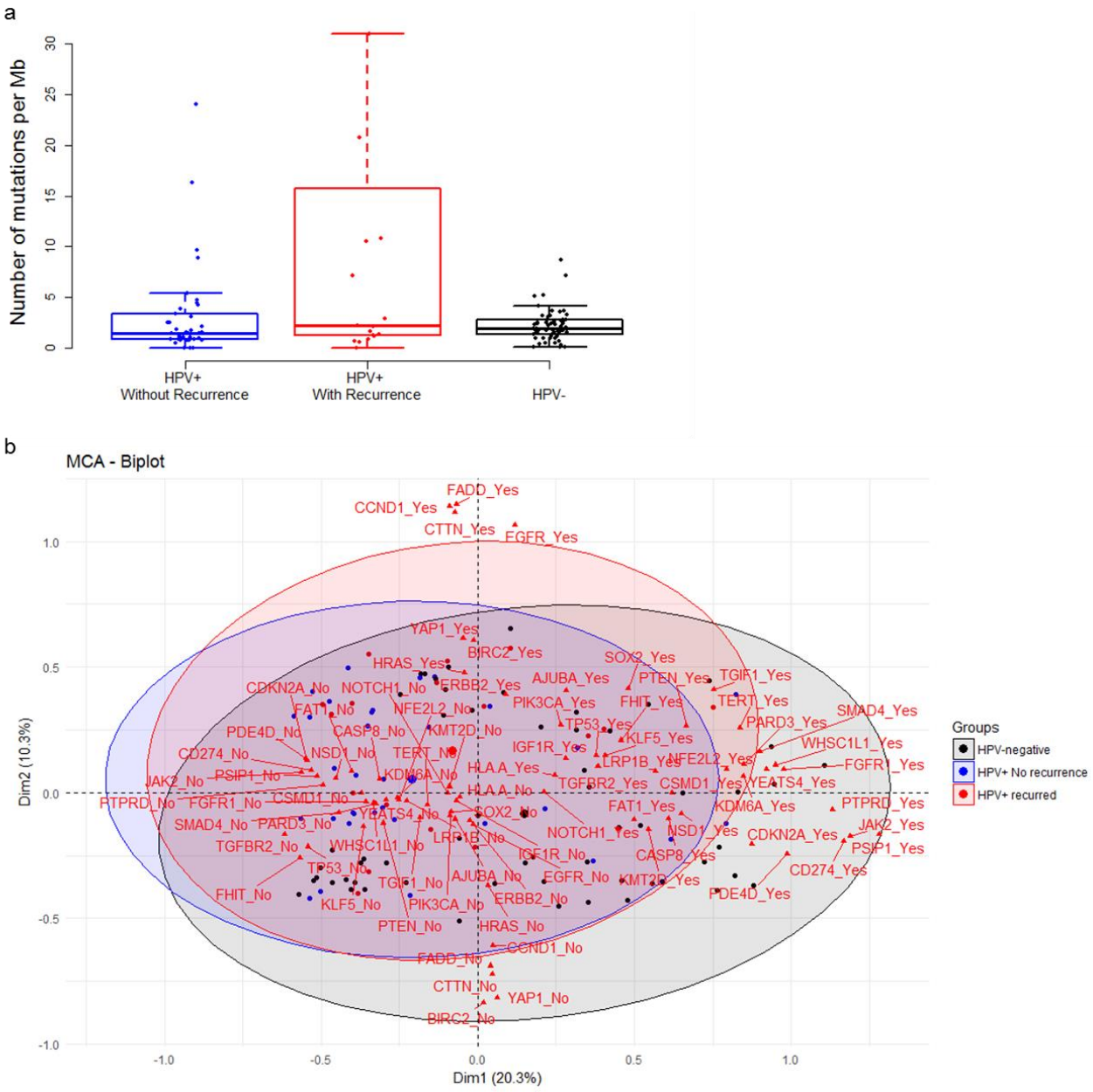

a
Number of mutations per Mb
0
5
10
15
20
25
30
HPV+
Without Recurrence
HPV+
With Recurrence
HPV-
b
MCA - Biplot
Dim2 (10.3%)
Dim1 (20.3%)
1.0
0.5
0.0
-0.5
-1.0
Groups
HPV-negative
HPV+ No recurrence
HPV+ recurred
FADD_Yes
CCND1_Yes
CTTN_Yes
EGFR_Yes
YAP1_Yes
BIRC2_Yes
HRAS_Yes
SOX2_Yes
AJUBA_Yes
PTEN_Yes
TGIF1_Yes
CDKN2A_No
NOTCH1_No
ERBB2_Yes
TERT_Yes
FAT1_No
NFE2L2_No
PIK3CA_Yes
FHIT_Yes
SMAD4_Yes
PDE4D_No
CASP8_No
KMT2D_No
TP53_Yes
PARD3_Yes
KLF5_Yes
WHSC1L1_Yes
CD274_No
NSD1_No
TERT_No
IGF1R_Yes
LRP1B_Yes
NFE2L2_Yes
FGFR1_Yes
JAK2_No
PSIP1_No
KDM6A_No
HLA.A_Yes
TGFBR2_Yes
CSMD1_Yes
YEATS4_Yes
PTPRD_No
FGFR1_No
CSMD1_No
HLA.A_No
FAT1_Yes
KDM6A_Yes
PTPRD_Yes
SMAD4_No
PARD3_No
YEATS4_No
SOX2_No
NOTCH1_Yes
NSD1_Yes
CDKN2A_Yes
JAK2_Yes
TGFBR2_No
WHSC1L1_No
LRP1B_No
IGF1R_No
CASP8_Yes
PSIP1_Yes
TP53_No
AJUBA_No
EGFR_No
KMT2D_Yes
CD274_Yes
FHIT_No
TGIF1_No
KLF5_No
PIK3CA_No
ERBB2_No
PDE4D_Yes
PTEN_No
HRAS_No
CCND1_No
FADD_No
CTTN_No
YAP1_No
BIRC2_No

Figure 4 Legend

Figure 4. Integrated analysis of genomic aberrations across primary HPV-related OPSCC tumors (n = 53) for which somatic mutation and copy number alteration data were available. *a*, Mutations per Mb (megabase pair) among primary HPV-related OPSCC tumors that did not recur (*blue box* and *points*) or did recur (*red box and points*) versus TCGA HPV-unrelated OCSCC/OPSCC tumors. The median (IQR) mutations per Mb was 1.41 (2.41) and 2.22 (14.5) among the non-recurrent and recurrent cases, respectively. The mutations per Mb was 1.95 (1.41) for HPV-negative OC and OPSCC tumors. Mutations per Mb did not vary between groups (*p-value* = 0.20, Kruskal-Wallis rank sum test). *b*, Multiple correspondence analysis (MCA) evaluating the association between somatic mutation and copy number variants between primary HPV-related OPSCC that did or did not recur compared to TCGA HPV-unrelated OCSCC/OPSCC. MCA coordinates encompassed by 95% confidence ellipses. *HPV-negative*, TCGA HPV-unrelated OCSCC/OPSCC data; *HPV+ No recurrence*, primary HPV-related OPSCC tumors that did not recur. *HPV+ recurred*, primary HPV-related OPSCC tumors that recurred. Analysis considers a list of 41 HPV-negative-like genes based on results from the TCGA head and neck cancer study. A sample that has a mutation and/or copy number aberration is defined as having a genomic aberration for a given gene. Gene names followed by "yes" or "no" represent the influence of either having a genomic aberration or not in one of these genes.

Figure 5

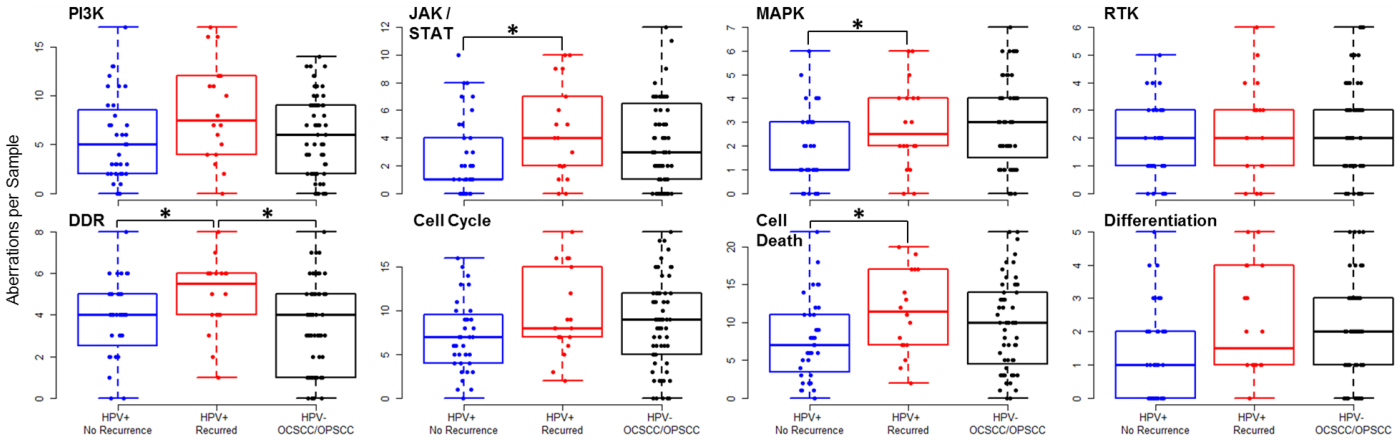

PI3K
JAK / STAT
MAPK
RTK
DDR
Cell Cycle
Cell Death
Differentiation
Aberrations per Sample
HPV+ No Recurrence
HPV+ Recurred
HPV- OCSCC/OPSCC
*

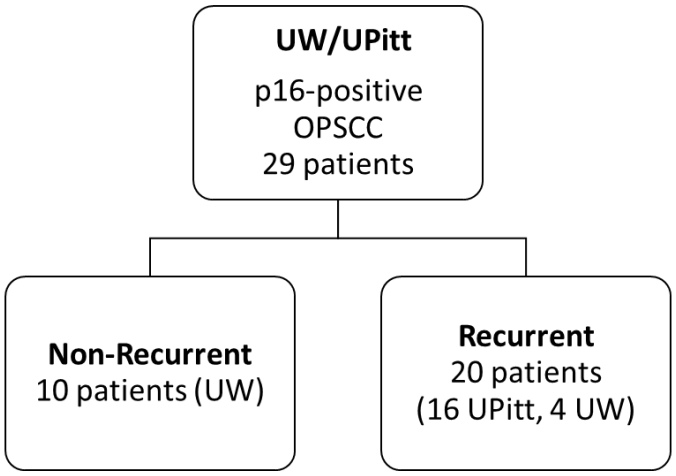
Figure 5 Legend

Figure 5. Somatic mutation and/or copy number alterations per tumor for eight pathways are illustrated across samples (n = 53). Pathway constituents are described in the text. Blue boxes and points represent primary HPV-related OPSCC tumors that did not recur and red boxes and points represent primary HPV-related OPSCC tumors that did recur. TCGA HPV-unrelated OCSCC/OPSCC tumors are represented by black boxes and points. *RTK*, receptor tyrosine kinase. *DDR*, DNA damage repair. Bars with an overlying asterisk represent statistically significant differences between pathways (*FDR q-value* <0.1, Dunn test).

Supplemental Figures and Figure Legends

Figure S1

a



b

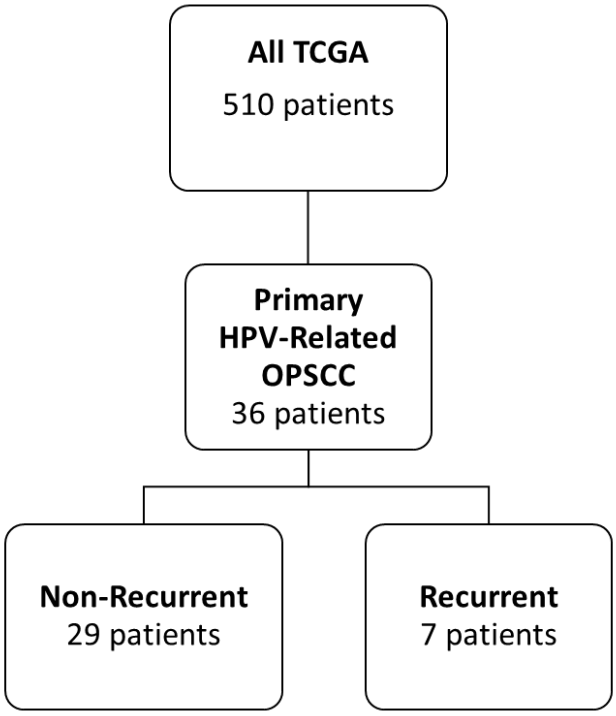


Figure S1 Legend

Figure S1. Study subjects sources. *a*, UW and UPitt primary HPV-related OPSCC tumors that did or did not recur. ([1]UW, University of Washington; UPitt, University of Pittsburgh *) b*, TCGA dataset primary HPV-related OPSCC tumors that did or did not recur.

Figure S2

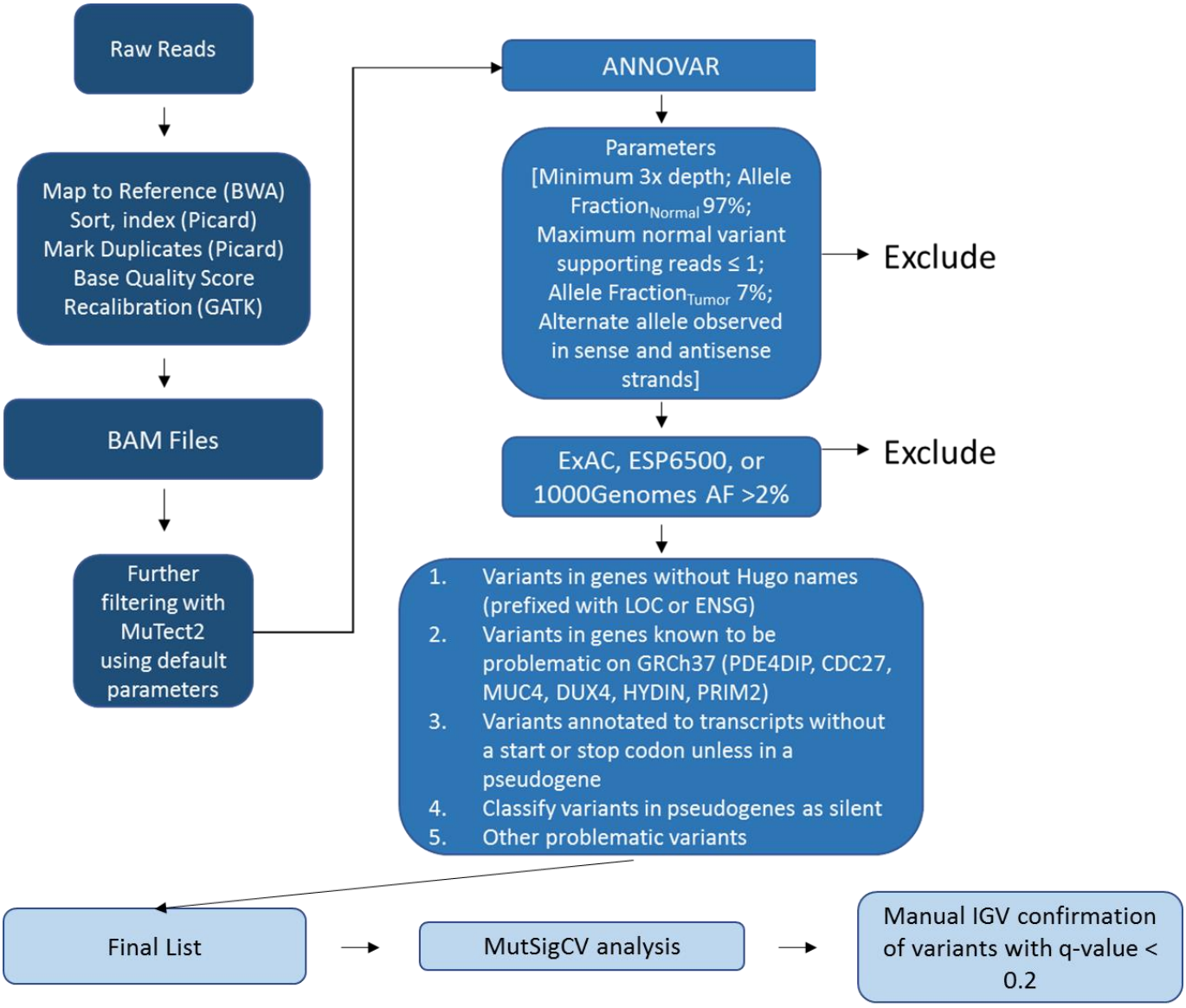

Raw Reads
Map to Reference (BWA)
Sort, index (Picard)
Mark Duplicates (Picard)
Base Quality Score Recalibration (GATK)
BAM Files
Further filtering with MuTect2 using default parameters
ANNOVAR
Parameters [Minimum 3x depth; Allele Fraction$_{Normal}$ 97%; Maximum normal variant supporting reads ≤ 1; Allele Fraction$_{Tumor}$ 7%; Alternate allele observed in sense and antisense strands]
Exclude
ExAC, ESP6500, or 1000Genomes AF >2%
Exclude
1. Variants in genes without Hugo names (prefixed with LOC or ENSG)
2. Variants in genes known to be problematic on GRCh37 (PDE4DIP, CDC27, MUC4, DUX4, HYDIN, PRIM2)
3. Variants annotated to transcripts without a start or stop codon unless in a pseudogene
4. Classify variants in pseudogenes as silent
5. Other problematic variants
Final List
MutSigCV analysis
Manual IGV confirmation of variants with q-value < 0.2

Figure S2 Legend

Figure S2. Mutation data curation methodology. Raw whole exome sequencing reads were generated and filtered to reduce false positive variants such as germline variants. A final list of variants was curated for downstream statistical and epidemiological analyses.

Figure S3

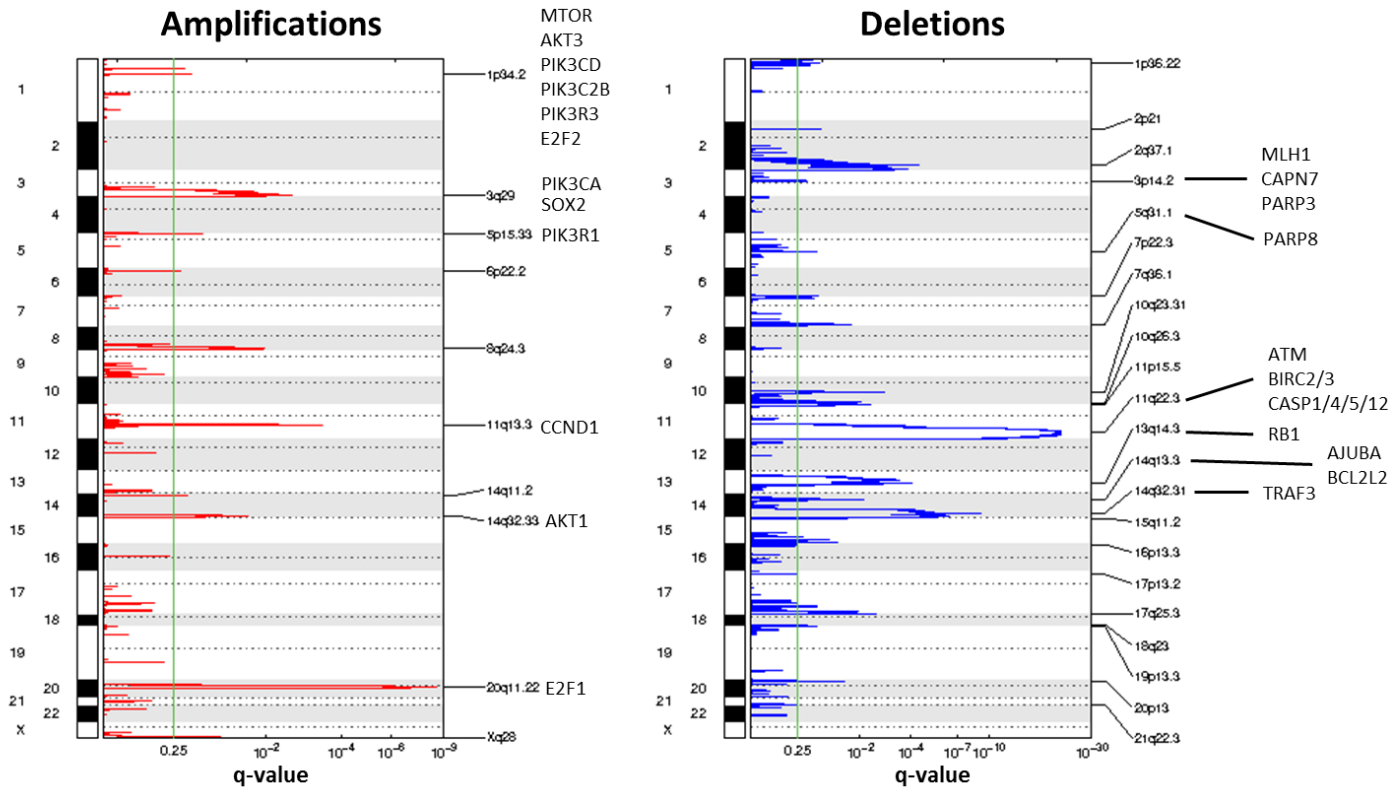

Amplifications
MTOR
AKT3
PIK3CD
PIK3C2B
PIK3R3
E2F2
PIK3CA
SOX2
PIK3R1
CCND1
AKT1
E2F1
1p34.2
3q29
5p15.33
6p22.2
8q24.3
11q13.3
14q11.2
14q32.33
20q11.22
Xq28
0.25
10−2
10−4
10−6
10−9
q-value
Deletions
1p36.22
2p21
2q37.1
3p14.2
5q31.1
7p22.3
7q36.1
10q23.31
10q26.3
11p15.5
11q22.3
13q14.3
14q13.3
14q32.31
15q11.2
16p13.3
17p13.2
17q25.3
18q23
19p13.3
20p13
21q22.3
MLH1
CAPN7
PARP3
PARP8
ATM
BIRC2/3
CASP1/4/5/12
RB1
AJUBA
BCL2L2
TRAF3
0.25
10−2
10−4
10−7
10−10
10−30
q-value

Figure S3 Legend

Figure S3. Significant amplifications or deletions across all primary HPV-related OPSCC tumors. Significant amplifications or deletions were detected using GISTIC2.0 and are depicted on the left or right, respectively. Selected genes are annotated on respective cytobands. *Green line*, significance threshold (*FDR q-value* = 0.25).

Figure S4

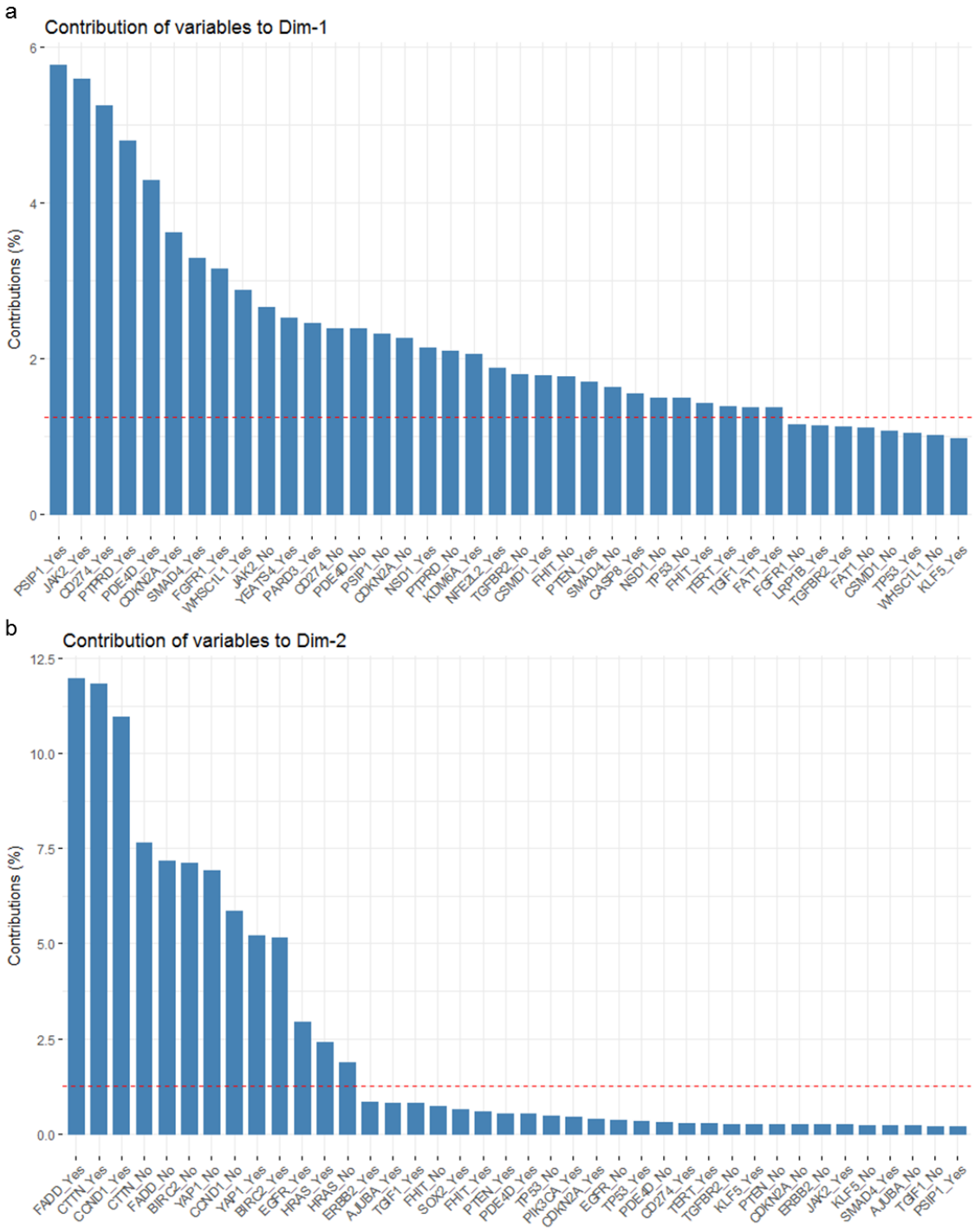

a
Contribution of variables to Dim-1
Contributions (%)
0
2
4
6
PSIP1_Yes
JAK2_Yes
CD274_Yes
PTPRD_Yes
PDE4D_Yes
CDKN2A_Yes
SMAD4_Yes
FGFR1_Yes
WHSC1L1_Yes
JAK2_No
YEATS4_Yes
PARD3_Yes
CD274_No
PDE4D_No
PSIP1_No
CDKN2A_No
NSD1_Yes
PTPRD_No
KDM6A_Yes
NFE2L2_Yes
TGFBR2_No
CSMD1_Yes
FHIT_No
PTEN_Yes
SMAD4_No
CASP8_Yes
NSD1_No
TP53_No
FHIT_Yes
TERT_Yes
TGFI_Yes
FAT1_Yes
FGFR1_No
LRP1B_Yes
TGFBR2_Yes
FAT1_No
CSMD1_No
TP53_Yes
WHSC1L1_No
KLF5_Yes
b
Contribution of variables to Dim-2
Contributions (%)
0.0
2.5
5.0
7.5
10.0
12.5
FADD_Yes
CTTN_Yes
CCND1_Yes
CTTN_No
FADD_No
BIRC2_No
YAP1_No
CCND1_No
YAP1_Yes
BIRC2_Yes
EGFR_Yes
HRAS_Yes
HRAS_No
ERBB2_Yes
AJUBA_Yes
TGFI_Yes
FHIT_No
SOX2_Yes
FHIT_Yes
PTEN_Yes
PDE4D_Yes
TP53_No
PIK3CA_Yes
CDKN2A_Yes
EGFR_No
TP53_Yes
PDE4D_No
CD274_Yes
TERT_Yes
TGFBR2_No
KLF5_Yes
PTEN_No
CDKN2A_No
ERBB2_No
JAK2_Yes
KLF5_No
SMAD4_Yes
AJUBA_No
TGFI_No
PSIP1_Yes

Figure S4 Legend

Figure S4. Contribution of variance to first and second dimensions of multiple correspondence analysis (MCA). Percent contribution to variance encompassed by the first (*a*) and second (*b*) dimensions of the MCA plot comparing genomic aberrations between primary HPV-related OPSCC tumors from cases that did or did not recur to HPV-unrelated OCSCC/OPSCC. Dashed red line at 1.25% represents the expected average contribution if the variable categories were uniform (1/number of categories = 1/80 = 1.25%). Gene names followed by “yes” or “no” represent the influence of either having a genomic aberration or not in one of these genes.

Figure S5

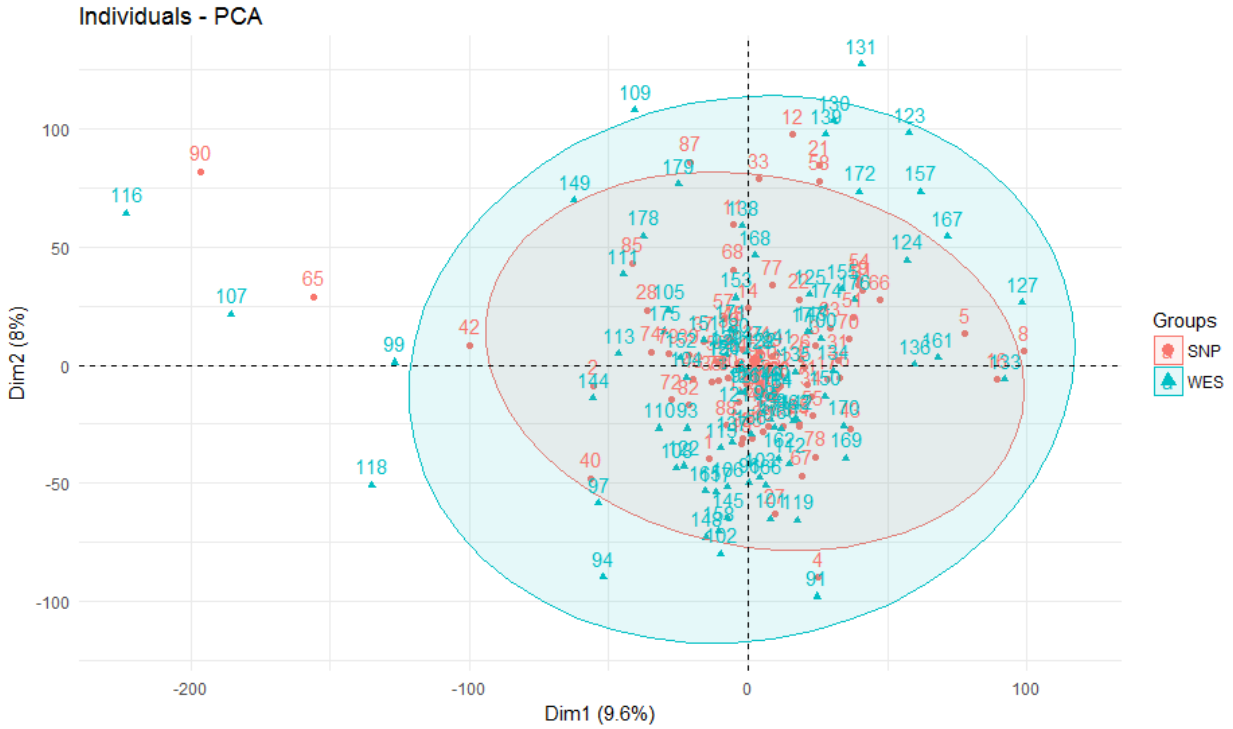

Individuals - PCA
Dim2 (8%)
Dim1 (9.6%)
100
50
0
-50
-100
-200
-100
0
100
Groups
SNP
WES

Figure S5 Legend

Figure S5. Principal component analysis (PCA) plot of TCGA copy number data including HPV-unrelated OCSCC/OPSCC and HPV-related OPSCC samples. Ellipses represent 95% confidence intervals. *SNP*, TCGA copy number data generated using data from a SNP array. *WES*, TCGA copy number data generated using data from whole exome sequencing results.

Figure S6

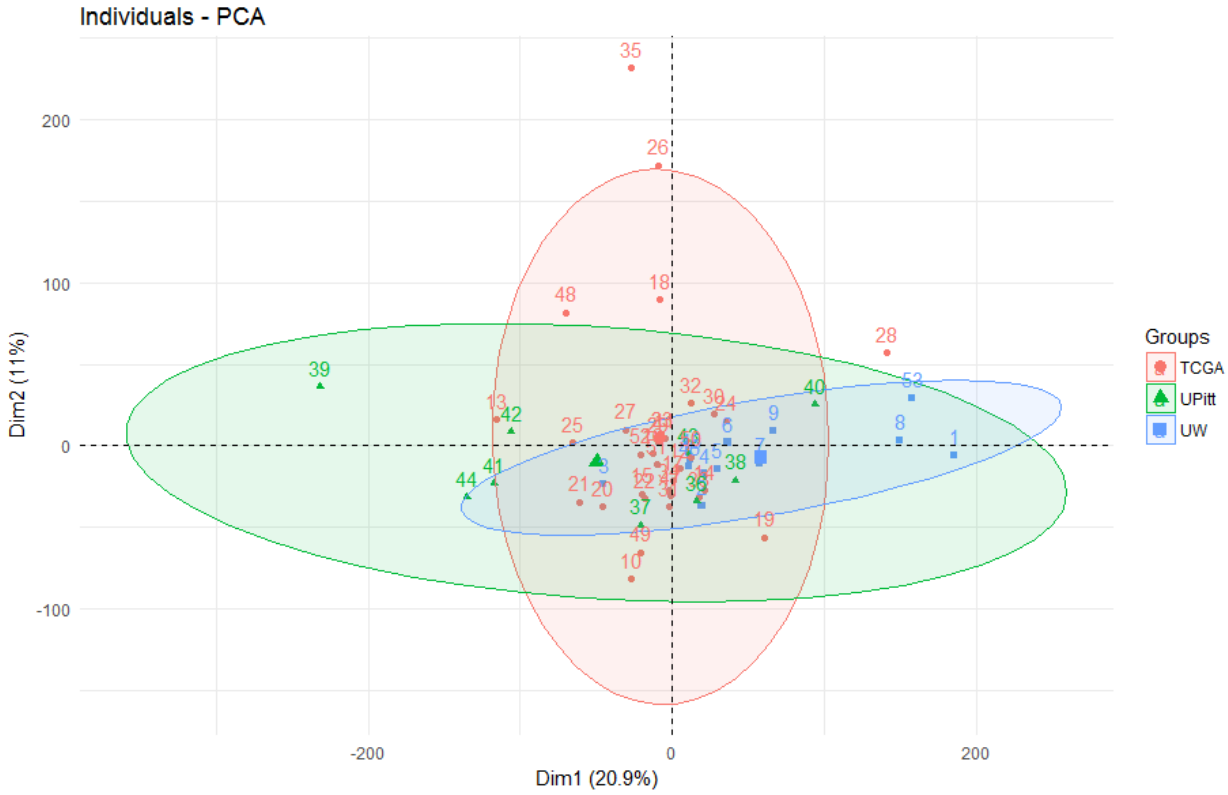

Individuals - PCA
Dim2 (11%)
Dim1 (20.9%)
200
100
0
-100
-200
0
200
Groups
TCGA
UPitt
UW

Figure S6 Legend

Figure S6. Principal component analysis (PCA) plot of whole exome-derived copy number data among HPV-related OPSCC samples by study site. Ellipses represent 95% confidence intervals. *TCGA*, TCGA head and neck copy number data. *UPitt*, UPitt head and neck copy number data. *UW*, UW head and neck copy number data.

Acknowledgements

The authors would like to thank Dr. Navonil De Sarkar of the Nelson lab for advice and guidance on selecting and implementing copy number analyses. We thank Dr. David Coffey of the Warren lab for advice and guidance with genomic analyses. We thank the members of our genomics shared resource at the Fred Hutchinson Cancer Research Center including Ryan Basom for his guidance on bioinformatics and programming, Andy Marty and Alyssa Dawson for their support with multiplexed sequencing. We thank David Gold and Dr. Paul Sampson of the UW Department of Biostatistics for biostatistical guidance and support in choosing appropriate regression models to test the association between genomic aberrations and tumor recurrence. We also thank Yuzheng Zhang for biostatistical and computational guidance.