HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2018 December 05.

Published in final edited form as:

Cell Rep. 2018 October 30; 25(5): 1332–1345.e5. doi:10.1016/j.celrep.2018.10.007.

Genomic and Transcriptomic Characterization Links Cell Lines with Aggressive Head and Neck Cancers

Hui Cheng[1], **Xinping Yang**[1], **Han Si**[2], **Anthony D. Saleh**[1], **Wenming Xiao**[3], **Jamie Coupar**[1], **Susanne M. Gollin**[4], **Robert L. Ferris**[5], **Natalia Issaeva**[6], **Wendell G. Yarbrough**[6], **Mark E. Prince**[7], **Thomas E. Carey**[7], **Carter Van Waes**[1,8,*], and **Zhong Chen**[1,*]

[1]Tumor Biology Section, Head and Neck Surgery Branch, National Institute on Deafness and Other Communication Disorders, NIH, Bethesda, MD 20892, USA

[2]Translational Bioinformatics, MedImmune, Gaithersburg, MD 20878, USA

[3]Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR 72079, USA

[4]Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA 15260, USA

[5]Division of Head and Neck Surgery, Departments of Otolaryngology, Radiation Oncology, and Immunology, University of Pittsburgh Cancer Institute, Pittsburgh, PA 15232, USA

[6]Department of Surgery, Division of Otolaryngology, Molecular Virology Research Program, Smilow Cancer Hospital, Yale Cancer Center, Yale University Medical School, New Haven, CT 06520, USA

[7]Cancer Biology Program, Program in the Biomedical Sciences, Rackham Graduate School, and the Department of Otolaryngology-Head and Neck Surgery, University of Michigan, Ann Arbor, MI 48109, USA

[8]Lead Contact

SUMMARY

Cell lines are important tools for biological and preclinical investigation, and establishing their relationship to genomic alterations in tumors could accelerate functional and therapeutic discoveries. We conducted integrated analyses of genomic and transcriptomic profiles of 15 human papillomavirus (HPV)-negative and 11 HPV-positive head and neck squamous cell carcinoma (HNSCC) lines to compare with 279 tumors from The Cancer Genome Atlas (TCGA). We identified recurrent amplifications on chromosomes 3q22–29, 5p15, 11q13/22, and 8p11 that drive increased expression of more than 100 genes in cell lines and tumors. These alterations, together with loss or mutations of tumor suppressor genes, converge on important signaling pathways,

*Correspondence: vanwaesc@nidcd.nih.gov (C.V.W.), chenz@nidcd.nih.gov (Z.C.).

AUTHOR CONTRIBUTIONS

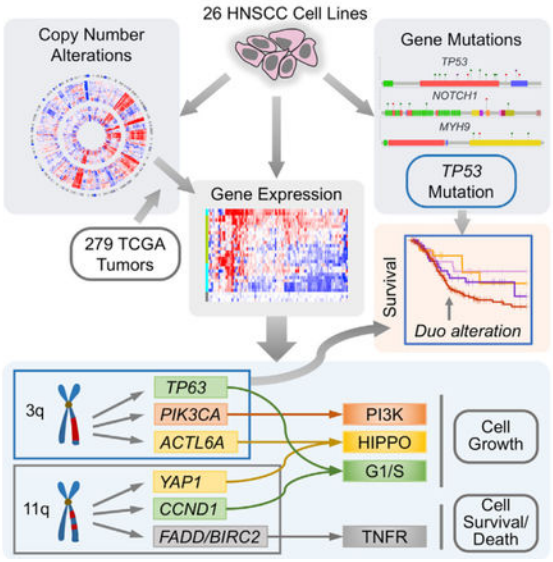
Conceptualization, C.V.W. and Z.C.; Resources, T.E.C., M.E.P., R.L.F., S.M.G., N.I., and W.G.Y.; Methodology, X.Y., H.C., J.C., A.D.S. and Z.C.; Sequencing, X.Y; Computational and Statistical Analysis, H.C., H.S., and W.X.; Figure and Table Preparation, H.C., Z.C.; Writing – Original Draft, H.C., Z.C., and C.V.W.; Writing – Review and Editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

recapitulating the genomic landscape of aggressive HNSCCs. Among these, concurrent 3q26.3 amplification and *TP53* mutation in most HPV(–) cell lines reflect tumors with worse survival. Our findings elucidate and validate genomic alterations underpinning numerous discoveries made with HNSCC lines and provide valuable models for future studies.

Graphical Abstract



In Brief

Cheng et al. reveal that head and neck squamous cell cancer cell lines display recurrent chromosome copy number alterations and mutations found in tumors with worse prognosis. The copy alterations correlate with expression of multiple genes implicated in squamous cell signaling pathways and programs that provide candidates for mechanistic studies and targeted therapy.

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) is the sixth leading cancer by incidence worldwide (Torre et al., 2015). Various chemical carcinogens (tobacco, alcohol, and betel nut), human papillomavirus (HPV) infection, and genetic predisposition contribute to the etiology of HNSCC and to the complex genetic alterations in tumor subsets that differ in prognosis and response to therapies (Van Waes and Musbahi, 2017). Establishing the functional role of genomic alterations in pathogenesis and as potential targets for therapy could be accelerated by genomic and transcriptomic characterization of a large panel of established cell line models differing in etiologic status. However, the extent to which HNSCC cell lines reflect the genomic and transcriptional alterations found in tumors remains unclear.

Human cancer cell lines have long fostered mechanistic, genetic, molecular, cellular, and pharmacologic studies that were not possible using tumor specimens alone. Several large cell line panels, such as the University of Michigan Squamous Cell Carcinoma (UM-SCC),

University of Pittsburgh Cancer Institute (UPCI:SCC), and other institutional series, have been authenticated genotypically, have been studied extensively worldwide (Brenner et al., 2010), and have led to important cancer discoveries and investigational drug development. However, until the recent advent of massively parallel sequencing, it has not been possible to comprehensively determine if and how well historic and newer cell line panels with germline controls model genomic and transcriptomic alterations in tumor subtypes that differ in etiology, origin, and prognosis.

Recently, a comprehensive landscape of genomic and transcriptomic alterations in HNSCC and other squamous tumors has emerged from The Cancer Genome Atlas (TCGA) Network (Campbell et al., 2018; Cancer Genome Atlas Network, 2015). TCGA revealed novel and previously recognized gene and chromosomal region copy number alterations (CNAs), mutations, and expression clusters and defined their frequency, co-occurrence, and relationship to common and rare subtypes of HPV(−) and HPV(+) tumors that vary in prognosis. Intriguingly, although a small subset of HNSCCs are mutation driven, both cytogenetic and TCGA studies indicate that most HNSCCs display multiple recurrent CNAs affecting broad and focal chromosome regions that harbor multiple candidate driver genes (Gollin, 2014). The frequent co-occurring events affecting multiple loci or genes suggest that alterations in associated pathways may elicit important complementary or synergistic effects important in tumorigenesis (Thomas et al., 2007). However, it remains to be fully determined whether these recurrent chromosomal alterations drive corresponding differential gene expression of one or multiple genes in both tumors and cell lines, where their potential functional interactions and roles in pathways of biologic and therapeutic interest may be investigated.

To this end, human cell line models in which the complex inter-section of CNAs, mutations, and gene expression has been defined are needed to investigate their functional role and therapeutic potential. In particular, the development of therapies for HNSCC patients with worse clinical outcomes and limited treatment options would be facilitated by identification of HNSCC lines modeling the genomic and expression alterations underlying these tumor subsets. Although the mutational landscape of HNSCC cell line panels has been characterized by targeted or exome sequencing (Kalu et al., 2017; Li et al., 2014; Pickering et al., 2013), these studies have not included a comprehensive and integrated characterization of CNAs, mutations, expression, and pathway analyses of a large panel of HPV(−) and HPV(+) cell lines for comparison with TCGA tumors.

Here, we used whole DNA exome and RNA transcriptome sequencing to characterize somatic CNAs, mutations, and expression profiles of 15 HPV(−) and 11 HPV(+) HNSCC cell lines. We found that the recurrent genomic alterations in cell lines are remarkably consistent with those found in more aggressive tumors, from which cell lines have traditionally been most readily adapted to culture (White et al., 2007). Genome-wide correlation of CN (copy number) with expression identified a suite of potential drivers or modifier genes that differ by HPV status and are of potential biologic and therapeutic relevance. Furthermore, our findings elucidate and validate genomic alterations underpinning numerous discoveries made with these widely used and recently derived

HNSCC lines and provide a roadmap for their potential use as models for future studies of tumor subtypes with worse prognosis.

RESULTS

DNA CNAs Identified in Cell Lines Recapitulate Those Found in Clinically Aggressive HNSCC Subsets in TCGA

The panel of 15 HPV(−) and 11 HPV(+) HNSCC cell lines included 18 widely studied and genotype-verified lines (Brenner et al., 2010; White et al., 2007), and 8 recently derived lines, which are those with matched blood controls (Table S1A, indicated in the table as blood: Y). CNA was determined from whole-exome sequence using CONTRA and significant focal CNAs were identified using GISTIC (Figure S1A; STAR Methods). HPV(+) cell lines were confirmed to have detectable HPV reads exceeding the threshold (of 100 reads for HPV(–) lines), and their HPV integration or non-integration status was determined (Figure S1B; Table S1B). We defined a recurrent CNA region as a chromosome region that shows a high probability of being altered in multiple samples (Figure S2A; GISTIC q value < 0.25). We identified recurrent chromosome gains in 3q, 5p, 7p, 8q, and 11q and losses in 3p, 5q, 8p, 9p, 10p, 18q, and 21q (Figures 1A and S2A), which are consistent with those observed in tumors in TCGA (Cancer Genome Atlas Network, 2015), previous cytogenetic analyses of tumors (Gollin, 2014), and published karyotypes of several of the UM-SCC lines (Van Dyke et al., 1994).

When comparing overall CNAs between recently derived lines and traditional cell lines, there was no significant difference be-tween recently derived and traditional HPV(−) lines ($t = 1.654$, $p = 0.198$, Welch two-sample t test), but there was a significant difference between recently derived and traditional HPV(+) lines ($t = 2.769$, $p = 0.023$, Welch two sample t test) (Figures S2B and S2C). Notably, there were fewer regions with significant CNAs in the subset of recently derived HPV(+) VUMC lines (VUMC-SCC 4149, 4755, and 4975), as summarized in Figures S2B and S2C. Interestingly, these VUMC lines were derived under different culture conditions using keratinocyte growth media and were noted to harbor lower HPV reads at baseline (Figure S1B) and lack of integrated HPV (Table S1B), as well as reduction in HPV and proliferative potential with further passages, consistent with episomal HPV. Exome sequencing (exome-seq) and CONTRA also displayed several broader regions with apparent increase in CN on chromosome (chr) 2, 4, 6, and 10 relative to control DNA that have not been broadly reported. Multiple amplification peaks on chr 2q were detected by GISTIC in TCGA HNSCCs (Cancer Genome Atlas Network, 2015), including 2q31, which harbors *NFE2L2*. NFE2L2 is a master transcription factor that regulates the genes involved in oxidative stress. Studies have found that gene signatures regulated by the NFE2L2 are associated with tumorigenesis and drug resistance in HNSCC (Namani et al., 2018). Broad CN gains on chr 4, 6, and 10 were not detected using GISTIC in TCGA tumors. Lower level or more focal gains in chr 2, 4, 6, and 10 can be sorted and visualized in subsets of TCGA tumor SNP array data using Integrated Genome Viewer (IGV) (Figure S3). Furthermore, although the apparent gains in chr 4 and 10 could be independently replicated by exome-seq using an Illumina platform with CONTRA analysis, those for chr 2 and chr 6 displayed a lack or narrower regions of gains, respectively (Figure S4). Thus, it appears CN

estimation based on exome-seq may overestimate CN for certain chromosomal regions (Fromer et al., 2012).

Several of the regions of broad or focal amplification and deletion harbor genes implicated in development and pathogenicity of HNSCC tumors (Figure 1A; Table S2). Notably, gain of 3q26–q28 is a dominant genomic alteration found in all of the HPV(−) and HPV(+) HNSCC cell lines included in this study (Figure 1A). This amplicon harbors more than 70 genes (Tables S3A and S3B), including several already experimentally implicated in HNSCC growth signaling (*PIK3CA*), stemness (*SOX2*), squamous cell differentiation, survival, migration, inflammation and angiogenesis (*TP63*), and cell immortalization (*TERC*) (Cao et al., 2008), as well as numerous less well characterized genes. Chr 11q harbored two recurrent focal amplifications found in TCGA (Figures 1B and 1C). Locus 11q13 (Huang et al., 2006) contains *FADD*, *CCND1*, *CTTN*, *FGF3/4/19*, *PPFIA1*, *MYEOV*, and *ANO1*, and 11q22 harbors *BIRC2/3*, *YAP1*, and *PDGFD*. Several of these genes have been reported to have individual or overlapping roles in deregulation of cancer cell cycle (*CCND1*, *FADD*), cell death and nuclear factor kB (NF-kB) signaling (*FADD*, *BIRC2/3*), growth factor (*FGFs*, *PDGFD*), and Hippo growth-regulatory (*YAP1*) pathways (Table S3A). Consistent with data from TCGA, 11q13/22 were predominately amplified in HPV(−) lines. Other common CNAs (Gollin, 2014) and genes potentially enhancing cell growth and survival include amplifications of oncogenes on 7p(*EGFR*) and 8q(*MYC*) (Table S2).

Focal CN losses found in HPV(−) lines and tumors include deletion of 9p21 (*CDKN2A/B/*p16) (Figure 1A), which has been implicated as an early event deregulating cell cycle control, and losses of other putative tumor suppressor genes (TSGs) on 8p (*TNFRSF10B*), distal 11q (*ATM*), and 18q (*SMAD4*, *GALR1*; Table S3). Notably, in the HPV(+) panel, we found only one line with low-level copy loss of 14q32.32 (*TRAF3*), in contrast to deletion and/or mutation observed in 20% of HPV(+) HNSCC tumors in TCGA (Cancer Genome Atlas Network, 2015), which are associated with better prognosis (Hajek et al., 2017). Several understudied focal amplifications (5p15, 8p11), deletions (5q, 10p, 21q), and associated genes were identified in HNSCC cell lines, which have also been associated with aggressive clinical-pathologic features of HNSCC (Tables S2 and S3A). Therefore, the CNAs identified in many of these HNSCC lines that were derived from patients with poor clinical outcomes similarly reflect those found among tumors with greater CNA and worse clinical features in TCGA.

Correlation between CNAs and mRNA Expression with Concordance in HNSCC Cell Lines and Tumors

To explore and compare how these CNAs affect the transcriptome in HNSCC tumors and lines, the associations between CNV and mRNA abundance for 17,040 genes were determined, and the overall concordance between CNA and gene expression for HNSCC tumors and cell lines was illustrated (Figure S5; $p < 2.2e\text{-}16$, $r_s = 0.56$, Spearman coefficient test). Among these, 4,500 genes displayed significant CNV-expression correlation in both tumors and cell lines (Figure S6; false discovery rate [FDR] < 0.05, Spearman co-efficient test). Figure 2 highlights examples of significant correlations between CN gains and expression for several oncogenes in HNSCC lines and tumors from TCGA. These include 3q

oncogenes *PIK3CA* and *TP63* in both HPV(−/+) subtypes; 11q13/22 genes *CCND1*, *FADD*/*BIRC2*, and *YAP1*, as well as 7p21 gene *EGFR*, mainly in HPV(−) lines. Conversely, CN loss of TSGs predominantly correlated with decreased expression, as illustrated by 8p gene tumor necrosis factor receptor superfamily member 10B/death receptor 5 (*TNFRSF10B*/DR5), also known as TRAIL receptor 2 (TRAILR2). The concordance between TCGA tumors and cell lines in terms of CNA and gene expression was also examined for individual chr 1, 2, 3, 4, 5, 6, 10, and 11, which were found to have broad or focal CNAs using exome-seq data and CONTRA analysis (Figure S5). Analysis of cell lines and TCGA tumors revealed higher concordance ($r > 0.6$) and statistical significance for chromosomes 3 and 11 with high CN gains, consistent with prior studies. Significant concordance but with lower r values (<0.6) were observed between cell lines and tumors for chr 1, 2, 4, 5, 6, and 10, which displayed CNAs in cell lines by exome-seq and CONTRA analysis. Overall, our analysis provides significant validation for concordance between CNA and mRNA expression and indicates that gene CN and expression of the broader transcriptome are highly consistent between the cell line panel and tumors from TCGA.

Gene Expression Driven by CN Gains Is Clustered on Chromosomes 3q, 5p, 8p, and 11q

We explored further how extensively gene expression in cell lines is driven by chromosomal amplifications. We conducted a supervised hierarchical clustering analysis of fold changes in gene expression between HNSCC cell lines and primary human oral keratinocytes (Figure 3). Remarkably, increased expression of the top 103 genes correlated with CN gain were found to co-cluster with the recurrent amplicons on chromosomes 3q22-qter (90 genes), 5p15 (7 genes), 11q13 (5 genes), and 8p11 (1 gene). Most lines clustered according to HPV status or patient origin. Three paired lines derived from primary and recurrent tumors from the same patient clustered together, while primary and metastatic lines UM-SCC22A/B clustered separately. Three recently derived HPV(+) VUMC lines clustered together and showed more limited expression of genes on chromosomes 3q, 5p, and 11q, consistent with overall lower number of significant CNAs detected in these lines (Figure S2B).

We performed a literature-based review, which revealed that many of these co-clustered genes represent understudied candidate oncogenes, which warrant further investigation (Table S3A). For example, expression of several genes at 5p appeared to be distributed among both HPV(−) and HPV(+) lines. The 5p15 amplicon cleft-lippalate gene *CLPTM1L* may promote resistance to mitochondrial cell death and cisplatin (Ni et al., 2012). *TRIP13* was recently reported to foster error prone DNA non-homologous end-joining and chemoresistance in HNSCC (Banerjee et al., 2014). *BRD9* is a bromo-domain family chromatin modifier reported to enhance *MYC* oncogene expression in AML. The 11q13 genes *CCND1*, *CTTN*, and *FADD* were more prominently expressed in a subset of HPV(−) than in HPV(+) lines, except for 93VU147T derived from a smoker, concurring with TCGA and another study (Ragin et al., 2006). Most HPV(+) lines exhibited higher expression of 3q27 gene *DVL3* implicated in Wnt-b-catenin pathway signaling. Other potentially interesting 3q26-qter candidates expressed across the lines include calmodulin kinase II *CAMK2N2*, implicated in nicotine-induced activation of transcription factor NF-kB and inflammatory and angiogenesis factor IL-8 (Tsunoda et al., 2016), and ephrin *EPHB3*, reported to promote migration in lung cancer (Efazat et al., 2016). In contrast, HPV(−) lines

and HPV(+) pair UPCI:SCC090/152 derived from a patient with primary and recurrent tumors who died within 22 months more often showed broader 3q arm gains with increased expression of *MAP3K13* (LZK), a kinase recently reported to stabilize mutant TP53 in HNSCC (Edwards et al., 2017); *ACTL6A*, which cooperates with 3q gene *TP63* and 11q22 Hippo pathway gene *YAP1*, linked to growth and poor patient survival in HNSCC (Saladi et al., 2017); PI3K-mTOR-eIF growth pathway (Iglesias-Bartolome et al., 2013) genes *PIK3CA*, *EIF4A2*, *EIF2A*, *EIF2B5*, and *EIF5A2*; and *ZNF639* (ZASC), a zinc finger transcription factor associated with recurrent oral squamous cell carcinoma (SCC) (Chiang et al., 2011). The *FGFR1* gene at 8p11 is a potential target for FGFR inhibitors (Dutt et al., 2011). Together, these findings suggest many of these CN-driven alterations in expression converge on common cancer-related pathways or functions. We also performed supervised clustering of mRNA expression for 832 classifier genes previously used to define HNSCC subtypes (Cancer Genome Atlas Network, 2015). Compared with the subtyping observed by mRNA expression in TCGA tumors (Cancer Genome Atlas Network, 2015), some HNSCC cell lines distinctly mapped to certain molecular subtypes (Figures S7A and S7B): UM-SCC11A, B and 74 A,B co-clustered with a mesenchymal-type expression signature; UM-SCC46 and UPCI:SCC154 were closest to the atypical sub-type; and UM-SCC104 was closest to the basal subtype. Alternatively, heatmap display of the correlation-based distances between centroids for HNSCC cell lines and expression sub-types in tumors from TCGA shows how cell lines overlap the profiles of one or more TCGA tumor subtypes (Figure S7B). Among these, the cell lines UPCI:SCC90 and 152 from the same patient showed the shortest distance from the classical subtype but also displayed expression pattern of the atypical subtype. Most of the remainder either were close to basal or did not have strong expression patterns close to any subtypes.

Integrated Analysis Reveals Differential Gene Expression in Distinct Signaling Pathways by HPV Status

To further characterize the potential relationship of concordant genomic and expression alterations to signaling path-ways, we selected 1,544 genes that showed significant and consistent amplification or deletion in both tumor specimens and HNSCC cell lines (Figure S6; GISTIC q values < 0.25). Empirically selecting as a cutoff an absolute expression fold change (FC) > 1.5, filtering yielded 492 and 440 genes from HPV(−) and HPV(+) lines, respectively. Ingenuity Pathway Analysis (IPA) identified distinct canonical pathways and networks enriched for genes displaying CNA-related differences in expression for HPV(−) and HPV(+) cell lines (Table S4). These overlapping IPA networks include growth factor receptor/PI3K-mTOR-eIF-protein synthesis and NF-kB pro-survival signaling; p53/death receptor/mitotic DNA damage checkpoint arrest, inflammation, angiogenesis, and viral entry; and stemness and differentiation (Figure 4; Table S4). HPV(−) and HPV(+) HNSCC cell lines overexpressed pathway components in PI3K signaling (*FGFR1*, *PIK3CA*) and WNT differentiation signaling pathways (*DVL3*). Increased *CCND1*, *FADD*, and *BIRC2/3* and decreased *CDKN2A* and *TNFRSF10B*, involved in cell cycle and death pathways, were predominant in HPV(−) cell lines (Figure 4A). In-contrast, greater increased *CDKN2A* (p16) and decreased death pathway-related genes *ATM* (11q22-qter), *ATR*, *RFC3*, *MRE11A*, and *PPP2R1B* occurred in HPV(+) cell lines (Figure 4B). Together, these findings

underscore potentially important differences in genomic and transcriptomic alterations in signaling pathways in HPV(−) and HPV(+) HNSCCs.

Mutations and Variants in HNSCC Lines Compared with Tumors from TCGA

To identify candidate somatic mutations in lines without or with matched DNA controls, we developed a custom pipeline that prioritized infrequent variants with minor allele frequency (MAF) < 1% and somatic and functionally deleterious mutations identified by COSMIC and Mutation Assessor (Figure S8). By filtering against the 396 top mutated genes discovered by MutSig in TCGA tumors, we identified 155 mutations occurring in 104 genes in the cell lines. Conservation of a footprint-like pattern of recurring mutations was observed in four cell line pairs derived from the same patient (Figure 5A). Significantly mutated genes in TCGA were also frequently mutated in cell lines, including established and putative TSGs involved in genomic stability, cell cycle, and survival (*TP53, RB1, CASP8*), squamous differentiation (*NOTCH1*), Hippo signaling (*FAT1*), growth factor receptor-PI3K-AKT/PKB signaling pathway (*IGF1R, MET, ERBB2, EPHA2, PTEN*), histone lysine methylation (*KMT2D*), and immunosurveillance (*HLA-B and HLA-A*) (Figures 5A, 5B, and S9; Tables S5 and S6A). Mutations in genes of those lines that overlap a recent whole-exome sequencing study showed high concordance (Kalu et al., 2017). Remarkably, whereas *PIK3CA* was amplified across both HPV(_) and HPV(+) cell lines in this panel, we did not find any of the hotspot activating mutations of *PIK3CA* that occur in _21% of the TCGA dataset. Lack of *PIK3CA* mutation in cell lines was validated by manual review of ExomeSeq and RNA sequencing (RNA-seq) data of these cell lines and independent targeted sequencing (Kalu et al., 2017). As *PIK3CA* hotspot mutations were also observed relatively infrequently by targeted sequencing in other HNSCC lines (Mazumdar et al., 2014), these observations suggest that HPV(_) and HPV(+) tumors harboring 3q amplifications associated with more aggressive tumors may be more readily selected and adapted to cell culture. Similarly, we also did not observe mutations in *NFE2L2*. As *NFE2L2* gains were observed in TCGA, we analyzed CN variation and gene expression of *NFE2L2* locus in the 26 cell lines. Interestingly, the IGV CN heatmap and scatterplot show that all the cell lines except UPCI:SCC90, 152 have at least one copy gain, and many of the lines show increased *NFE2L2* expression. The expression and CN of *NFE2L2* are highly correlated (r = 0.63) and significant (p = 0.00058) (Pearson correlation).

Our HNSCC panel includes cell lines with rare gene mutations observed in HNSCC tumors in TCGA or other studies (Tables S5 and S6A). A homozygous Q1186* nonsense mutation in *MYH9* was discovered in UM-SCC 110, derived from an unusually young 39-year-old male smoker with an HPV(−) anterior tongue tumor who died of early recurrence (Figures 5B and 6; Tables S1A, S5, and S6A). Heterozygous mutation was observed in matched blood, and a loss of heterozygosity (LOH) was observed in the cancer cell line, consistent with a second hit loss of the wild-type allele. This *MYH9* associated cancer also had *TP53*, *NOTCH1*, and *EP300* mutations associated with smoking-related HPV(−) HNSCC. TGFb pathway *SMAD4* nonsense mutations were seen in the primary and metastatic pair UM-SCC-22A, B. Other less frequent mutations (Figure 5A; Table S5), and seven genetic polymorphisms implicated as prognostic factors were also identified (Table S6B). For more

abundantly expressed genes and common alterations, we found high concordance between mutations detected by exome-seq and RNA-seq (Tables S5 and S6A).

Integration of Genetic Alterations Highlights Potential Therapeutic Targets for Preclinical Studies

We summarized the major CNAs and mutations for key oncogenes and TSGs in HNSCC (Figure 6). Significantly, these lines included genomic alterations in 3q,11q, and *TP53* observed frequently in TCGA dataset. Our previous publications demonstrated the effects of genetic and/or pharmacologic modulation of 3q *PIK3CA*, the Δ*Np63* isoform of *TP63*, and *SOX2*, as well as 11q *FADD* and *BIRC2* in distinct subsets of lines from this panel (Mohan et al., 2015; Yang et al., 2011; Eytan et al., 2016). We also analyzed the genomic and transcriptionally altered genes from Figures 3 and 5 against the drug databases in MetaCore and IPA. Altered genes with potential as therapeutic targets are represented among the cell lines as potential models for drug screens (Table S7). Potentially interesting examples of drug targets include recently characterized oncogenes, such as *FGFR1* at 8p (AZD4547 and BGJ398) (Chae et al., 2017) and *LZK* at 3q26 (lestaurtinib) (Davis et al., 2011).

3q26.3 CN Gain and TP53 Mutation Co-occurrence Is Associated with Worse Clinical Outcome in HPV ( ) Tumors

It is evident that many HNSCC cell lines are preferentially adapted from tumors of patients with a poorer prognosis and who died within 2 years (Table S1A). Interestingly, all cell lines displayed 3q26–q28 CN gain, whereas 3q CN gain is observed in 35% of tumors in TCGA (low-level CN gain, 28%; high level CN gain, 6.5%), and overall, 3q26–q28 CN gain is observed in 69% tumors, whereas others appear to be driven by 3q gene *PIK3CA* mutations and/or other alterations. To confer a selective advantage in tumors and cell culture, genetic alterations often cluster in a set of essential pathways and act in concert (Ciriello et al., 2012; Gross et al., 2014). In this study, we observed that most of the HPV(−) cell lines display 3q26.3 CN gain and *TP53* mutation, which also significantly co-occurred in 159 of 243 HPV(−) TCGA tumors (p = 3.3e-06, two-sided Fisher's exact test; Figure 7A). We explored how 3q amplification that is linked with expression of multiple putative oncogenes and mutations of 3q gene *PIK3CA*, and mt *TP53*, affect prognosis. Among 243 TCGA HPV(−) HNSCCs, those containing $TP53^{Mut}$-$3q26.3^{Amp}$ had substantially worse prognosis than those with $TP53^{Mut}$ or $3q26.3^{Amp}$ alone, with a hazard ratio (HR) of 2.2 (p = 0.04, log rank test) (Figure 7B). The 3q26.3 segment encompasses 30 protein-coding genes, including *PIK3CA*, *ACTL6A*, *SOX2*, *TP63*, and others implicated in cancer stemness, growth, proliferation, and differentiation (Table S3B). Among these genes, three have been reported to be p53 targets (Fischer, 2017): *ECT2* (epithelial cell transforming 2) (Scoumanne and Chen, 2006), *ZMAT3/WIG1* (zinc finger, matrin-type 3) (Bersani et al., 2014), and *PIK3CA* (Astanehe et al., 2008). When we compared the overall survival for patients with different combinations of $3q26.3^{Amp}$ and *PIK3CA* mutation (Figure 7C), $3q26.3^{Amp}$ with wild-type *PIK3CA* was associated with a worse prognosis than *PIK3CA* mutation alone or $PIK3CA^{Mut}$-$3q26.3^{Amp}$ (HR = 1.4, p = 0.3, log rank test; Figure 7D). We further investigated how known activating *PIK3CA* hotspot mutations and $3q26.3^{Amp}$ comparatively affected survival. We identified patients with *PIK3CA* hotspot mutations (c.1624G > A[E542K], c.1633G > A[E545K] and c.3140A > G [H1047R]) in TCGA data and compared

survival between patients with 3q26.3 amplification versus those with *PIK3CA* hotspot mutations. We found that patients with 3q26.3 amplification also had worse overall survival than patients with *PIK3CA* hotspot mutations (HR = 2.1, p = 0.29, log rank test). Interestingly, HPV(–) HNSCC patients with the *PIK3CA* hotspot mutations trended toward a better prognosis compared with patients without either alteration events, but the significance of the difference could not be established with the limited sample size in our analysis (HR = 0.7, p = 0.64, n = 7, survival > 60 months). These results support a role for other co-occurring genes on 3q or other chromosomes that are altered in higher CNA sub-sets of HNSCC that display worse prognosis.

DISCUSSION

We report a comprehensive genomic and transcriptomic characterization of 26 HNSCC cell lines, including 15 with HPV(−) and 11 with HPV(+) status, and performed a systematic comparison with genomic alterations in tumors from TCGA. These cell lines reflect common and rare genomic and transcriptomic alterations uncovered in clinically advanced and aggressive tumors by TCGA and other genomic studies. Through integrated genomic, expression, and pathway analysis, we elucidate deregulated gene expression involved in critical oncogenic pathways, under-pinned by 3q amplification and other shared or subset-specific alterations characteristic of HNSCC with different HPV status. The biological, clinical, and therapeutic relevance of some of these alterations, pathways, and functions has been anticipated and validated by hypothesis-driven experimental studies reported by us and others using long-established cell lines included in this panel (Conti et al., 2015; Eytan et al., 2016; Herzog et al., 2013; Mohan et al., 2015). Recent reports also support the utility of several HNSCC lines in this panel with mtTP53 in functional genetic or pharmacologic screens demonstrating pre-clinical activity of WEE1 kinase inhibitors (Moser et al., 2014). Together, these studies have highlighted the importance of a network of pathways in promotion of the malignant phenotype and therapeutic resistance, and of investigating the contributions of genomic and transcriptomic alterations, which was here-tofore limited to selected gene candidates.

TCGA HNSCC study showed that patients from a subset with higher CNAs, including 3q, 5p, and 11q amplification, have a worse prognosis (Cancer Genome Atlas Network, 2015). Here, we find that these alterations are highly represented and associated with broad alterations in gene expression across biologically and therapeutically relevant pathways in both tumors and the long-and recently established HNSCC lines included in this study. It has been evident that cell lines from HNSCC and other cancers are often more readily adaptable from tumors associated with worse clinical outcomes, suggesting they harbor alterations conducive to growth both *in vivo* and *in vitro* (White et al., 2007). Strikingly among these, we observed 3q amplification with higher frequency in both HPV(−) and HPV(+) lines in this panel, than was observed in tumors by TCGA. Furthermore, none of our 26 HNSCC cell lines harbored hotspot activating mutations of the 3q gene *PIK3CA*, consistent with our finding that TCGA HNSCC patients with 3q26.3 amplification had worse overall survival than those with mutations of *PIK3CA*.

These observations suggest that 3q26 and other CNAs may harbor and drive altered expression of multiple genes that contribute to worse prognosis as well as cell culture adaptability. Consistent with this hypothesis, integration of mRNA expression and CNA data from the cell lines and TCGA tumors uncovered significant overlap, supporting the contribution of 3q, 5p, 8p, and 11q amplifications to increased expression of diverse genes that potentially contribute to HNSCC pathogenesis. Remarkably, among the 103 most significantly amplified genes with elevated expression in HNSCC cell lines and tumors, 90 of these genes were from 3q22-qter, with the remainder from 5p15, 11q13, and 8p11. We annotated multiple interesting growth and signaling pathway-related gene alterations, which merit further biological and preclinical investigation. These cell lines may therefore represent useful models for investigation of the role of CN-driven expression of clustered genes that contribute to biologic features and therapeutic resistance of aggressive HNSCC subtypes.

There are multiple candidate gene drivers on 3q, which are implicated in growth, metabolism, and stemness in cell line models. Early studies indicated *TP53* family (Yamaguchi et al., 2000) oncogene *TP63* (Yang and McKeon, 2000), PI3-kinase subunit *PI3KCA,* protein kinase C *PRKCI* (Justilien et al., 2014), stemness gene *SOX2* (Justilien et al., 2014) as well as *ECT2* (Justilien and Fields, 2009), and *FXR1* (Majumder et al., 2016) genes as drivers in lung, esophagus, or HNSCC. It is increasingly appreciated that these and recently characterized 3q genes could collaborate with one another or with genes encoded by other genomic alterations in SCCs (Justilien and Fields, 2009; Majumder et al., 2016). We have shown that both the DNp63 iso-form of *TP63* and *PIK3CA* overexpressed by UM-SCC lines in the current panel promote activation of NF-kB transcription factors and gene programs that promote proliferation, migration, and inflammation (Yang et al., 2011). Furthermore, the *PRKCI* and *SOX2* oncogenes are co-amplified and have been implicated in cooperative activation of Hedgehog signaling in lung SCC (Justilien et al., 2014). Recently, the *ACTL6A* gene has been reported to be co-amplified with *TP53* family oncogene *TP63* on 3q, and together with Hippo pathway gene *YAP1* amplified on 11q13, cooperate to drive proliferation and poor prognosis of HNSCC (Saladi et al., 2017). Several other components of the PI3K-mTOR (*EIF4A2*, *EIF2B5*, *EIF2A*, *EIF5A2*) and WNT (*DVL3*) pathways that are co-amplified on 3q and overexpressed in HNSCC cell lines and tumors may also merit further investigation.

Our analysis of TCGA data reveals that patients with aggregate events of 3q26.3 gain and *TP53* mutation are associated with a significantly worse prognosis. Intriguingly, supplemental data in a study of 250 TCGA HPV(−) HNSCCs (Gross et al., 2014) also support a stronger association between mutation of *TP53* with 3q26.3 gain than with gains in other chromosomal regions. The aforementioned study focused upon *TP53* mutation and 3p loss, although the question remained as to what genes encoded on chr 3p or other co-occurring chromosomal alterations are responsible for the interaction with *TP53* (Gross et al., 2014). Numerous cytogenetic studies have found that the coordinate loss of 3p, and gain of 3q26.3, often co-occur as a consequence of an isochromosome formation, resulting in duplication of distal 3q and loss of 3p (Gollin, 2014). Strikingly, we found that deletion of 3p significantly co-occurred with CN gain of 3q26.3 in 135 of 243 HPV(−) TCGA tumors ($p = 1.07e-06$, two-sided Fisher's exact test) but not with overall amplification of 3q ($p = 0.59$,

two-sided Fisher's exact test). Our findings of significant co-occurrence of 3q26.3 gain and *TP53* mutation, and reports identifying p53 target genes on 3q26.3, provide a possible explanation for the highly significant relationship between 3q26.3 amplification, *TP53* mutation, and worse prognosis. Gain of 3q26.3, loss of 3p, and mutation of *TP53* appear to be acquired early in the pathogenesis of HNSCC, suggesting an important role in early oncogenesis (Leemans et al., 2011), and the potential for broader gene interactions among these loci merit further investigation. *TP53* mutation has long been associated with poorer prognosis for patients but has been insufficient alone in predicting outcome (Cancer Genome Atlas Network, 2015). The role of inactivating mutations of *TP53* in enhancing the overexpression of the *PIK3CA* gene has been reported (Astanehe et al., 2008). In addition, we uncovered a mitochondrial biosynthesis gene signature among 3q genes. Patients with Li-Fraumeni syndrome (LFS) carrying germline *TP53* mutations have increased abnormal mitochondria function compared with healthy volunteers (Wang et al., 2013). Mean-while, genetic or pharmacologic disruption of mitochondrial respiration improved cancer-free survival in a mouse model of LFS (Wang et al., 2017). These studies, together with our observation that amplification and overexpression of multiple 3q26.3 genes are associated with *TP53* mutation or inactivation by HPV, suggest *TP53* may modulate broader 3q26.3 gene programs and signal networks that regulate bioenergetic homeostasis, growth, stemness, and differentiation.

HNSCC cell lines recapitulated other frequent CNAs and mutations observed in tumors from TCGA that affect death pathways and that were recently implicated as therapeutic targets using these cell lines. Intriguingly, HPV(−) tumors and lines exhibit alterations of the TRAILR2-FADD-IAP-CASP8 axis, while HPV(+) tumors and lines displayed copy loss and decrease in expression of ATM, ATR, and other death pathway components. Using several HPV(−) lines from this panel, we recently showed that *FADD* amplification and overexpression can serve as an Achilles' heel that sensitizes HNSCC to IAP antagonists in combination with exogenous death factor TNF or radiation (Eytan et al., 2016). We observed *NFE2L2* gains and expression more frequently than mutations compared with TCGA. We did not detect mutations such as *HRAS* or co-occurrence with the heterozygous *CASP8* mutation found in UM-SCC-1, which are seen in small subset of TCGA HNSCC tumors lacking 3q and 11q gains.

These cell lines displayed rare alterations as well as recurrent mutations and CN losses of putative TSGs consistent with TCGA and prior literature reports. A unique line derived from a tongue cancer displayed a rare homozygous *MYH9* gene mutation compared with heterozygous germline mutation, consistent with loss of the wild-type (WT) allele. As we showed that *MYH9* knockout mice develop aggressive oral HNSCC (Conti et al., 2015), this line may facilitate studies of how *MYH9* functionally contributes as a TSG. HPV(+) HNSCC cell lines, pre-dominantly from oropharyngeal tumors, exhibited infrequent mutation or genetic loss in *CDKN2A/RB1* or *TP53* components, which are targeted by HPV oncoproteins. However, these cell lines displayed copy loss and decreased expression of DNA damage response genes ATM and ATR, which function as up-stream activators of WT TP53 and could augment the death defects attributed to *TP53* inactivation by HPV. Interestingly, somatic aberrations in DNA-repair genes, including *ATM*, were also reported in HPV(+) tumors (Seiwert et al., 2015). In addition, one HPV(+) line harbored a *RB1*

frameshift deletion, which could also magnify the effects of RB inactivation by HPV E7 protein.

The apparent gains on broad regions of chromosomes 2, 4, 6, and 10 were investigated using several approaches. Broad low-level or focal gains were also observed in subsets of TCGA tumors (chr 2, 4, 6, and 10); these apparent gains observed in cell lines and subsets of TCGA tumors also significantly affect expression of genes in these regions in HNSCC lines and TCGA tumors, as demonstrated by concordance between cell lines and TCGA tumors. Some of these chr alterations were observed in previous studies. Amplifications of 2q, 4q, 6p, and 6q were described in the invasive front area of laryngeal carcinoma and associated with aggressive tumor progression (Ambrosio et al., 2013). The 6p21 region harbors oncogenes *CCND3*, *HMGA1*, cell cycle regulators *CDKN1A*, *CDC5L*, human leukocyte antigen (*HLA*), and cytidine deaminase gene *APOBEC2* genes. Among these, the enrichment of APOBEC mutations were also observed in HNSCC and other SCCs (Roberts et al., 2013). Although some of the gains were replicated with Illumina ExomeSeq (chr 4 and chr 10), those for chr 2 and chr 6 only partially overlapped. This apparent overestimation of CNA from whole-exome sequencing (WES) could stem from differences in read fragment length or depth resulting from DNA isolation, exome capture, and/or PCR amplification, and therefore WES-based CN methods for some regions may not always reflect true CN variation.

Interestingly, the recently derived HPV(+) VUMC lines showed significantly fewer CNAs and HPV reads compared with traditional HPV(+) lines. CNA and corroborating RNA expression data indicate that these lines with lower HPV reads, integration, and proliferative potential also harbor fewer CNAs and corresponding expressed genes. Of potential interest, the VUMC lines were cultured in keratinocyte growth medium, which differs in the lower concentration of calcium, serum, and other factors, which are used to favor and select for growth rather than differentiation of normal keratinocytes. In comparison, high calcium serum-supplemented medium has been used for culture of the most SCC cancer lines, including the others in this study. Although our sample is limited, these unexpected observations suggest that culture conditions could be another important factor determining the success rate of culture and selection of tumors differing in HPV status, genomic CNAs and prognosis.

In summary, these HNSCC cell lines exhibit common and rare genetic alterations observed in TCGA tumors, providing a valuable resource as head and neck tumor models. HNSCC cell lines display concurrent genetic alterations in multiple signaling pathways, consistent with those derived from clinically advanced and aggressive tumors, supporting the concerted involvement of these pathways in tumorigenesis, progression, and metastasis. The characterization of CNAs and mutations has already identified important genetic alterations underpinning sensitivity or resistance found in prior and recent preclinical studies. Identifying new targets and potential mechanisms of resistance may facilitate mechanistic and preclinical studies to advance personalized cancer medicine for HNSCC.

STAR*METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Raw, processed and clinical data	TCGA Network	https://portal.gdc.cancer.gov/legacy-archive/search/f
TCGA publication page	TCGA Network	https://gdc.cancer.gov/about-data/publications/hnsc_2014
TCGA publication page	TCGA Network	https://gdc.cancer.gov/about-data/publications/mc3-2017
COSMIC Census	Forbes et al., 2017	https://cancer.sanger.ac.uk
FireBrowse portal	Broad Institute	http://gdac.broadinstitute.org
cBioPortal	Memorial Sloan Kettering Cancer Center	http://www.cbioportal.org
dbSNP	NCBI dbSNP	http://www.ncbi.nlm.nih.gov/SNP
1000 Genomes Project	The International Genome Sample Resource (IGSR)	http://www.internationalgenome.org/
Critical Commercial Assays		
SOLiD 5500	ThermoFisher	https://www.thermofisher.com/us/en/home/brands/product-brand/5500-series-genetic-analysis-systems.html
Deposited Data		
Phenotype, Exome DNASeq and RNASeq data (deposited in the Short Read Archive (SRA: http://www.ncbi.nlm.nih.gov/sra/) of NCBI under BioProject)	This paper	SRA: PRJNA453457; dbGaP: phs001581
Experimental Models: Cell Lines		
UM-SCC cells	Brenner et al., 2010	N/A
VUMC-SCC cells	Provided by Dr. Wendell Yarbrough	N/A
UPCI:SCC cells	Busch et al., 2013	N/A
UD-SCC cells	Busch et al., 2013	N/A
VU cells	Busch et al., 2013	N/A
Software and Algorithms		
GISTIC2	Mermel et al., 2011	http://software.broadinstitute.org/cancer/software/genepattern/modules/docs/GISTIC_2.0

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Integrative Genomics Viewer	Robinson et al., 2011	http://gdac.broadinstitute.org/igv
MutSig	Lawrence et al., 2013	http://software.broadinstitute.org/software/cprg/?q=node/35
LifeScope v.2.5	LifeScope Genomic Analysis Solutions	https://www.thermofisher.com/search/results?query=lifescope&refinementAction=true&persona=DocSupport&focusarea=Search%20All
CONTRA	Li et al., 2012	http://contra-cnv.sourceforge.net/
ANNOVAR	Wang et al., 2010	http://annovar.openbioinformatics.org/en/latest/
OmicCircos	N/A	http://bioconductor.org/packages/release/bioc/html/OmicCircos.html
MutationAssessor	Reva et al., 2011	http://mutationassessor.org/r3/
NovoalignCSMPI	Wang et al., 2011	http://www.novocraft.com/documentation/novoaligncs-3/
STAR: ultrafast universal RNA-seq aligner	Dobin et al., 2013	https://code.google.com/archive/p/rna-star/
USeq	Nix et al., 2008	http://useq.sourceforge.net/
SAMtools	Li et al., 2009	http://samtools.sourceforge.net/
SAMtools: bcftools	Li, 2011	http://samtools.sourceforge.net
ComplexHeatmap	Gu et al., 2016	http://bioconductor.org/packages/devel/bioc/html/ComplexHeatmap.html
Ingenuity Pathway Analysis	N/A	https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/
ConsensusClusterPlus R package	Wilkerson and Hayes, 2010	https://bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html
Survival R package	N/A	https://cran.r=project.org/package=survival

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be facilitated by the Lead Contact, Carter VanWaes (vanwaesc@nidcd.nih.gov).

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Human Subjects—Tumor tissue, adjacent normal tissue, and normal whole blood samples were obtained from patients at contributing centers with informed consent according to their local Institutional Review Boards (IRBs, see below). Biospecimens were centrally processed and DNA, RNA, and protein were distributed to TCGA analysis centers.

TCGA Project Management has collected necessary human subjects documentation to ensure the project complies with 45-CFR-46 (the “Common Rule”). The program has obtained documentation from every contributing clinical site to verify that IRB approval has been obtained to participate in TCGA. Such documented approval may include one or more of the following: 1) An IRB-approved protocol with Informed Consent specific to TCGA or a substantially similar program. In the latter case, if the protocol was not TCGA-specific, the clinical site PI provided a further finding from the IRB that the already-approved protocol is sufficient to participate in TCGA; 2) A TCGA-specific IRB waiver has been granted; 3) A TCGA-specific letter that the IRB considers one of the exemptions in 45-CFR-46 applicable. The two most common exemptions cited were that the research falls under 46.102(f)(2) or 46.101(b)(4). Both exempt requirements for informed consent, because the received data and material do not contain directly identifiable private information; 4) A TCGA-specific letter that the IRB does not consider the use of these data and materials to be human subjects research. This was most common for collections in which the donors were deceased.

A total of 279 patients were analyzed in TCGA HNSCC analysis with clinical data, whole exome sequencing (WES), RNA sequencing, miRNA sequencing, methylation arrays, and DNA SNP chips. Most patients were male (203 males, 73% and 76 females, 27%), heavy smokers (mean pack years = 51) and used alcohol (188 out of 279, 67%). The median age of diagnosis was 61 years of age.

Cell lines were established in accordance with IRB, exemption, or consent requirements of 45-CFR-46 (the “Common Rule”) or Declaration of Helsinki in place at the time of acquisition. The genomic and transcriptomic analysis of the cell lines from de-identified individuals was exempted from IRB review by The NIH Office of Human Subjects Research Protection.

Clinical Samples, Data Types, and Genomic Platforms—Details about sample collection, tissue-specific sample selection criteria, clinical annotations, and the genomic data pipelines can be found in the TCGA HNSCC publication (Cancer Genome Atlas Network, 2015). Human papilloma virus (HPV) status was assessed by clinical assay (p16 immunohistochemsitry staining or *in situ* hybridization) and multiple nucleic acids-based sequencing platforms including RNA-Seq (n = 279 samples), WES (n = 279 samples), “high coverage” WGS (n = 29) and “low pass” WGS (n = 98). Samples were classified as HPV positive using an empiric definition of detection of > 1000 mapped RNA-Seq reads,

primarily aligning to viral genes E6 and E7. In general there was good agreement between all measures of HPV assessment, indicating that 36 tumors were HPV(+) and 243 were HPV(−). Although 64% of the 33 oropharyngeal tumors were HPV(+), 14 of the 36 cases in which the molecular data strongly support an HPV(+) diagnosis were from non-oropharyngeal tumors. TCGA clinical and platform data are available on the GDC website (https://gdc.cancer.gov/).

Cell lines—We selected a panel of 15 HPV(−) and 11 HPV(+) HNSCC cell lines, including 18 historically widely studied and genotypically validated lines (Brenner et al., 2010; White et al., 2007), and 8 recently derived lines with matched blood DNA controls (Table S1A). Cell line stocks from earliest passages available were genotyped and preserved in liquid nitrogen and cultured for less than 1 month before sequencing. UM-SCC lines (University of Michigan) originated from Drs. Thomas E. Carey, Mark E. Prince, and Carol R. Brad-ford; VUMC-SCC lines (Vanderbilt University) from Dr. Wendell Yarbrough; UPCI:SCC lines (University of Pittsburgh) from Dr. Susanne M. Gollin; UD-SCC line (University of Dusseldorf,€ Germany) from Drs. Thomas K. Hoffman and Henning Bier; and VU line (Vreije University, Netherlands) from Johan de Winter. Authentication of UM-SCC and other lines was done at the University of Michigan by DNA genotyping of alleles for 9 loci, D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11, FGA, vWA, and the amelogenin locus in 2015, as described (Brenner et al., 2010). The HPV expression and integration status of the lines was verified by RNA and DNA sequencing (Figure S1A & B, Table S1B).

METHODS DETAILS

Sequencing

Whole Exome Sequencing (WES): Genomic DNA from HNSCC lines and matched blood samples was isolated from frozen cell pellets using the DNeasy Blood & Tissue Kit (QIAGEN). Multiplexed libraries for exome capture sequencing were constructed utilizing the 5500 SOLiD Fragment Library Core Kit, TargetSeq Exome Enrichment Kits and 5500 SOLiD Fragment Library Barcode Adaptors 1–16 (Life Technologies). The exome enriched genomic DNA libraries were clonally amplified by emulsion PCR. Whole-exome sequencing achieved an average of 87x coverage across targeted bases, with 85% of target bases above 20x coverage.

RNA sequencing: The total RNAs of HNSCC lines and 3 keratinocyte lines were isolated by combination of Trizol method (Life Technology) and QIAGEN RNeasy kit (QIAGEN). Ribosomal RNA was depleted using Ribo-Zero kit (Epicenter). Multiplexed whole transcriptome libraries were generated by SOLiD Total RNA-Seq Kit and SOLiD RNA Barcoding Kit (Life TechnologiesTM), and fragmented cDNA libraries were clonally amplified by emulsion PCR.

The multiplex DNA and cDNA libraries were sequenced utilizing 75 bp forward and 35 bp reverse paired-end sequencing chemistry on the ABI SOLiD system. Color space reads were mapped into human NCBI Build 37 reference genome (Hg19) using LifeScope v.2.5

Genomic Analysis Software (https://www.appliedbiosystems.com/lifescope) that was developed for the 5500 Series Genetic Analysis Systems.

Copy number (CN) analysis—The exome capture area comprised 45.1 Mb in the coding regions of over 19,000 genes (https://tools.thermofisher.com/content/sfs/manuals/TargetSeq_Exome_System_man.pdf) identified in VEGA (http://vega.sanger.ac.uk), CCDS (https://www.ncbi.nlm.nih.gov/CCDS) and RefSeq (http://www.ncbi.nlm.nih.gov/refseq) databases. The bam files were analyzed with CONTRA (COpy Number Targeted Resequencing Analysis) v2.0.4 with reference to a pooled baseline control set comprising 8 matched blood samples. A bed file showing the primary target regions defined by hg19-based RefSeq database was used as reference. Copy number variation in terms of gene-level log-ratios were computed by taking the mean of the adjusted region-level log-ratios (RLRs) of all the targeted regions within a gene. Significant focal copy number alterations were identified from segmented data using GISTIC (Genomic Identification of Significant Targets in Cancer) 2.0. The confidence level was set as 90%.

The copy-number profiles of TCGA samples and HNSCC cell lines were visualized using the Integrative Genomics Viewer (IGV, http://www.broadinstitute.org/igv) (version 2.3) and R package OmicCircos (http://bioconductor.org/packages/OmicCircos/).

Whole Exome Sequencing with Illumina HiSeq—Whole Exome DNA sequencing of independently prepared libraries from the same samples of DNA from UM-SCC 1, UM-SCC 6, UM-SCC 11A and UM-SCC 74A were performed using the Illumina HiSeq platform by Dr. Robert J. Morell from the Genomics and Computational Biology Core of NIDCD. The multiplex DNA of four cell line samples were sequenced in one lane utilizing 126 3 126 paired-end sequencing chemistry on the Illumina HiSeq system, and mapped to human NCBI Build 37 reference genome (Hg19) with BWA-mem (http://bio-bwa.sourceforge.net/). The bam files were analyzed with CONTRA v2.0.4 with reference to a pooled baseline control set comprising 13 unrelated human blood samples.

Mutation analysis—Single nucleotide variants (SNVs) were called from exome sequence data by the LifeScope platform(Tang et al., 2008) algorithm using medium call stringency. Small insertions and deletions were detected using the LifeScope Small InDel Tool. SNV and indel data were annotated with gene-base annotation, population frequency, database annotation and predicted effects at the protein level using ANNOVAR (Wang et al., 2010). For prioritization we excluded intronic and synonymous exonic SNVs. We then filtered SNVs based on known single nucleotide polymorphisms (SNPs) from dbSNP (http://www.ncbi.nlm.nih.gov/SNP) v132 and population frequency data from the 1000 Genomes project (http://www.1000genomes.org/), excluding variants with a population minor allele frequency (MAF) > 1%. SNVs were further filtered, retaining variants present in the Catalogue of Somatic Mutations in Cancer (COSMIC) v67 (http://cancer.sanger.ac.uk/cosmic), and/or predicted to have high or medium (H, M) functional impact by MutationAssessor(Reva et al., 2011). The mutations were merged from all cell lines and filtered against the 395 top TCGA mutated genes identified with Mut-Sig (http://www.broadinstitute.org/cancer/cga/mutsig)($p < 0.05$).

RNA-seq expression analysis—Reads were filtered using mapping quality (threshold 10). On average, we obtained about 168 million mapped reads per sample. Read counts of the genes were normalized using DESeq (estimateSizeFactors) R package (version 3.2.0) and the upper quartile normalization method. The expression heatmap was generated using fold changes (FCs) in expression between HNSCC cell line and the mean of three oral keratinocyte lines as normal control. To select highly amplified and expressed genes, we combined significantly amplified genes in both HPV(+) and ( ) HNSCCs from cell lines and TCGA, identified by GISTIC with residual *q*-value of 0.05 as cutoff in cell lines, and confidence level of 99% in TCGA. We then identified the genes with expression significantly correlated with CN by intersecting with CNV-gene-expression significantly correlated genes, as described for Integrative and pathway analysis below. The genes were further filtered by removing those with maximum FC % 2 and less than 3 samples with FC R 1.5. For the heatmap, columns (samples) and rows (gene FCs compared to normal) were hierarchically clustered using Canberra distance and Ward's minimum variance method of agglomeration and supervised to enhance the expression comparison between HPV(−) and HPV(+) lines.

RNA-seq variant detection—To validate SNVs identified in Exome seq, we detected variants with RNA-seq. Raw RNA-seq data in fastq format were aligned to the NCBI human genome build 37 (hg19) and splice junctions database using multi-threading NovoalignCSMPI V1.04. An alignment threshold for polyclonal filter was used (-p 7,15 0.3,15) and quality calibration enabled (-k). Output was in SAM format (-o SAM) and 10 repeat matches for each read (-r A 10) were allowed. Header lines (@SQ lines) and non-aligning reads were removed. The splice-junction coordinates were converted to genomic coordinates by running Useq software SamTranscriptomeParser (http://useq.sourceforge.net/). The resulting alignments were sorted with Picard (http://broadinstitute.github.io/picard/) tool SortSam and the duplicated reads were removed by SAMtools (http://samtools.sourceforge.net/) rmdup. Variants were called with the SAMtools bcftools "view" with settings -vcg. Variants were filtered with vcfutils "varFilter" using default settings except retaining all variants with less than 300 reads. Where sufficient reads enabled detection of variants by RNA-seq and Exome seq, cross-validated SNVs are reported.

HPV detection—To identify HPV reads, color space RNA-seq data in fastq format were aligned to an HPV reference genome with NovoalignCSMPI V1.04 using parameters described in previous section of "RNA-seq variant detection." The HPV reference genome includes genome sequences of 37 HPV subtypes downloaded from NCBI. The HPV virus reads were extracted, filtered and counted using custom Python scripts. Chimeric human and HPV read pairs were identified by aligning the paired end reads in bam format to a combined human and HPV16/18 reference genome, using STAR (http://code.google.com/p/rna-star/).

Comparison of CNV-gene expression association in TCGA and HNSCC cell lines—TCGA mRNA abundance and CNV for 279 tumor specimens (243 HPV(−) and 36 HPV(+) tissue samples) were extracted from Level 3 data associated with the data freeze

(Cancer Genome Atlas Network, 2015). For Cell lines, CN alterations were analyzed by GISTIC, and gene expression was quantified using RNASeq read count normalized with DESeq (estimateSizeFactors) R package and within-sample to a fixed upper quartile of total reads. For gene level analyses, expression values of zero were set to the overall minimum value, and all data were $\log_2$ transformed. To compare CNV-gene expression association in HNSCC tumor specimen and cell lines, first, for each gene, a Spearman coefficient test (one-tailed, greater) was conducted to check for positive correlation between its CNV and expression. The strength of association between CNV and expression for each gene measured as Spearman correlation coefficients (r_s) were recorded. The same tests were carried out in both TCGA and cell lines. Second, to compare the CNV-expression association for all the genes in TCGA and cell lines, another Spearman coefficient test was conducted between Spearman's rho for the common 17,040 genes in TCGA and cell lines. The concordance of the two datasets was analyzed with correlation analysis and visualized with a scatterplot. The concordance between TCGA tumors and cell lines in terms of correlation between CNA and gene expression was also examined and visualized for individual chromosome 1, 2, 3, 4, 5, 6, 10 and 11 (Figure S5).

Expression subtype analysis—Cell lines expression data was log2-transformed, genes with missing values were removed and resulting expression were median centered by gene. Using subtype predictor centroids, each cell line was assigned an expression subtype using a nearest centroid predictor. Not all genes from the prior predictor were available thus the predictor was adjusted from 838 genes to the overlap set of 832. The 832 predictor genes were used to define centroids for each of the four classes of HNSCC. Distance between the centroids was empirically evaluated and visualized using 1 minus the Pearson correlation coefficient for the centroids. Cell lines that were mapped to both Classical and Atypical subtypes ($r > 0.125$) were labeled as Classical/Atypical. Cell lines with low Pearson correlation with all subtypes ($r < 0.125$) were labeled as "No type."

Integrative and pathway analysis—TCGA amplified and deleted genes with 99% confidence intervals in HPV(−) and HPV(+) samples for 279 tumor specimens were extracted from Level 3 data associated with the data freeze (Cancer Genome Atlas Network, 2015). Cell line significantly amplified or deleted genes were identified by GISTIC with residual q-value of 0.05 as cutoff. Merge significantly amplified genes in HPV(−) and (+) and intersect the resulted genes in TCGA and cell lines. The same procedure was done for significantly deleted genes. Merge significantly amplified genes and deleted genes.

CNV-gene-expression significantly correlated genes were selected using these steps: For each gene, a Spearman coefficient test (one-tailed, greater) was used to check for positive correlation between its CNV and log2 transformed expression. P values were adjusted for false discovery rate (FDR) using Benjamini-Hochberg correction. An FDR value of < 0.05 was considered statistically significant. This procedure was carried out in both TCGA and cell lines and the intersection of significantly correlated genes were determined.

CN significantly altered genes intersected with CNV-gene-expression significantly correlated genes in both TCGA and HNSCC cell lines. The genes in the intersection were filtered with their expression in cell lines by comparison to diploid. Genes with 1.5 fold

difference in HPV(−) or HPV(+) respectively were selected and analyzed with Ingenuity Pathway Analysis (IPA) (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis) to determine affected molecular and cellular functions and the enrichment of canonical pathways.

Survival analysis—To compare overall survival time by *TP53* mutation or 3q26.3 amplification events, subjects were categorized as four groups: cases with wild-type (WT) *TP53* and 3q26.3 CN neutral, cases with WT *TP53* and amplified 3q26.3, cases with *TP53* mutation and 3q26.3 CN neutral, cases with *TP53* mutation and 3q26.3 amplification. Five types of non-silent mutations were included in *TP53* mutation: missense, nonsense, frameshift insertion, frameshift deletion and splice-site mutations. The comparison of survival between PIK3CA mutation and 3q26.3 amplification events was carried out in a similar fashion.

The survival analyses were performed with R package. The overall survival curves were obtained using Kaplan-Meier method and were compared using the log-rank test. The Cox proportional hazards model was used to estimate Hazard Ratios (HRs) with 95% Confidence Intervals (CIs).

Comparison of CNAs in recently derived and traditional HNSCC cell lines—To compare CNAs in recently derived and traditional cell lines, we analyzed GISTIC regions with significant alterations using *all_lesions.conf_90.csv* file. The regions with significant alterations in each cell line were compared and visualized with bar chart. Cell lines were grouped by recently-derived or traditional lines and by HPV status, and regions with significant alterations were compared by groups and HPV status using Welch two sample t test and visualized with box and whisker plots.

DATA AND SOFTWARE AVAILABILITY

The accession numbers for the phenotype, exome DNA-seq, and RNA-seq data reported in this paper are SRA: PRJNA453457 and dbGaP: phs001581.

Software tools—All statistical analyses were performed in R software version 3.2.0, and Bioconductor. The OncoPrints and lollipop mutation plots were generated using cBioPortal Tools, OncoPrinter and MutationMapper (Gao et al., 2013). The Bioconductor package ComplexHeatmap (Gu, 2015) was used to draw the mutation heatmap.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank TCGA HNSCC workgroup, especially Alison Taylor of the Broad Institute for providing the sorted IGV display of CN amplifications in TCGA HNSCCs. We thank Andrew Cherniack of the Broad Institute, Vonn Walter of Penn State Milton Hershey Medical Center, Jianjiong Gao of Memorial Sloan Kettering Cancer Center for expert technical assistance and advice, and Medha Bhagwat of the NIH library for technical support. We thank Dr. Robert Morell of the National Institute on Deafness and Other Communication Disorders (NIDCD) Sequencing Core for WES of a subset of lines and controls by Illumina-based methods. This work used the computational resources of

the NIH High-Performance Computing (HPC) Biowulf cluster (http://hpc.nih.gov). This study was supported by the Intramural Research Program (grants ZIA-DC-000073 and ZIA-DC-000074 from the NIH, National Institute on Deafness and Other Communication Disorders).

REFERENCES

Ambrosio EP, Silveira CG, Drigo SA, Sacomano Vde.S., Molck MC, Rocha RM, Domingues MA, Soares FA, Kowalski LP, and Rogatto SR (2013). Chromosomal imbalances exclusively detected in invasive front area are associated with poor outcome in laryngeal carcinomas from different anatomical sites. Tumour Biol 34, 3015–3026. [PubMed: 23749487]

Astanehe A, Arenillas D, Wasserman WW, Leung PC, Dunn SE, Davies BR, Mills GB, and Auersperg N (2008). Mechanisms underlying p53 regulation of PIK3CA transcription in ovarian surface epithelium and in ovarian cancer. J. Cell Sci 121, 664–674. [PubMed: 18270270]

Banerjee R, Russo N, Liu M, Basrur V, Bellile E, Palanisamy N, Scanlon CS, van Tubergen E, Inglehart RC, Metwally T, et al. (2014). TRIP13 promotes error-prone nonhomologous end joining and induces chemoresistance in head and neck cancer. Nat. Commun 5, 4527. [PubMed: 25078033]

Bersani C, Xu LD, Vilborg A, Lui WO, and Wiman KG (2014). Wig-1 regulates cell cycle arrest and cell death through the p53 targets FAS and 14-3-3s. Oncogene 33, 4407–4417. [PubMed: 24469038]

Brenner JC, Graham MP, Kumar B, Saunders LM, Kupfer R, Lyons RH, Bradford CR, and Carey TE (2010). Genotyping of 73 UM-SCC head and neck squamous cell carcinoma cell lines. Head Neck 32, 417–426. [PubMed: 19760794]

Busch CJ, Kriegs M, Laban S, Tribius S, Knecht R, Petersen C, Dikomey E, and Rieckmann T (2013). HPV-positive HNSCC cell lines but not primary human fibroblasts are radiosensitized by the inhibition of Chk1. Radiother. Oncol 108, 495–499. [PubMed: 23932155]

Campbell JD, Yau C, Bowlby R, Liu Y, Brennan K, Fan H, Taylor AM, Wang C, Walter V, Akbani R, et al. (2018). Genomic, pathway network, and immunologic features distinguishing squamous carcinomas. Cell Rep 23, 194–212.e6. [PubMed: 29617660]

Cancer Genome Atlas Network (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 517, 576–582. [PubMed: 25631445]

Cao Y, Bryan TM, and Reddel RR (2008). Increased copy number of the TERT and TERC telomerase subunit genes in cancer cells. Cancer Sci 99, 1092–1099. [PubMed: 18482052]

Chae YK, Ranganath K, Hammerman PS, Vaklavas C, Mohindra N, Kalyan A, Matsangou M, Costa R, Carneiro B, Villaflor VM, et al. (2017). Inhibition of the fibroblast growth factor receptor (FGFR) pathway: the current landscape and barriers to clinical application. Oncotarget 8, 16052–16074. [PubMed: 28030802]

Chiang WF, Hung PS, Liu SY, Yuan TC, Chang KW, Chen YP, Liu YC, and Lin SC (2011). Increase of ZASC1 gene copy number in recurrent oral carcinoma. Oral Dis 17, 53–59. [PubMed: 20646234]

Ciriello G, Cerami E, Sander C, and Schultz N (2012). Mutual exclusivity analysis identifies oncogenic network modules. Genome Res 22, 398–406. [PubMed: 21908773]

Conti MA, Saleh AD, Brinster LR, Cheng H, Chen Z, Cornelius S, Liu C, Ma X, Van Waes C, and Adelstein RS (2015). Conditional deletion of nonmuscle myosin II-A in mouse tongue epithelium results in squamous cell carcinoma. Sci. Rep 5, 14068. [PubMed: 26369831]

Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, and Zarrinkar PP (2011). Comprehensive analysis of kinase inhibitor selectivity. Nat. Biotechnol 29, 1046–1051. [PubMed: 22037378]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Dutt A, Ramos AH, Hammerman PS, Mermel C, Cho J, Sharifnia T, Chande A, Tanaka KE, Stransky N, Greulich H, et al. (2011). Inhibitor-sensitive FGFR1 amplification in human non-small cell lung cancer. PLoS ONE 6, e20351. [PubMed: 21666749]

Edwards ZC, Trotter EW, Torres-Ayuso P, Chapman P, Wood HM, Nyswaner K, and Brognard J (2017). Survival of head and neck cancer cells relies upon LZK kinase-mediated stabilization of mutant p53. Cancer Res 77, 4961–4972. [PubMed: 28760853]

Efazat G, Novak M, Kaminskyy VO, De Petris L, Kanter L, Juntti T, Bergman P, Zhivotovsky B, Lewensohn R, Haag P, and Viktorsson K (2016). Ephrin B3 interacts with multiple EphA receptors and drives migration and invasion in non-small cell lung cancer. Oncotarget 7, 60332–60347. [PubMed: 27533087]

Eytan DF, Snow GE, Carlson S, Derakhshan A, Saleh A, Schiltz S, Cheng H, Mohan S, Cornelius S, Coupar JF, et al. (2016). SMAC mimetic birinapant plus radiation eradicates human head and neck cancers with genomic amplifications of cell death genes FADD and BIRC2. Cancer Res 76, 5442–5454. [PubMed: 27469115]

Fischer M (2017). Census and evaluation of p53 target genes. Oncogene 36, 3943–3956. [PubMed: 28288132]

Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, et al. (2017). COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res 45 (D1), D777–D783. [PubMed: 27899578]

Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM, Handsaker RE, McCarroll SA, O'Donovan MC, Owen MJ, et al. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am. J. Hum. Genet 91, 597–607. [PubMed: 23040492]

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal 6, pl1.

Gollin SM (2014). Cytogenetic alterations and their molecular genetic correlates in head and neck squamous cell carcinoma: a next generation window to the biology of disease. Genes Chromosomes Cancer 53, 972–990. [PubMed: 25183546]

Gross AM, Orosco RK, Shen JP, Egloff AM, Carter H, Hofree M, Choueiri M, Coffey CS, Lippman SM, Hayes DN, et al. (2014). Multi-tiered genomic analysis of head and neck cancer ties TP53 mutation to 3p loss. Nat. Genet 46, 939–943. [PubMed: 25086664]

Gu Z (2015). ComplexHeatmap: Making Complex Heatmaps. R package version 1.6.0, https://github.com/jokergoo/ComplexHeatmap.

Gu Z, Eils R, and Schlesner M (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32, 2847–2849. [PubMed: 27207943]

Hajek M, Sewell A, Kaech S, Burtness B, Yarbrough WG, and Issaeva N (2017). TRAF3/CYLD mutations identify a distinct subset of human papillomavirus-associated head and neck squamous cell carcinoma. Cancer 123, 1778–1790. [PubMed: 28295222]

Herzog A, Bian Y, Vander Broek R, Hall B, Coupar J, Cheng H, Sowers AL, Cook JD, Mitchell JB, Chen Z, et al. (2013). PI3K/mTOR inhibitor PF-04691502 antitumor activity is enhanced with induction of wild-type TP53 in human xenograft and murine knockout models of head and neck cancer. Clin. Cancer Res 19, 3808–3819. [PubMed: 23640975]

Huang X, Godfrey TE, Gooding WE, McCarty KS, Jr., and Gollin SM (2006). Comprehensive genome and transcriptome analysis of the 11q13 amplicon in human oral cancer and synteny to the 7F5 amplicon in murine oral carcinoma. Genes Chromosomes Cancer 45, 1058–1069. [PubMed: 16906560]

Iglesias-Bartolome R, Martin D, and Gutkind JS (2013). Exploiting the head and neck cancer oncogenome: widespread PI3K-mTOR pathway alterations and novel molecular targets. Cancer Discov 3, 722–725. [PubMed: 23847349]

Justilien V, and Fields AP (2009). Ect2 links the PKCiota-Par6alpha complex to Rac1 activation and cellular transformation. Oncogene 28, 3597–3607. [PubMed: 19617897]

Justilien V, Walsh MP, Ali SA, Thompson EA, Murray NR, and Fields AP (2014). The PRKCI and SOX2 oncogenes are coamplified and cooperate to activate Hedgehog signaling in lung squamous cell carcinoma. Cancer Cell 25, 139–151. [PubMed: 24525231]

Kalu NN, Mazumdar T, Peng S, Shen L, Sambandam V, Rao X, Xi Y, Li L, Qi Y, Gleber-Netto FO, et al. (2017). Genomic characterization of human papillomavirus-positive and-negative human squamous cell cancer cell lines. Oncotarget 8, 86369–86383. [PubMed: 29156801]

Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499, 214–218. [PubMed: 23770567]

Leemans CR, Braakhuis BJ, and Brakenhoff RH (2011). The molecular biology of head and neck cancer. Nat. Rev. Cancer 11, 9–22. [PubMed: 21160525]

Li H (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993. [PubMed: 21903627]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R; 1000 Genome Project Data Processing Sub-group (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]

Li J, Lupat R, Amarasinghe KC, Thompson ER, Doyle MA, Ryland GL, Tothill RW, Halgamuge SK, Campbell IG, and Gorringe KL (2012). CONTRA: copy number analysis for targeted resequencing. Bioinformatics 28, 1307–1313. [PubMed: 22474122]

Li H, Wawrose JS, Gooding WE, Garraway LA, Lui VW, Peyser ND, and Grandis JR (2014). Genomic analysis of head and neck squamous cell carcinoma cell lines and human tumors: a rational approach to preclinical model selection. Mol. Cancer Res 12, 571–582. [PubMed: 24425785]

Majumder M, House R, Palanisamy N, Qie S, Day TA, Neskey D, Diehl JA, and Palanisamy V (2016). RNA-binding protein FXR1 regulates p21 and TERC RNA to bypass p53-mediated cellular senescence in OSCC. PLoS Genet 12, e1006306. [PubMed: 27606879]

Mazumdar T, Byers LA, Ng PK, Mills GB, Peng S, Diao L, Fan YH, Stemke-Hale K, Heymach JV, Myers JN, et al. (2014). A comprehensive evaluation of biomarkers predictive of response to PI3K inhibitors and of resistance mechanisms in head and neck squamous cell carcinoma. Mol. Cancer Ther 13, 2738–2750. [PubMed: 25193510]

Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, and Getz G (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 12, R41. [PubMed: 21527027]

Mohan S, Vander Broek R, Shah S, Eytan DF, Pierce ML, Carlson SG, Coupar JF, Zhang J, Cheng H, Chen Z, and Van Waes C (2015). MEK inhibitor PD-0325901 overcomes resistance to PI3K/mTOR inhibitor PF-5212384 and potentiates antitumor effects in human head and neck squamous cell carcinoma. Clin. Cancer Res 21, 3946–3956. [PubMed: 25977343]

Moser R, Xu C, Kao M, Annis J, Lerma LA, Schaupp CM, Gurley KE, Jang IS, Biktasova A, Yarbrough WG, et al. (2014). Functional kinomics identifies candidate therapeutic targets in head and neck cancer. Clin. Cancer Res 20, 4274–4288. [PubMed: 25125259]

Namani A, Matiur Rahaman M, Chen M, and Tang X (2018). Gene-expression signature regulated by the KEAP1-NRF2-CUL3 axis is associated with a poor prognosis in head and neck squamous cell cancer. BMC Cancer 18, 46. [PubMed: 29306329]

Ni Z, Tao K, Chen G, Chen Q, Tang J, Luo X, Yin P, Tang J, and Wang X (2012). CLPTM1L is overexpressed in lung cancer and associated with apoptosis. PLoS ONE 7, e52598. [PubMed: 23300716]

Nix DA, Courdy SJ, and Boucher KM (2008). Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. BMC Bioinformatics 9, 523. [PubMed: 19061503]

Pickering CR, Zhang J, Yoo SY, Bengtsson L, Moorthy S, Neskey DM, Zhao M, Ortega Alves MV, Chang K, Drummond J, et al. (2013). Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers. Cancer Discov 3, 770–781. [PubMed: 23619168]

Ragin CC, Taioli E, Weissfeld JL, White JS, Rossie KM, Modugno F, and Gollin SM (2006). 11q13 amplification status and human papillomavirus in relation to p16 expression defines two distinct etiologies of head and neck tumours. Br. J. Cancer 95, 1432–1438. [PubMed: 17003776]

Reva B, Antipin Y, and Sander C (2011). Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 39, e118. [PubMed: 21727090]

Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, et al. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat. Genet 45, 970–976. [PubMed: 23852170]

Robinson JT, Thorvaldsdo´ttir H, Winckler W, Guttman M, Lander ES, Getz G, and Mesirov JP (2011). Integrative genomics viewer. Nat. Biotechnol 29, 24–26. [PubMed: 21221095]

Saladi SV, Ross K, Karaayvaz M, Tata PR, Mou H, Rajagopal J, Ramaswamy S, and Ellisen LW (2017). ACTL6A is co-amplified with p63 in squamous cell carcinoma to drive YAP activation, regenerative proliferation, and poor prognosis. Cancer Cell 31, 35–49. [PubMed: 28041841]

Scoumanne A, and Chen X (2006). The epithelial cell transforming sequence 2, a guanine nucleotide exchange factor for Rho GTPases, is repressed by p53 via protein methyltransferases and is required for G1-S transition. Cancer Res 66, 6271–6279. [PubMed: 16778203]

Seiwert TY, Zuo Z, Keck MK, Khattri A, Pedamallu CS, Stricker T, Brown C, Pugh TJ, Stojanov P, Cho J, et al. (2015). Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. Clin. Cancer Res 21, 632–641. [PubMed: 25056374]

Tang S, Hyland F, Wessel T, Sorenson J, Peckham H, and De La Vega F (2008). DiBayes: a SNP detection algorithm for next-generation dibase sequencing http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_057817.pdf

Thomas RK, Baker AC, Debiasi RM, Winckler W, Laframboise T, Lin WM, Wang M, Feng W, Zander T, MacConaill L, et al. (2007). High-throughput oncogene mutation profiling in human cancer. Nat. Genet 39, 347–351. [PubMed: 17293865]

Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, and Jemal A (2015). Global cancer statistics, 2012. CA Cancer J. Clin 65, 87–108. [PubMed: 25651787]

Tsunoda K, Tsujino I, Koshi R, Sugano N, Sato S, and Asano M (2016). Nicotine-mediated Ca(2+)-influx induces IL-8 secretion in oral squamous cell carcinoma cell. J. Cell. Biochem 117, 1009–1015. [PubMed: 26418512]

Van Dyke DL, Worsham MJ, Benninger MS, Krause CJ, Baker SR, Wolf GT, Drumheller T, Tilley BC, and Carey TE (1994). Recurrent cytogenetic abnormalities in squamous cell carcinomas of the head and neck region. Genes Chromosomes Cancer 9, 192–206. [PubMed: 7515662]

Van Waes C, and Musbahi O (2017). Genomics and advances towards precision medicine for head and neck squamous cell carcinoma. Laryngoscope Investig. Otolaryngol 2, 310–319.

Wang K, Li M, and Hakonarson H (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164. [PubMed: 20601685]

Wang W, Wei Z, Lam T-W, and Wang J (2011). Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. Sci. Rep 1, 55. [PubMed: 22355574]

Wang PY, Ma W, Park JY, Celi FS, Arena R, Choi JW, Ali QA, Tripodi DJ, Zhuang J, Lago CU, et al. (2013). Increased oxidative metabolism in the Li-Fraumeni syndrome. N. Engl. J. Med 368, 1027–1032. [PubMed: 23484829]

Wang PY, Li J, Walcott FL, Kang JG, Starost MF, Talagala SL, Zhuang J, Park JH, Huffstutler RD, Bryla CM, et al. (2017). Inhibiting mitochondrial respiration prevents cancer in a mouse model of Li-Fraumeni syndrome. J. Clin. Invest 127, 132–136. [PubMed: 27869650]

White JS, Weissfeld JL, Ragin CC, Rossie KM, Martin CL, Shuster M, Ishwad CS, Law JC, Myers EN, Johnson JT, and Gollin SM (2007). The influence of clinical and demographic risk factors on the establishment of head and neck squamous cell carcinoma cell lines. Oral Oncol 43, 701–712. [PubMed: 17112776]

Wilkerson MD, and Hayes DN (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 26, 1572–1573. [PubMed: 20427518]

Yamaguchi K, Wu L, Caballero OL, Hibi K, Trink B, Resto V, Cairns P, Okami K, Koch WM, Sidransky D, and Jen J (2000). Frequent gain of the p40/p51/p63 gene locus in primary head and neck squamous cell carcinoma. Int. J. Cancer 86, 684–689. [PubMed: 10797291]

Yang A, and McKeon F (2000). P63 and P73: P53 mimics, menaces and more. Nat. Rev. Mol. Cell Biol 1, 199–207. [PubMed: 11252895]

Yang X, Lu H, Yan B, Romano RA, Bian Y, Friedman J, Duggal P, Allen C, Chuang R, Ehsanian R, et al. (2011). DNp63 versatilely regulates a Broad NF-kB gene program and promotes squamous epithelial proliferation, migration, and inflammation. Cancer Res 71, 3688–3700. [PubMed: 21576089]

Highlights

- Genomic and transcriptomic landscape of 26 HNSCC lines compared with 279 TCGA tumors
- Copy gains of chromosomes 3q, 5p, and 11q drive increased oncogene expression
- Concurrent 3q26.3 amplification and TP53 mutation are associated with worse survival
- Cell lines recapitulate genomic alterations of more aggressive HNSCC tumor subtypes

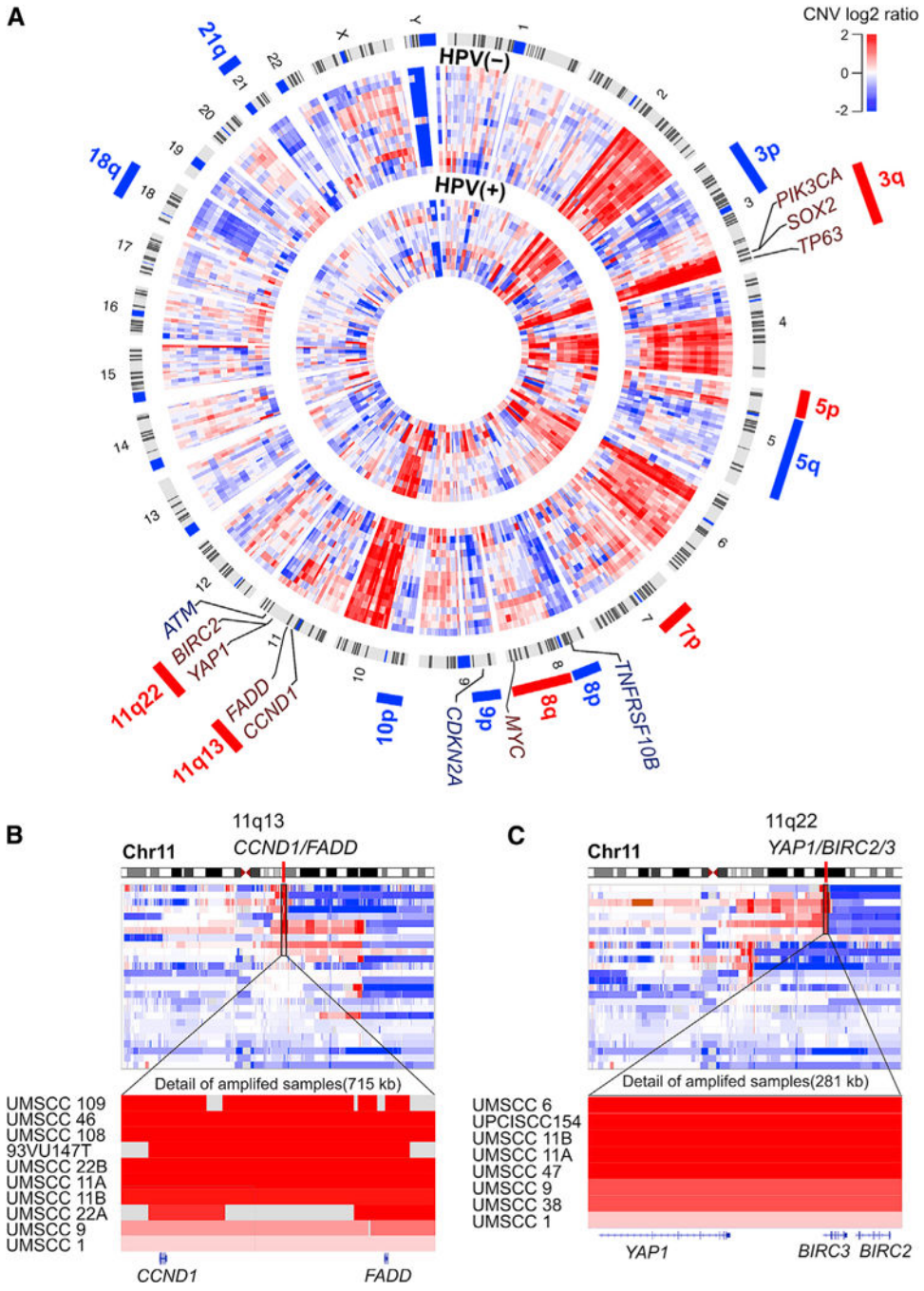


Figure 1. DNA CN Alterations of HNSCC HPV(−) and HPV(+) Cell Lines Highlighting Several Alterations Observed in TCGA Datasets

(A) micCircos plot showing landscape of CNAs analyzed using GISTIC for HPV(−) (n = 15) and HPV(+) (n = 11) HNSCC cell lines. Chromosomes displaying cytobands are plotted on the outer ring, with the inner rings showing segmented CNAs of 15 HPV(−) lines (outer band) and 11 HPV(+) lines (inner band). Regions of recurrent CN gains (red) or losses (blue) and selected genes observed in TCGA are indicated. See Figure S2A for detailed GISTIC analysis results and Table S2 for common CNAs in cell lines and TCGA. (B and C) Modified Integrative Genomics Viewer (IGV) plots illustrating ordered CNAs for chr 11q13

region (B) for *CCND1* and *FADD* or 11q22 region (C) for *YAP1* and *BIRC2/3*. Red, amplification; blue, deletion; white, neutral.

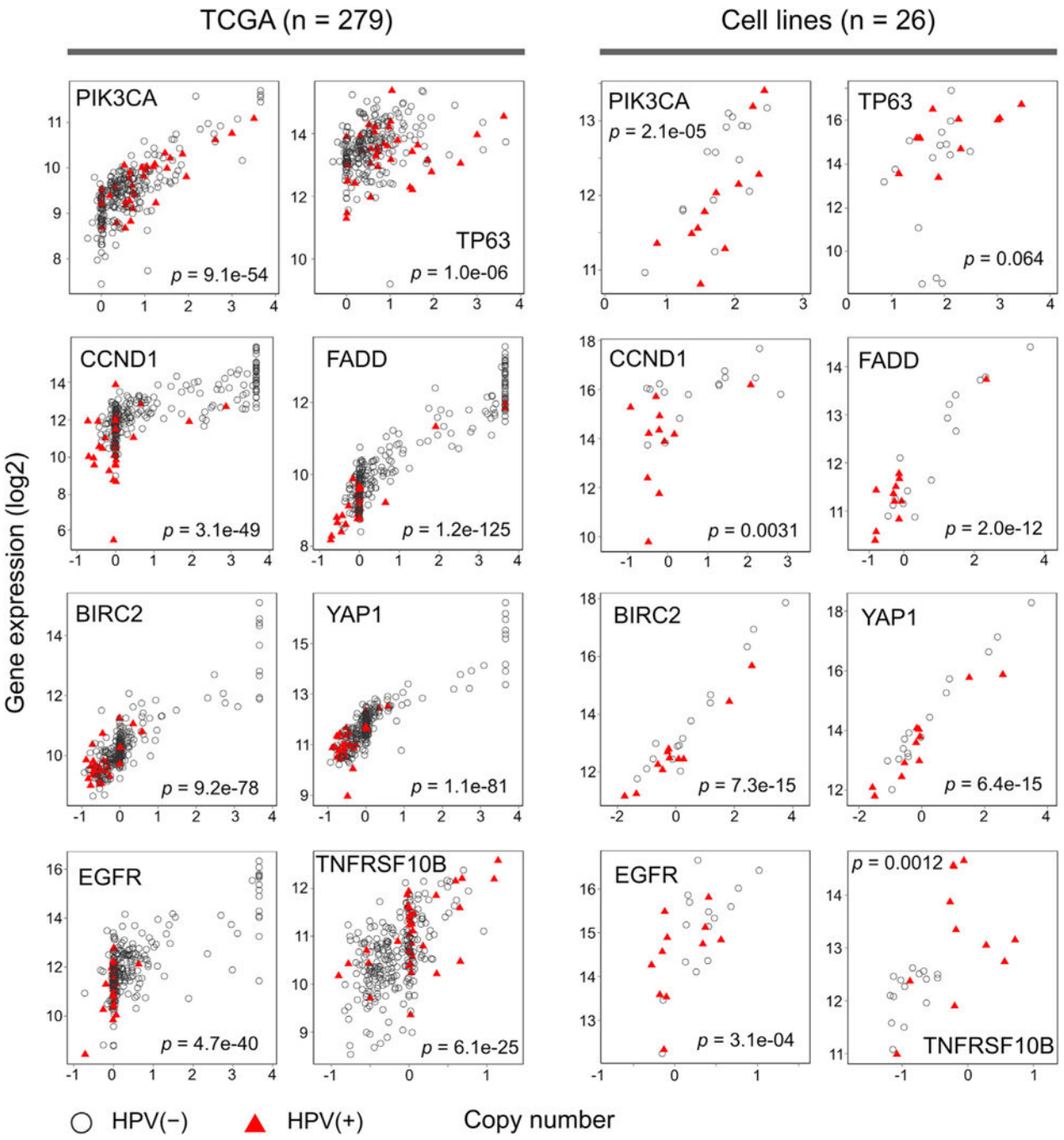


Figure 2. Correlation between CNA and mRNA Expression Observed in TCGA Datasets and HNSCC Cell Lines

Scatterplots for selected genes showing significant CNA and gene expression correlation in both HNSCC TCGA datasets (left, n = 279) and cell lines (right, n = 26). X axis, $\log_2$ CN ratio; 0 is diploid, 1 is one-copy loss, 0.585 is one-copy gain, and values larger than 1 are amplifications. Y axis, $\log_2$ gene expression from normalized RNA-seq read count. Spearman coefficient test, with p values presented. HPV(+), red triangles; HPV(−), gray circles. See Figure S5 for concordance between TCGA and cell lines in the correlation of CNV and gene expression.

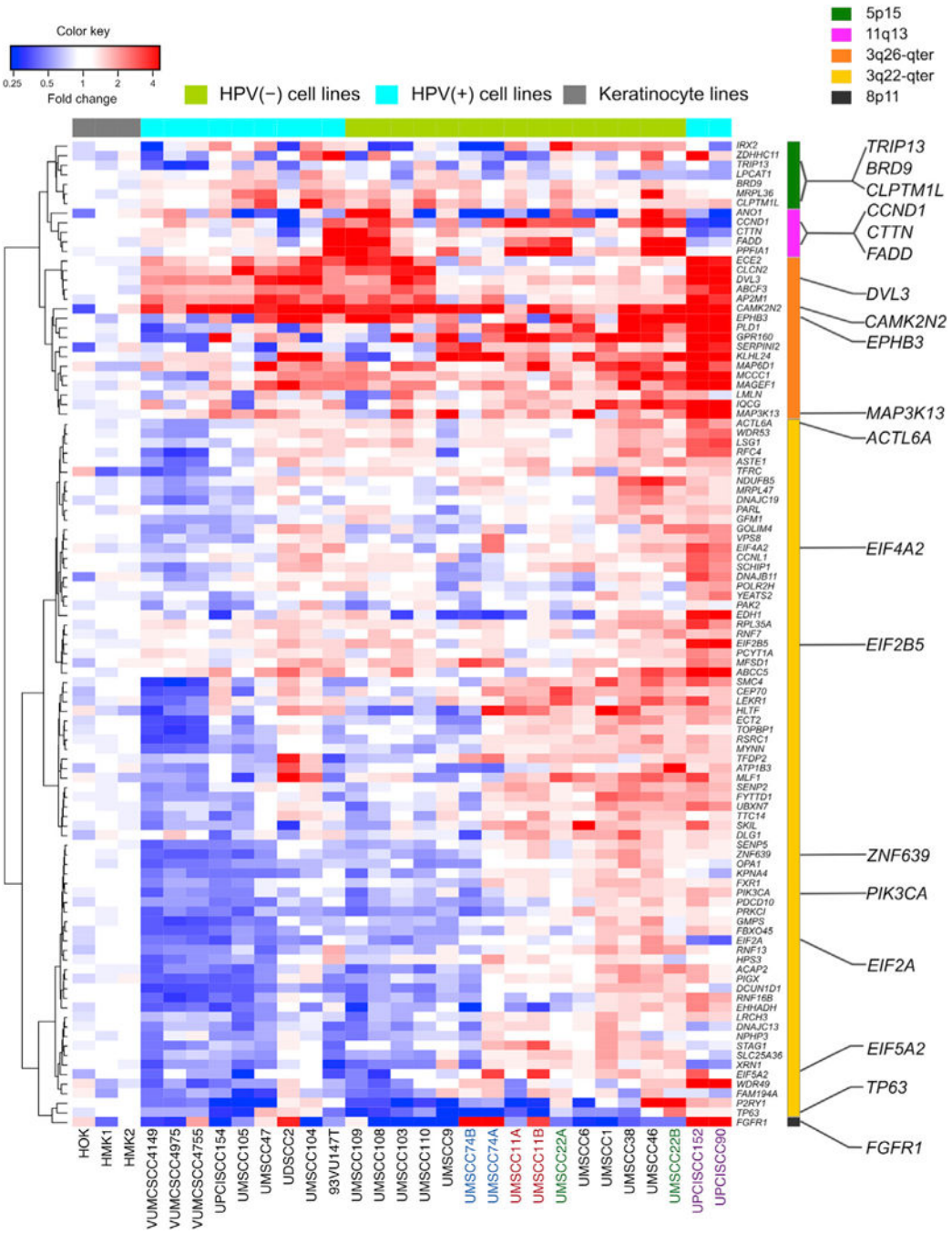


Figure 3. Hierarchically Clustered Heatmap of Significantly Amplified and Expressed Genes in HNSCC Cell Lines and TCGA Datasets

Heatmap of supervised hierarchical clustering from RNA-seq data shows expression significantly correlated with amplification for top 103 concordant genes in 26 HNSCC cell lines and 279 tumors from TCGA dataset. Columns, HNSCC cell lines; rows, genes; top, HPV status; key and colored bar on the right, chr regions. Fold expression compared with normal: red, increased; blue, decreased); white, no change. Cell lines with colored font represent paired cell lines derived from the same patient, with different color for each pair.

See Figure S6 for workflow of the integrated analysis and Table S3A for annotation of genes identified in this analysis.

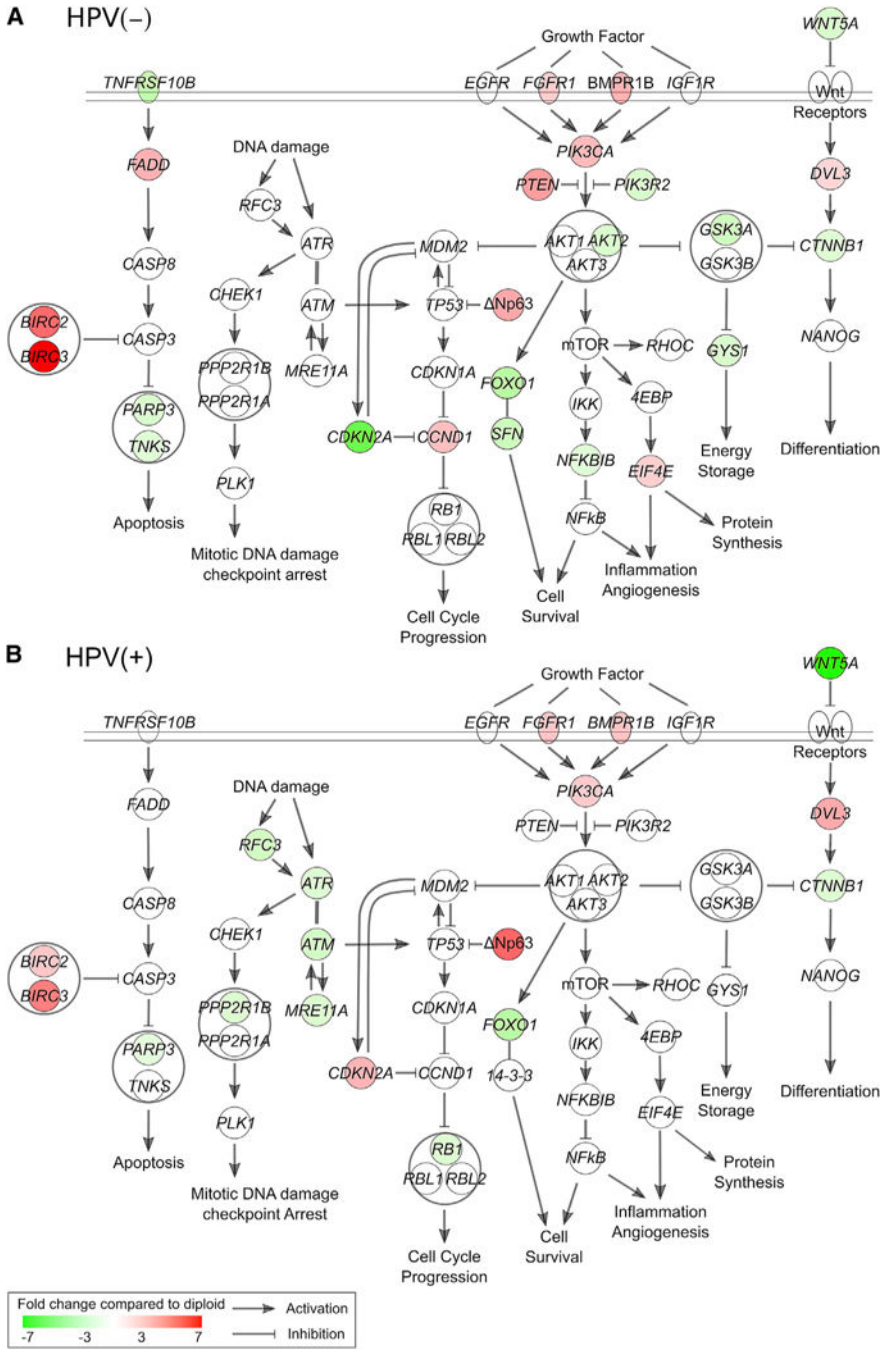


Figure 4. Integrated Analysis Reveals Genomic and Expression Alterations in Signaling Pathways in HNSCC Differing in HPV Status

See Figure S6 for workflow of analysis and Table S4 for detailed pathway analysis results. (A and B) The key affected pathways, components, and cellular functions in (A) HPV(−) and (B) HPV(+) cell lines are presented.

The altered genes were analyzed for corresponding pathways by Ingenuity Pathway Analysis (IPA), and the workflow is summarized in Figure S6. Additional information is summarized in Table S4. Color key, fold expression compared with diploid: red, increased;

green, decreased. Function, activation (arrows), inhibition (block). Large circles indicate biological complex, and ellipses indicate transmembrane receptors.

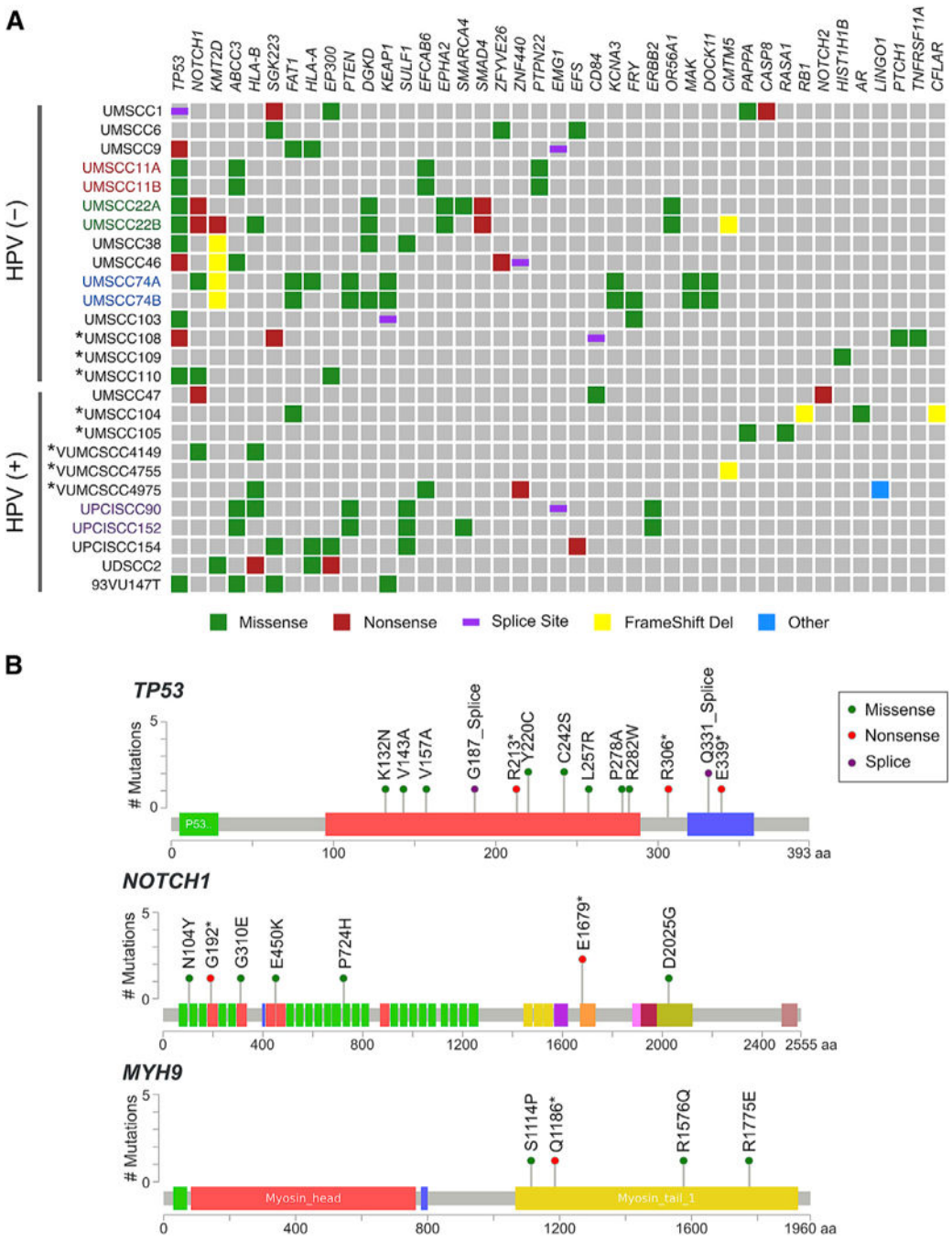


Figure 5. Mutated Genes in 26 HNSCC Cell Lines Consistent with TCGA Datasets

(A) Frequently mutated genes (columns) are presented by mutation frequency, as well as additional genes of biological interest in HNSCC. Cell lines (rows, n = 26) are arranged by HPV status. Cell lines with colored font represent paired cell lines derived from the same patient, with different color for each pair. Cell lines annotated with asterisks had matched blood DNA for comparison. Color shading indicates mutation types: green, missense; red, nonsense; purple, splice site; yellow, frameshift deletion; blue, other mutations. See Figure S8 for mutation detection process and Tables S5 and S6A for detailed mutation results. (B)

Lollipop plots show the distribution and classes of mutations in TSGs: *TP53*, *NOTCH1*, and *MYH9* across HNSCC cell lines. The predicted impact of mutations detected is shown by transcript base position and functional domain, where color of the lollipop represents mutation types corresponding to the mutation classes in (A). Colored bars represent protein domains or motifs. The y axis indicates the observed number of mutations in 26 HNSCC cell lines. See Figure S9 for mutation diagrams of other genes implicated in HNSCC.

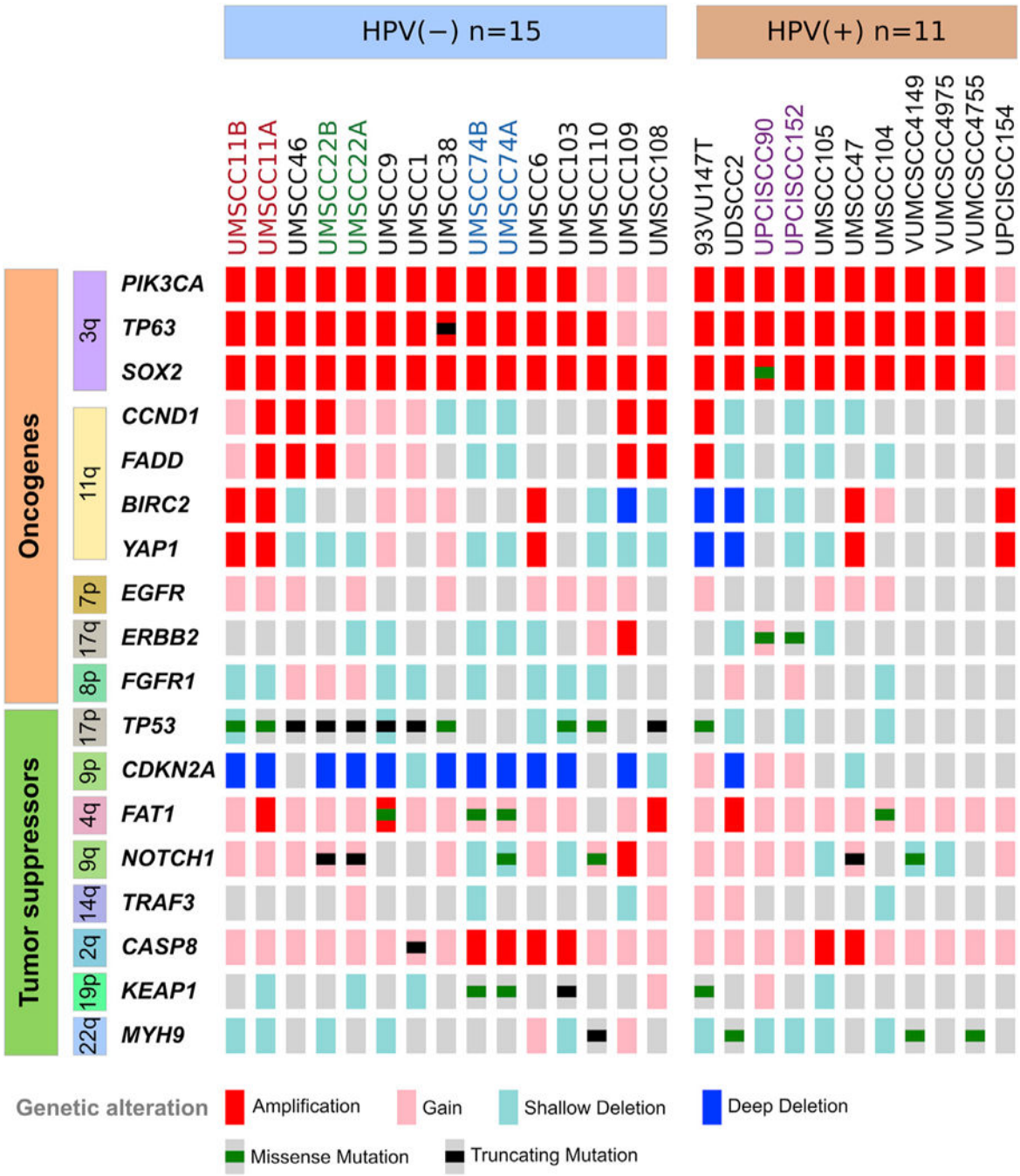


Figure 6. HNSCC Cell Lines Recapitulate the Spectrum of CNAs and Mutations Found in HNSCC Tumor Subtypes by TCGA

Integrated summary of chr and oncogene amplification and TSG mutation and deletion for 26 HNSCC cell lines are illustrated by an oncoprint. Cell lines are displayed in columns and grouped by HPV status. Cell lines with colored font represent paired cell lines derived from the same patient, with different color for each pair. Genes with chr location are displayed in rows and grouped by putative function (oncogenes and tumor suppressors) and similarity of alteration patterns. Copy gains: red, high-level; pink, low-level; blue, homozygous deletions;

aqua, heterozygous deletions; green bar, mutations, missense; black bar, truncating mutations.

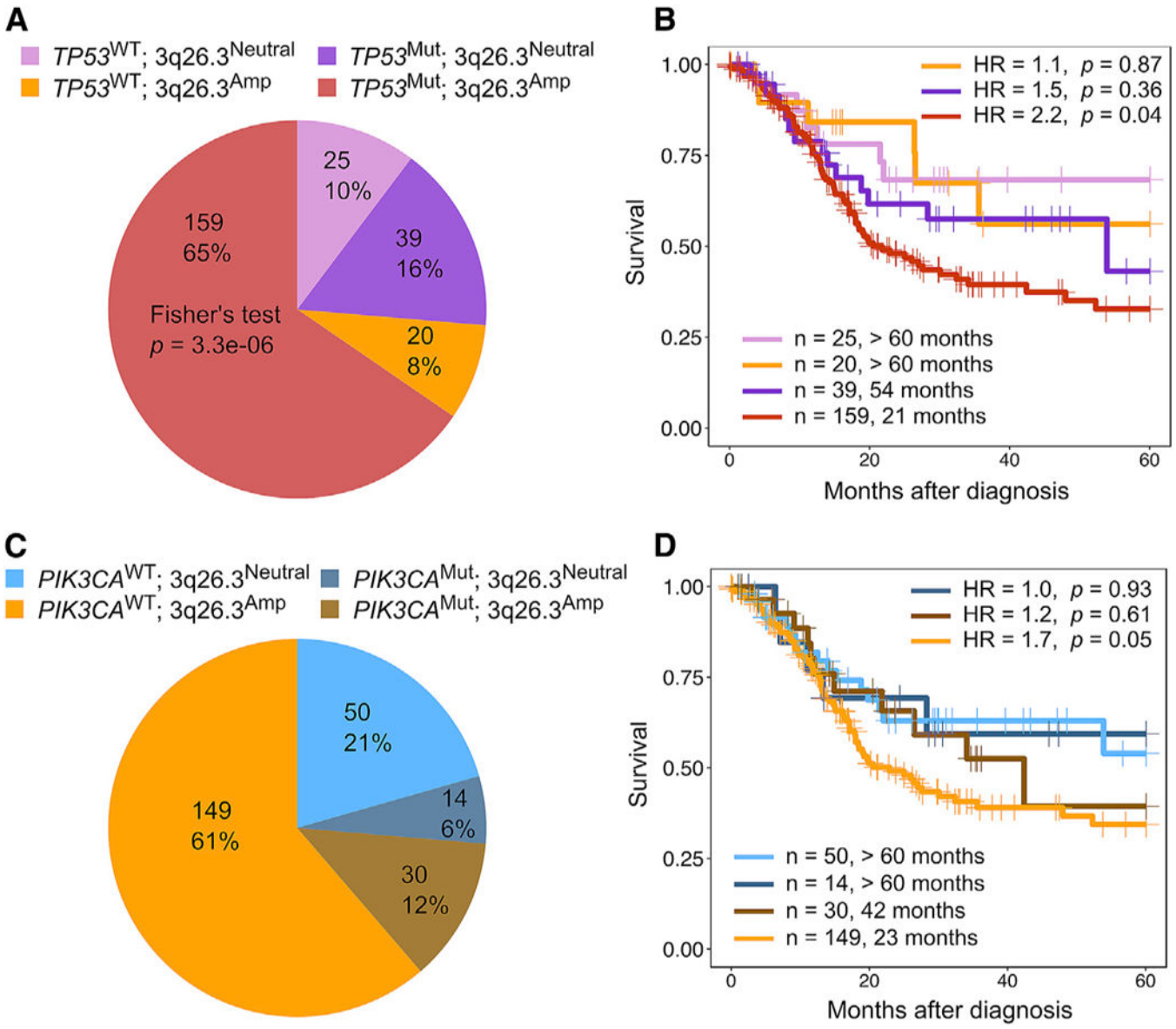


Figure 7. Co-occurrence of TP53 Mutation and 3q26.3 Amplification Is Associated with Poor Survival in HPV(−) HNSCC Patients from TCGA Datasets

See Table S3B for annotation of amplified gene on chr 3q26.3.

(A) The pie chart represents the percentage of HNSCC patients with co-occurrence of *TP53* mutation (*TP53*^{Mut}) and 3q26.3 amplification (3q26.3^{Amp}), either alteration alone or none of the event. A significant co-occurrence was observed between *TP53* mutation and 3q26.3 amplification (p = 3.3e-06, Fisher's exact test).

(B) Overall survival outcome for four combinations of 3q26.3 amplification and *TP53* mutation events in patients by Kaplan-Meier analysis (colors correspond to the case subsets in A). P values were determined using a log rank test.

(C) Distribution of patients with different combinations of 3q26.3^{Amp} and *PIK3CA* mutation.

(D) Assessment of the overall survival outcome for four combinations of 3q26.3 amplification and *PIK3CA* mutation events in HNSCC patients.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Raw, processed and clinical data	TCGA Network	https://portal.gdc.cancer.gov/legacy-archive/search/f
TCGA publication page	TCGA Network	https://gdc.cancer.gov/about-data/publications/hnsc_2014
TCGA publication page	TCGA Network	https://gdc.cancer.gov/about-data/publications/mc3-2017
COSMIC Census	Forbes et al., 2017	https://cancer.sanger.ac.uk
FireBrowse portal	Broad Institute	http://gdac.broadinstitute.org
cBioPortal	Memorial Sloan Kettering Cancer Center	http://www.cbioportal.org
dbSNP	NCBI dbSNP	http://www.ncbi.nlm.nih.gov/SNP
1000 Genomes Project	The International Genome Sample Resource (IGSR)	http://www.internationalgenome.org/
Critical Commercial Assays		
SOLiD 5500	ThermoFisher	https://www.thermofisher.com/us/en/home/brands/product-brand/5500-series-genetic-analysis-systems.html
Deposited Data		
Phenotype, Exome DNASeq and RNASeq data (deposited in the Short Read Archive (SRA: http://www.ncbi.nlm.nih.gov/sra/) of NCBI under BioProject)	This paper	SRA: PRJNA453457; dbGaP: phs001581
Experimental Models: Cell Lines		
UM-SCC cells	Brenner et al., 2010	N/A
VUMC-SCC cells	Provided by Dr. Wendell Yarbrough	N/A
UPCI:SCC cells	Busch et al., 2013	N/A
UD-SCC cells	Busch et al., 2013	N/A
VU cells	Busch et al., 2013	N/A
Software and Algorithms		
GISTIC2	Mermel et al., 2011	http://software.broadinstitute.org/cancer/software/genepattern/modules/docs/GISTIC_2.0

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Integrative Genomics Viewer	Robinson et al., 2011	http://gdac.broadinstitute.org/igv
MutSig	Lawrence et al., 2013	http://software.broadinstitute.org/software/cprg/?q=node/35
LifeScope v.2.5	LifeScope Genomic Analysis Solutions	https://www.thermofisher.com/search/results?query=lifescope&refinementAction=true&persona=DocSupport&focusarea=Search%20All
CONTRA	Li et al., 2012	http://contra-cnv.sourceforge.net/
ANNOVAR	Wang et al., 2010	http://annovar.openbioinformatics.org/en/latest/
OmicCircos	N/A	http://bioconductor.org/packages/release/bioc/html/OmicCircos.html
MutationAssessor	Reva et al., 2011	http://mutationassessor.org/r3/
NovoalignCSMPI	Wang et al., 2011	http://www.novocraft.com/documentation/novoaligncs-3/
STAR: ultrafast universal RNA-seq aligner	Dobin et al., 2013	https://code.google.com/archive/p/rna-star/
USeq	Nix et al., 2008	http://useq.sourceforge.net/
SAMtools	Li et al., 2009	http://samtools.sourceforge.net/
SAMtools: bcftools	Li, 2011	http://samtools.sourceforge.net
ComplexHeatmap	Gu et al., 2016	http://bioconductor.org/packages/devel/bioc/html/ComplexHeatmap.html
Ingenuity Pathway Analysis	N/A	https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/
ConsensusClusterPlus R package	Wilkerson and Hayes, 2010	https://bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html
Survival R package	N/A	https://cran.r=project.org/package=survival