

RESEARCH ARTICLE

Multiple Routes to Oncogenesis Are Promoted by the Human Papillomavirus–Host Protein Network

Manon Eckhardt^{1,2,3}, Wei Zhang⁴, Andrew M. Gross⁴, John Von Dollen^{1,3}, Jeffrey R. Johnson^{1,2,3}, Kathleen E. Franks-Skiba^{1,3}, Danielle L. Swaney^{1,2,3,5}, Tasha L. Johnson³, Gwendolyn M. Jang^{1,3}, Priya S. Shah^{1,3}, Toni M. Brand⁶, Jacques Archambault⁷, Jason F. Kreisberg^{4,5}, Jennifer R. Grandis^{5,6}, Trey Ideker^{4,5}, and Nevan J. Krogan^{1,2,3,5}

ABSTRACT

We have mapped a global network of virus–host protein interactions by purification of the complete set of human papillomavirus (HPV) proteins in multiple cell lines followed by mass spectrometry analysis. Integration of this map with tumor genome atlases shows that the virus targets human proteins frequently mutated in HPV[−] but not HPV⁺ cancers, providing a unique opportunity to identify novel oncogenic events phenocopied by HPV infection. For example, we find that the NRF2 transcriptional pathway, which protects against oxidative stress, is activated by interaction of the NRF2 regulator KEAP1 with the viral protein E1. We also demonstrate that the L2 HPV protein physically interacts with the RNF20/40 histone ubiquitination complex and promotes tumor cell invasion in an RNF20/40-dependent manner. This combined proteomic and genetic approach provides a systematic means to study the cellular mechanisms hijacked by virally induced cancers.

SIGNIFICANCE: In this study, we created a protein–protein interaction network between HPV and human proteins. An integrative analysis of this network and 800 tumor mutation profiles identifies multiple oncogenesis pathways promoted by HPV interactions that phenocopy recurrent mutations in cancer, yielding an expanded definition of HPV oncogenic roles. *Cancer Discov*; 8(11): 1474–89. ©2018 AACR.

¹Quantitative Biosciences Institute (QBI), UCSF, San Francisco, California. ²Gladstone Institute of Data Science and Biotechnology, J. David Gladstone Institutes, San Francisco, California. ³Department of Cellular and Molecular Pharmacology, UCSF, San Francisco, California. ⁴Department of Medicine, UCSF, La Jolla, California. ⁵The Cancer Cell Map Initiative (CCMI), UCSF and UCSD, San Francisco and La Jolla, California. ⁶Department of Otolaryngology–Head and Neck Surgery, UCSF, San Francisco, California. ⁷Department of Microbiology and Immunology, McGill University, Montreal, Quebec, Canada.

Note: Supplementary data for this article are available at Cancer Discovery Online (<http://cancerdiscovery.aacrjournals.org/>).

M. Eckhardt and W. Zhang contributed equally to this article.

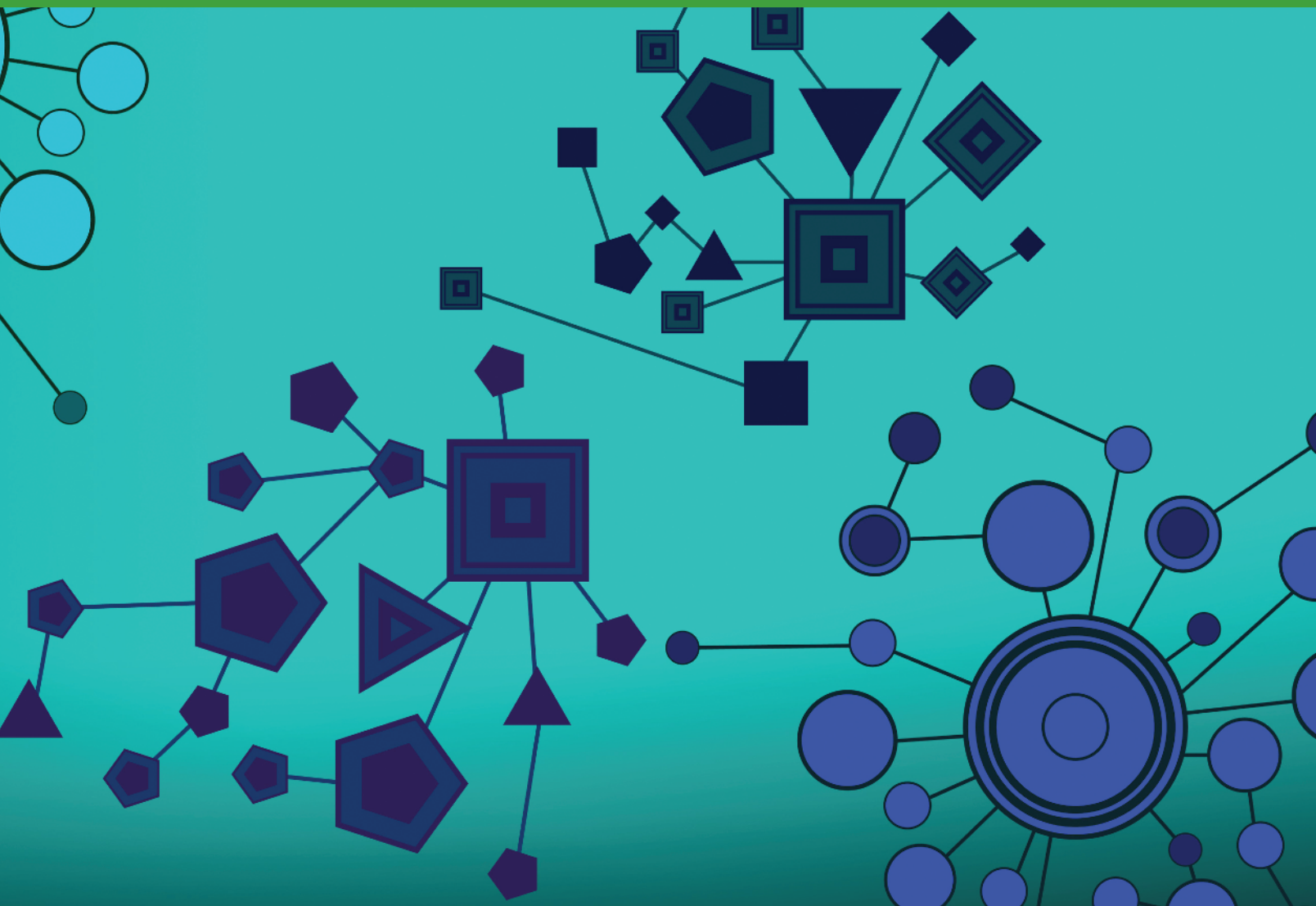
All data and code are publicly available. HPV–Human PPI raw data are available online at the Chorus project (project ID: 1529). Custom code for

identifying mutated genes in each tumor, assessing difference in mutation rates by HPV status, and network propagation are publicly available through GitHub (https://github.com/wzhang1984/Eckhardt-Zhang_HPV).

Corresponding Authors: Nevan J. Krogan, University of California, San Francisco, 1700 4th Street, BH 308D, San Francisco, CA 94158. Phone: 415-476-2980; Fax: 415-514-9736; E-mail: nevan.krogan@ucsf.edu; and Trey Ideker, University of California, San Diego, School of Medicine, 9500 Gilman Drive, MC 0688, La Jolla, CA 92093-0688. Phone: 858-534-3578; Fax: 858-822-4246; Email: tideker@ucsd.edu

doi: 10.1158/2159-8290.CD-17-1018

©2018 American Association for Cancer Research.



INTRODUCTION

Infection with human papillomavirus (HPV) is associated with a variety of cancers, including nearly all cases of cervical squamous cell carcinoma (CESC) and the majority of oropharyngeal cancers, a subset of head and neck squamous cell carcinomas (HNSCC; ref. 1). CESC is among the most common cancers in the developing world, accounting for 7.5% of all female cancer deaths, and HPV⁺ HNSCC is projected to outnumber cases of CESC by the year 2020 (2). Although vaccines that protect against HPV infection are becoming available (3), these measures are not yet widespread, especially in men, and they do not address the large number of currently infected patients with cancer. To treat these patients as well as HPV⁻ versions of similar cancers, a thorough understanding of the molecular events underlying HPV infection and tumor initiation is needed.

The HPV proteins E6 and E7 have long been appreciated as drivers of tumorigenesis in HPV-related cancers, with potential roles for E5 also proposed (4). Both E6 and E7 exert oncogenic activity through protein–protein interactions (PPI) with tumor

suppressor proteins. For example, E6 binds to and promotes ubiquitination of p53, leading to its subsequent degradation (5). Similarly, E7 binds to and promotes the inactivation of RB1, thus leading to cell-cycle progression through activation of E2F-driven transcription (6). Beyond p53 and RB1, a number of other tumor suppressors and proto-oncogenes have been established as viral targets, suggesting that establishment of infection and tumorigenesis share a set of common pathways (7–9). Focusing on these known oncogenic viral proteins, previous studies have used affinity purification–mass spectrometry (AP-MS) or yeast two-hybrid methods to construct catalogs of HPV–human PPI networks (9–12). However, a complete interactome between all HPV and human proteins is still lacking, and the functions of most HPV–human PPIs with respect to their oncogenic roles are still unknown.

Another major source of information about virally induced tumors comes from DNA sequence. Comprehensive tumor genome analysis has revealed that HPV⁺ and HPV⁻ cancers from the same tissue can have markedly different incidences of mutation in particular genes (13). For example, the gene

encoding p53, *TP53*, is mutated in 82% of smoking-related HPV⁻ HNSCC cases but only 3% of HPV⁺ cases (13). The tumor suppressor gene *CDKN2A* and several members of the oxidative stress pathway, including *CUL3*, *KEAP1*, and *NFE2L2*, also show significantly increased mutation rates in HPV⁻ tumors (13). However, with the exception of *TP53*, the mechanistic basis for these discrepancies is not clear.

Here, we develop an integrated strategy based on protein network mapping and tumor genome analysis to define the set of oncogenic interactions between virus and host. First, we apply AP-MS (14–20) to map the HPV–human PPI network in multiple cell lines. These data are then combined with the somatic mutation profiles of HPV⁺ and HPV⁻ tumors using the technique of network propagation, leading to the identification of previously unappreciated oncogenic pathways promoted by HPV–human PPIs. This overall approach, whereby viral proteins are initially used as proteomic probes, followed by comparing the resulting data to cancer genomic data sets through network propagation, has the potential to uncover the biology underlying other virally induced cancers.

RESULTS

Systematic Mapping of HPV–Human PPIs

To systematically map the physical interactions between HPV and human proteins, we individually cloned codon-optimized versions of all 9 viral genes into expression vectors containing a 2x Strep-tag (Fig. 1A). We chose the well-established model of HPV-31 as a high-risk HPV strain with all constructs overexpressed individually (Supplementary Notes), initially in the cervical cell line C33A (21), a widely used model system to study HPV infection (Supplementary Notes; Supplementary Fig. S1). Streptavidin-coated beads were used for AP to enrich for human proteins associated with viral proteins, and MS was used to detect interacting proteins (Fig. 1B). Our MiST software (MS interaction STatistics; refs. 15, 22) was used to assign a quantitative confidence score to each putative PPI, a system that takes into account the abundance, reproducibility, and specificity of the copurifying proteins. To select a cutoff that would provide a set of high-confidence interactions, we benchmarked our data against 27 well-characterized HPV–human PPIs (Supplementary Table S1; Methods). At a MiST cutoff of 0.75, we detected 18 of these PPIs (Fig. 1C), including the well-characterized oncogenic interactions between E6 and p53 and between E7 and RB-like proteins. Of note, we did not detect an interaction between E7 and RB1 (6) as C33A cells express a truncated and unstable RB1 protein (23, 24). Application of this threshold (MiST > 0.75) to the entire data set identified 137 high-confidence interactions across all 9 HPV proteins (Supplementary Table S2), with E7 showing the highest number of interactions (Fig. 1D).

Initial functional analysis of the HPV–human protein interaction network revealed significant enrichment for human proteins involved in protein processing, chromatin remodeling, cell proliferation, and host defense (Fig. 1E–F; Supplementary Table S3). For example, E5-interacting proteins were enriched for cellular components of the endoplasmic reticulum (Fig. 1E), consistent with the localization of E5 to

the endoplasmic reticulum (25). E1-interacting proteins were enriched for protein deubiquitination (Fig. 1F), in line with the requirement of this process for HPV DNA replication (26). We also ran an enrichment analysis for known human protein complexes, in which we highlighted complexes for which the virus directly binds multiple subunits (Fig. 2). For example, E1 interacts with proteins in the USP1/UAF1 deubiquitinating enzyme complex (26, 27), whereas E7 interacts with members of the DREAM complex, including RB-like proteins RBL1 and RBL2 (28). E6 interacts not only with the tumor suppressor p53 but also with members of the ubiquitin–proteasomal system, including six proteasomal subunits and the ubiquitin ligase E6AP (encoded by *UBE3A*; refs. 5, 29). Such strong enrichment for biological functions and protein complexes is indicative of a high-quality protein interaction network.

Validation in Multiple Cell Types

We next sought to validate the network detected in C33A (human cervical cancer cells) by generating parallel interaction maps in two additional cellular contexts: HEK293 (human embryonic kidney cells; Fig. 2, blue nodes), which have been used for previous studies of host–pathogen PPIs (30), and Het-1A (human esophageal epithelial cells; Fig. 2, orange nodes), a commonly used nonneoplastic model for normal head and neck mucosa (31). We detected 405 and 84 high-confidence interactions in HEK293 and Het-1A cells, respectively. Viral protein expression levels were lower in Het-1A cells compared with the others, resulting in the detection of a smaller number of high-confidence interactions (Supplementary Fig. S1; Supplementary Table S2). Notably, the majority of high-confidence interactions that had been originally detected in C33A (80/137) could be validated in either one or both additional contexts (Fig. 2; Supplementary Fig. S2). It is worth noting that the highly pathogenic interaction between E7 and RB1 was detected in both HEK293 and Het-1A cell lines.

HPV–Human PPIs Phenocopy Recurrent Genetic Alterations in Cancer

Our next goal was to determine which HPV–human interactions were most directly involved in cancer. For this purpose, we adopted an integrative strategy combining the complete HPV–human interactome with the genomic mutational landscape of tumors. Genomic data were obtained from 177 cervical squamous cell carcinomas and endocervical adenocarcinomas (CESC) and 505 HNSCC from The Cancer Genome Atlas (TCGA; refs. 13, 32) along with 118 HNSCC samples from the University of Chicago (33), yielding a combined cohort of 295 HPV⁺ and 505 HPV⁻ tumor samples (Fig. 3A). The genes affected by either single-nucleotide alterations or copy-number alterations (CNA) in each sample were determined based on subtractive analysis of paired tumor and normal whole-exome sequences or single-nucleotide polymorphism (SNP) arrays, respectively.

From these mutation calls, we used logistic regression to identify genes for which genetic alteration is strongly dependent on tumor HPV status. In particular, we identified eight genes that are altered frequently in HPV⁻ tumors but rarely in HPV⁺ tumors: *TP53*, *CDKN2A*, *CDKN2B*, *C9orf53*, *CCND1*,

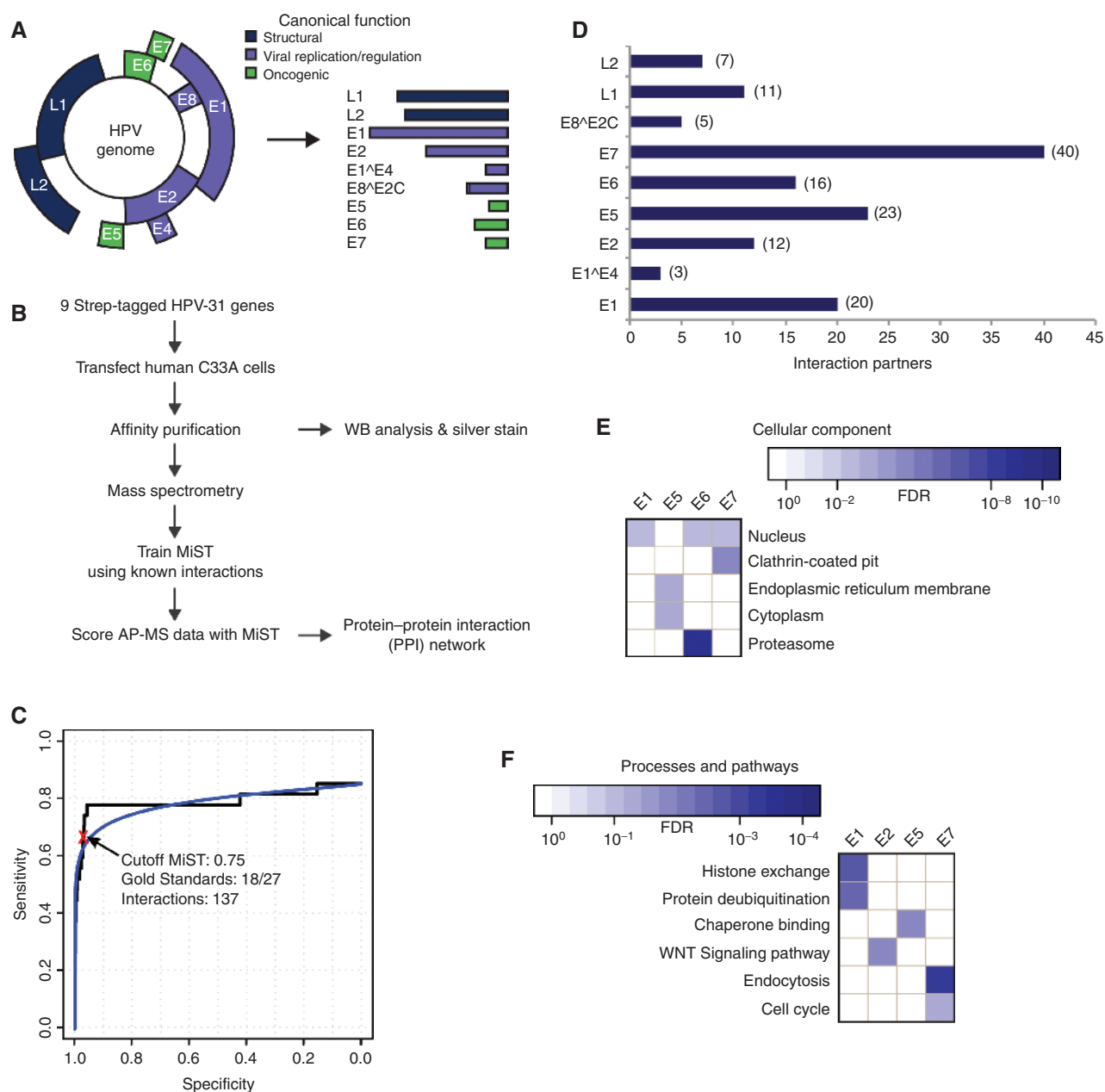


Figure 1. Generation of an HPV-human PPI map. **A**, Representation of the HPV genome, as well as the individual viral proteins used in this study (to scale), colored according to their canonical function during the viral replication cycle. **B**, Summary of the workflow used to construct the PPI network. **C**, Receiver operating characteristic curve illustrating MiST prediction accuracy of a set of 27 gold-standard interactions (see Supplementary Table S1) with the curve inflection point marked with a red "X". The threshold of MiST 0.75 recalled 18 gold-standard interactions and a total of 137 high-confidence interactions (see also Supplementary Table S2). **D**, Number of human proteins interacting with each individual viral protein. **E** and **F**, Heat maps representing manually curated enriched cellular components (**E**) and biological pathways and processes (**F**) of human proteins interacting with individual HPV bait proteins (see also Supplementary Table S3). Colors represent statistical significance as indicated by the respective color scale.

FAT1, *NOTCH2*, and *EGFR* (Fig. 3A; FDR < 10%; Supplementary Table S4). The frequent genetic alterations in these genes suggest their functions are under selection as cancer drivers (34), whereas the mutual exclusivity (35, 36) of these alterations with HPV, a type of epistatic genetic interaction, suggests the virus is able to hijack the same or related pathways. Providing immediate support for this idea, we found that two

of the eight genes, *TP53* and *CDKN2A*, encoded tumor suppressors that were also bound physically by HPV proteins in our interaction data. That is, these two genes were involved in both genetic and physical interactions directly with the virus. One of these, *TP53*, recovered a well-known route of viral infection and tumor development (5). The other tumor suppressor gene, *CDKN2A*, had not been previously reported to

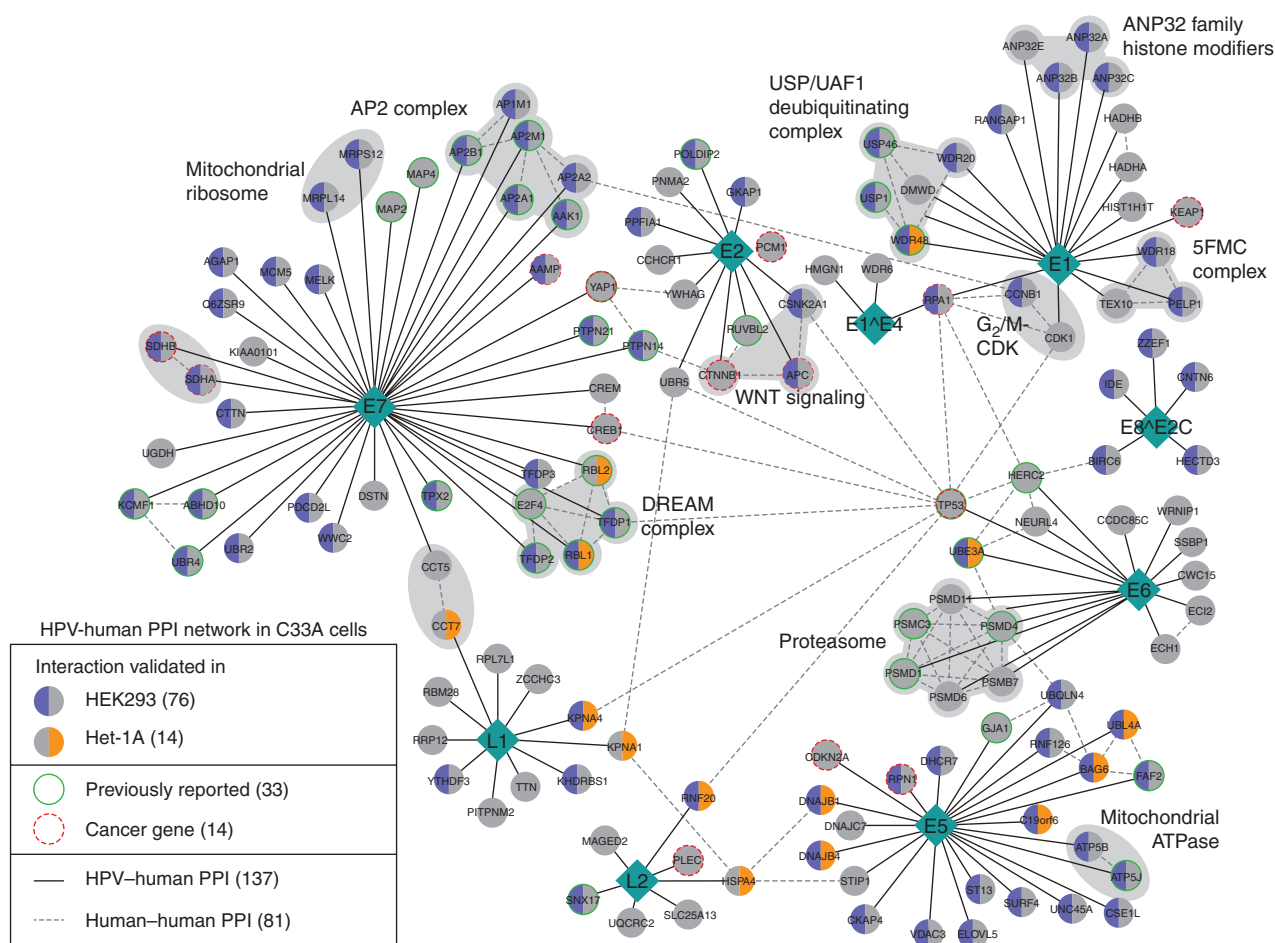


Figure 2. HPV-human protein network map. Network representation of the HPV-human interaction map in C33A cells (<http://doi.org/10.18119/N9J53T>; see also Supplementary Table S2). Diamond-shaped nodes represent the 9 individually expressed HPV-31 proteins, and circles represent the 137 high-confidence interactors. HPV-human interactions are depicted in solid black lines, and manually curated human-human interactions are shown with broken gray lines. Select human complexes are highlighted by gray shadows and annotation. Interactors are colored based on detection in HEK293 (blue), Het-1A (orange), or both (see also Supplementary Table S2; Supplementary Figs. S1 and S2). Green rings around interactors indicate previously described HPV-human PPIs (see also Supplementary Table S1). Cancer genes are highlighted by a broken red ring around the respective human prey node.

interact with HPV proteins. The observed interaction between HPV E5 and a protein product of *CDKN2A*, p16-INK4a, may affect its tumor-suppressive role in the cell cycle (37), phenocopying the recurrent inactivating mutations seen in this gene in HPV⁺ cancers.

To alleviate concerns of some viral genes being expressed at a lower level in HPV⁺ cancers, we examined HPV mRNA expression in TCGA patient samples. Based on RNA-sequencing (RNA-seq) data, the majority of HPV genes are expressed in most of the patients with HPV⁺ CESC (Supplementary Fig. S3a). Of note, E1 is expressed at comparable levels to the oncogene E6 (median RPKM = 240.7 vs. 241.8, respectively).

Common Pathways Affected by Tumor Mutation and Viral Interaction

Motivated by these observations, we reasoned that tumors and viruses might alter the same pathways even in cases where the implicated genes were different. To more broadly understand these common pathways, we downloaded the

Reactome functional interaction (ReactomeFI) reference network (38), a public catalog of 229,300 known pathway relationships among human proteins, including PPIs, protein-gene transcriptional regulatory interactions, and metabolic reactions. We then used the framework of network propagation (39) to score proteins based on their proximity within the ReactomeFI map to HPV-interacting proteins or proteins preferentially mutated in HPV⁺ tumors (Fig. 3B; Methods). In particular, the MiST scores without threshold, characterizing the confidence of HPV protein binding, and the deviances, characterizing the significance of preferential HPV⁺ mutations, were propagated separately across the network. In each case, the effect of network propagation was to spread the starting data type across the ReactomeFI network neighborhood (Methods). Next, the two propagated scores were integrated as a single value for each protein, representing its network proximity to both lines of evidence. This combined score was then compared with what was expected by chance, based on 20,000 permutations in which the scores of proteins were randomly interchanged while retaining the

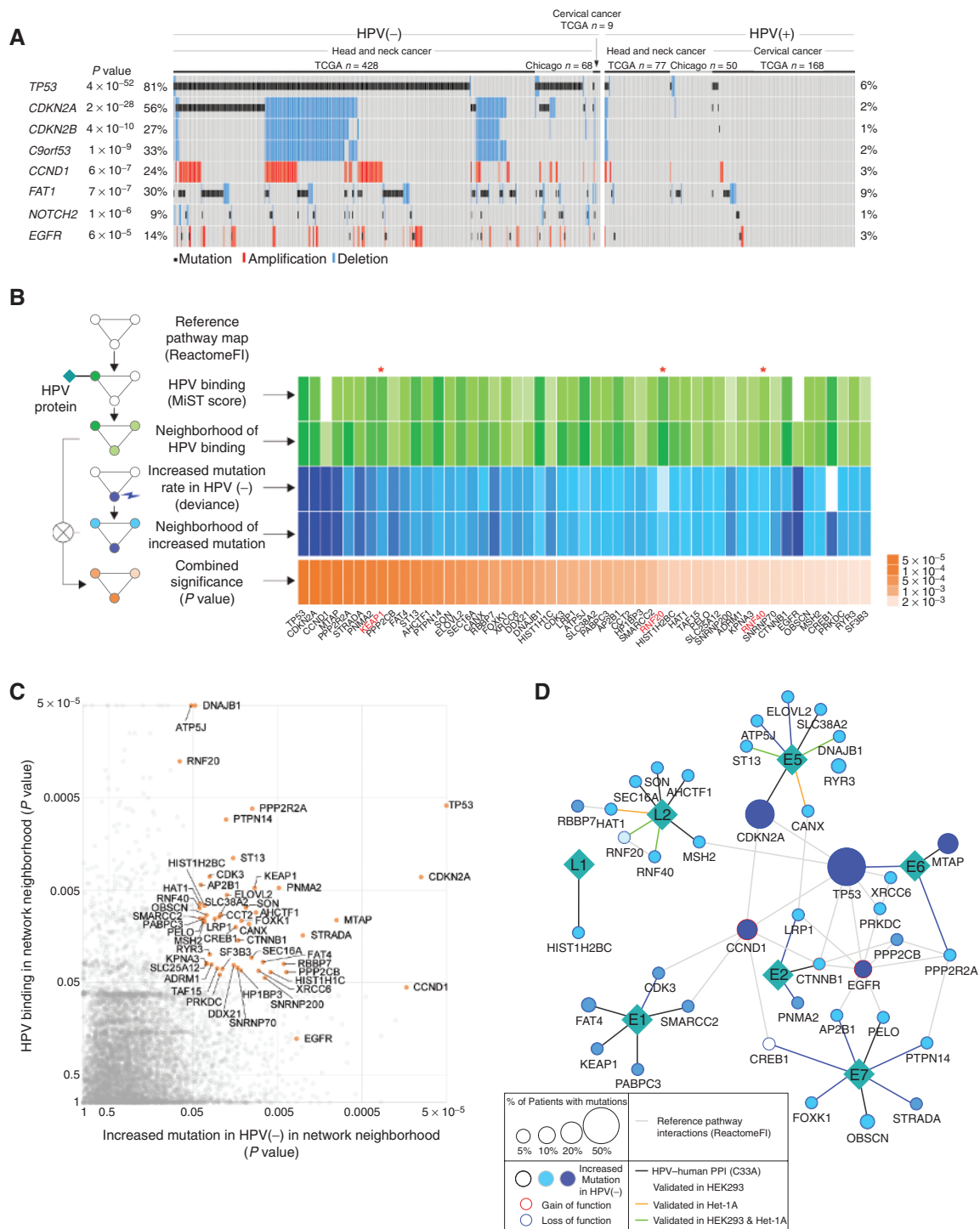


Figure 3. HPV-binding and recurrent genetic alterations complement each other in carcinogenesis. **A**, Genes with significantly increased mutation rates in HPV⁻ tumors are shown with types of mutations present shown for each tumor (see also Supplementary Table S4). Columns are ordered based on HPV status, cancer type, and cohort. **B**, Network propagation is used to identify the network neighborhoods that are enriched in both HPV interactors and an increased mutation rate in HPV⁻ tumors. The original and network propagated scores for each gene are indicated by color intensity, with green representing HPV interaction scores (MiST), blue representing differential mutation scores (D , deviance) and orange representing combined significance (P value). **C**, Scatterplot depicting the empirical P values of MiST scores and differential mutation scores after network propagation. Empirical P values were derived from 20,000 permutations of the HPV interaction scores and the differential mutation scores. Orange (labeled) nodes are genes whose neighborhoods are enriched in both events ($n = 51$; combined significance with empirical FDR < 25%; see Methods and Supplementary Table S5). **D**, Network regions extracted from ReactomeFI by network propagation, containing genes with combined significance (FDR < 25%) together with relevant HPV-human interactions (MiST > 0.68; see also Supplementary Tables S2 and S5). Graphic keys including node and edge colors, node size, and node border color are indicated in the graphical figure legend. Figures 4A and 5A are similarly annotated.

network structure (Methods). This analysis identified 51 human proteins with significant pathway proximity to both proteins with preferential HPV⁻ mutations and those with strong physical binding to HPV proteins (Fig. 3B and C; Supplementary Table S5, FDR < 25%).

The ReactomeFI interactions among these 51 proteins, together with relevant HPV–host interactions from our PPI data, define an integrated pathway map (Fig. 3D). This map depicts how each of the proteins emerges as a top hit: For instance, not only does cyclin D1 (encoded by *CCND1*) have a high differential mutation rate, it is near other proteins encoded by genes with high differential mutation rates (*TP53*, *CDKN2A*, and *EGFR*) and also near proteins with direct physical interactions with HPV (E6–p53, E5–p16, E2– β -catenin). As another example, β -catenin (encoded by *CTNNB1*) not only is an interesting target of HPV E2, but it also lies in common pathways with several differentially mutated genes (*TP53*, *CCND1*, and *EGFR*).

E1-KEAP1 Interaction Phenocopies Inactivating Mutations in the KEAP1–NRF2 Pathway

The above analyses identified several pathways that contain both physical and genetic interactions with HPV but were not previously implicated in viral oncogenesis. These findings led us to focus on two pathways in particular. First, we investigated KEAP1–NRF2 signaling, which governs a major cytoprotective response to stress that can also be activated by somatic mutation during tumorigenesis (40–43). Under basal conditions, KEAP1 binds to NRF2 and acts as a substrate receptor for the CUL3 ubiquitin ligase complex, leading to subsequent degradation of NRF2 by the proteasome. In response to oxidative stress, KEAP1 loses its inhibitory function on NRF2, which is thus able to translocate to the nucleus and transcriptionally activate gene promoters containing antioxidant response elements (ARE; refs. 44, 45).

In our pathway analysis, KEAP1 was assigned a high combined significance score due to strong physical interaction with the HPV E1 protein (C33A MiST 0.81) and preferential mutation in HPV⁻ cancers in multiple genes in its pathway neighborhood ($P = 2 \times 10^{-4}$, Fig. 3B). These genes were *NFE2L2* (encoding NRF2) and *CUL3*, for which the odds of mutation were 2.6-to-1 and 3.3-to-1 in HPV⁻ and HPV⁺ cancers, respectively, as well as *KEAP1* itself with 7.4-to-1 odds (Fig. 4A). Surprisingly, none of these three genes had met the significance cutoff for increased mutation frequency when analyzed individually (Fig. 3A), due to the strong FDR correction required for multiple hypothesis testing of all genes.

To further explore the HPV–KEAP1 association, we first confirmed E1 expression in a panel of HPV⁺ cell lines (Supplementary Fig. S3b). Next, we validated that the physical

interaction of KEAP1 with HPV E1 could be confirmed by immunoprecipitation (IP)–western blotting (WB) analysis using both streptavidin-coated beads targeting the Strep-tag (Fig. 4B) or an antibody against the endogenous KEAP1 protein for IP (Fig. 4C), and confirmed that E1 from HPV-16 also binds to KEAP1 (Fig. 4D). Importantly, we also validated the E1–KEAP1 interaction in the HPV⁺ head and neck cancer cell line UD-SCC-2 (hypopharynx; Supplementary Fig. S4). Furthermore, we found that KEAP1 binds at the N-terminus of E1 as demonstrated by decreased interaction with both a C-terminal fragment of E1, as well as further decreased binding to a mutant lacking the p80/UAF1-binding domain (E1 Δ 10–40) in C33A cells (Fig. 4E and F). To determine the functional consequences of the E1–KEAP1 interaction, we tested for NRF2 activity in the presence or absence of E1. Using an NRF2 reporter assay, we found that expression of HPV E1, but not an unrelated control gene (GFP), increased ARE-dependent transcription of luciferase (Fig. 4G). These results suggest that HPV E1 inactivates this pathway through a direct PPI between E1 and KEAP1, thus leading to the expression of cytoprotective target genes (Fig. 4H).

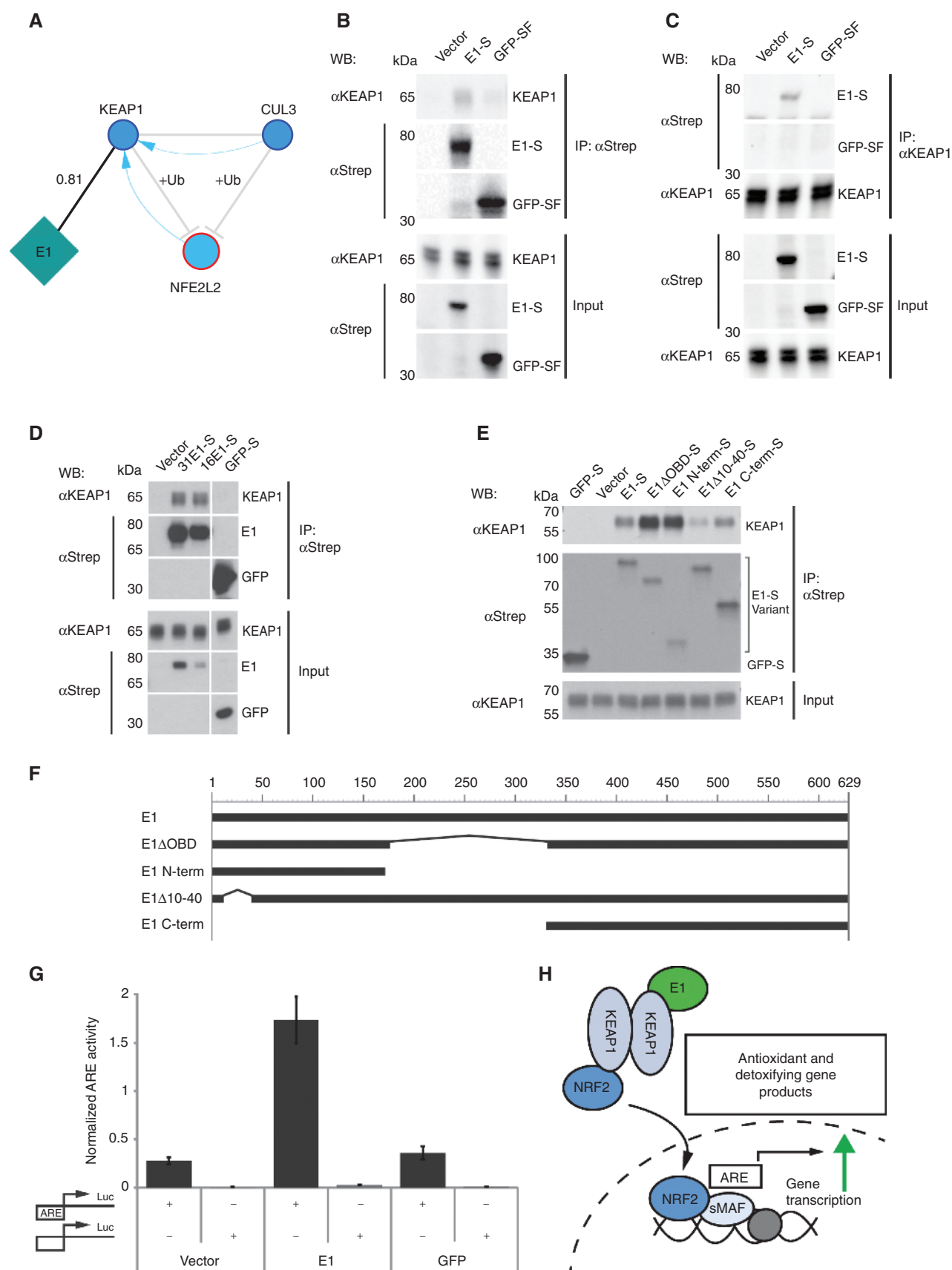
RNF20/40 Histone Modifiers: Targets of HPV L2 versus Mutational Inactivation

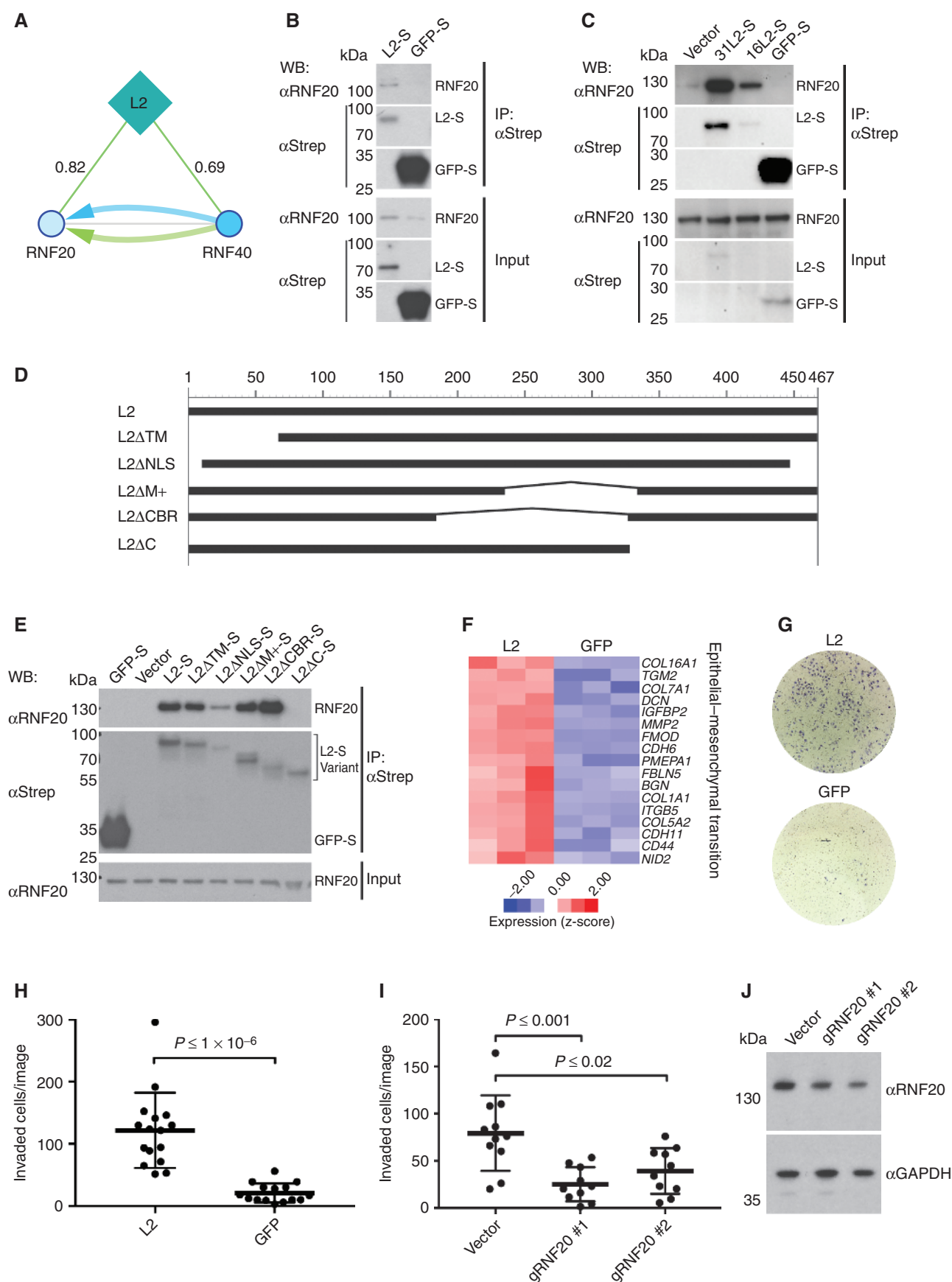
The second pathway we investigated was a chromatin modification pathway governed by RNF20 and RNF40, which form a heterodimeric ubiquitin ligase complex that monoubiquitinates histone H2B at lysine 120. This “H2Bub1” histone mark induces an open chromatin structure that is accessible to transcription factors and DNA repair factors (46–48) and plays a role in tumor suppression (49). Accordingly, decreased activity of the RNF20/40 complex is associated with increased invasiveness and tumor progression (49, 50), presumably related to the epithelial–mesenchymal transition (EMT; ref. 51).

In our pathway analysis, both RNF20 and RNF40 were assigned highly significant combined *P* values ($P = 0.0016$ and 0.0018 , respectively; Fig. 3B). This significance was achieved because in all three cell lines tested HPV L2 was found to bind RNF20 and, to a more variable extent, RNF40 (Fig. 2; MiST scores of 0.69, 0.85, and 0.98 in C33A, Het-1A, and HEK293 cells, respectively). Moreover, *RNF20* and *RNF40* together had increased odds of mutation in HPV⁻ versus HPV⁺ cancers, although these odds were substantially greater for *RNF40* (6.4-to-1) than for *RNF20* (3.4-to-1; Fig. 5A). The known cooperativity between these two proteins in ReactomeFI meant that the entire complex could be associated with HPV by combining the physical and genetic evidence.

To investigate the impact of HPV L2 on RNF20-dependent gene regulation, we first confirmed L2 expression in a panel

Figure 4. Functional consequences of the E1–KEAP1 interaction. **A**, Subnetwork of interaction between HPV E1 and KEAP1. Blue arrows denote the main contribution of differential mutation scores to KEAP1 through network propagation. MiST score of virus–human interaction in C33A cells is indicated (0.81). The mechanism of interaction is represented by edge shapes, with +Ub indicating ubiquitination. **B–E**, Western blot analysis of the virus–host interaction by IP using streptavidin-coated beads to bind the Strep-tag (**B**, **D**, **E**) or an antibody against endogenous KEAP1 (**C**). Proteins indicated on the right of each blot from IP and input samples were detected using the antibodies indicated. Bands were cropped from the same original membrane. Distinction between different strains of HPV is made using strain number preceding the respective viral bait (HPV-31 = 31, HPV-16 = 16; **D**). **F**, Cartoons depicting the truncation/deletion mutants of HPV-31 E1. Scale is provided in amino acid numbers of the full-length protein at the top. **G**, Luciferase reporter assay for the antioxidant response element (ARE). Relative luciferase light units were normalized based on transfection control. The mean \pm standard deviation of technical triplicates are depicted. **H**, Model depicting how E1 influences the KEAP1–NRF2 pathway.





of HPV⁺ cell lines (Supplementary Fig. S3b). Next, the physical interaction between RNF20 and L2 was confirmed in Het-1A (Fig. 5B) and C33A cells (Fig. 5C) using IP followed by WB analysis. We demonstrated this interaction to be conserved with HPV-16 L2 (Fig. 5C) and mapped the interaction between L2 and RNF20 in 293T cells. Although previous reports have shown nuclear localization and tethering of L2 to chromatin to be dependent on a central region within L2 [chromosome-binding region (CBR); ref. 52], we show that RNF20 does not bind to this region of L2, but rather to the C-terminus of L2 (Fig. 5D and E). Next, we performed RNA-seq analysis of cells expressing L2 or GFP as a control. In total, 305 upregulated and 253 downregulated genes were found in response to L2 expression (Supplementary Table S6, FDR < 5%). Focusing on the upregulated genes, we found that EMT and tumor invasiveness were among the most highly enriched pathways (FDR < 10⁻³; Fig. 5F; Supplementary Table S6). Key genes in these pathways included extracellular matrix components (collagens: *COL16A1*, *COL7A1*, *COL1A1*, and *COL5A2*; integrin: *ITGB5*), regulators such as *IGFBP2*, and genes regulating TGFβ signaling (*PMEPA1* and *FMOD*; Fig. 5F). These findings suggest a link between HPV L2 and RNF20/40-mediated induction of EMT and tumor invasion.

We next sought to examine in more detail the functional relationship between L2 expression in human cells and invasive cell migration. Overexpression of HPV-31 L2 in the human esophagus cell line Het-1A led to a statistically significantly higher percentage of cells invading through a Matrigel matrix in a Transwell invasion assay (Fig. 5G and H; $P \leq 1.0 \times 10^{-6}$, two-tailed *t* test). Additionally, HPV⁺ UD-SCC2 cells with CRISPR-mediated polyclonal disruptions of RNF20 showed a decrease in invasion (Fig. 5I and J). These results suggest that HPV L2 expression allows cells to more efficiently invade the local microenvironment through its interaction with RNF20. Finally, consistent with these data, we also showed a decrease in the invasive phenotype of HPV⁺ HNSCC cells when HPV-16 L2 was targeted by a CRISPR guide RNA (Supplementary Notes and Supplementary Fig. S5).

DISCUSSION

This work describes the first systematic HPV-human interactome combined with an integrative genomics approach to recognize interactions that drive viral oncogenesis. As

the majority of this interactome has not been previously described, this data set creates a powerful resource of HPV-human connections for future study and hypothesis building. It is also, to our knowledge, the first HPV-human interaction map generated across multiple human cellular contexts, enabling the further study of how pathway architecture is modulated by cell type while providing strong validation for the majority of interactions that persist despite these contextual changes.

Our HPV-human interactome recovered the expected interactions linking HPV to cancer, including RB-E7 and p53-E6, as well as previously unrecognized connections such as interactions between E7 and YAP1 (Supplementary Notes), as well as E2 and proteins of the WNT/β-catenin signaling pathway (Fig. 2). Dysregulation of WNT signaling is a hallmark of cancer (53, 54) and has been recently shown to play a role in the maintenance of cancer stem cells (55). Although previous reports link the classic HPV oncogenes E6 and E7 to dysregulation of WNT signaling (56), a role for E2 has not been well appreciated, although it is corroborated by the observation that E2 expression levels affect this pathway (57). Interestingly, the E2 gene is often lost upon HPV genome integration (58, 59) but can be maintained episomally in cancer cells (60).

We have shown that integrating this HPV-human interactome with tumor genome data, focusing on the genes that are recurrently mutated in HPV⁻ but not HPV⁺ tumors, constitutes a powerful approach to identifying proteins that serve as both viral targets and drivers of cancer. Moreover, we also observed that this integrated proteome/genome analysis can be expanded from the level of proteins to pathways. An interesting example highlighted by this type of analysis relates to interactions between proteins of HPV and the host *CDKN2A-TP53-CCND1* pathway. The tumor suppressor gene *CDKN2A* encodes multiple proteins through alternative splicing, including p16-INK4a and p14ARF. p16 inhibits G₁-S cyclin-dependent kinase (CDK) complexes containing cyclin D1 (encoded by *CCND1*; ref. 37), whereas p14 activates p53 through the inhibition of MDM2 (61). This pathway is further regulated through the p53-mediated transcription of p21, which again inhibits the activity of cyclin D1-CDK complexes (37). In HPV⁻ patients, the inactivation of tumor suppressors *CDKN2A* and p53 thus releases two key inhibitions on the activity of cyclin D1. In HPV⁺ patients, the virus achieves similar effects, in which the well-established E6-p53 interaction leads to proteasomal degradation of p53, and

Figure 5. Functional consequences of the L2-RNF20 interaction. **A**, Subnetwork of HPV L2 interacting with RNF20 and RNF40. Green/blue arrows represent the main contribution of MiST/differential mutation scores through network propagation, respectively. MiST scores of virus-human interaction in C33A cells is indicated (0.82, 0.69). **B, C, E**, Western blot analysis of virus-host interaction by IP using streptavidin-coated beads to bind the Strep-tag. Proteins indicated on the right of each blot from IP and input samples were detected using the antibodies indicated. Bands were cropped from the same original membrane. Distinction between different strains of HPV is made using strain number preceding the respective viral bait (HPV-31 = 31, HPV-16 = 16). **D**, Cartoons depicting the truncation/deletion mutants of HPV-31 L2. Scale is provided in amino acid numbers of the full-length protein at the top. **F**, Expression level of upregulated EMT genes in response to L2 overexpression (see also Supplementary Table S6). The z-score normalized expression of each gene is indicated by color intensity according to the color scale. **G**, Transwell invasion assay of Het-1A human esophagus cells overexpressing HPV-31 L2 or GFP. Representative images of the bottom side of the Transwell after invasion of Het-1A cells overexpressing HPV-31 L2 (top) and GFP (bottom). **H** and **I**, Quantification of Transwell invasion assays. Individual dots with centerline (mean) and error bar (standard deviation) show the number of invaded cells per image (*P* value from two-tailed *t* test) of Het-1A cells overexpressing HPV-31 L2 and GFP (**H**) and control UD-SCC2 hypopharyngeal cancer cells compared with polyclonal disruptions of RNF20 (**I**). **J**, Western blot analysis of UD-SCC2 hypopharyngeal cancer cells subjected to CRISPR/Cas9 gene editing using the indicated guide RNAs. Proteins indicated on the right of each blot were detected using the antibodies indicated. Bands were cropped from the same original membrane.

the newly identified E5-p16 interaction likely phenocopies *CDKN2A* loss of function in this same pathway.

Our integrated approach also led to the discovery of previously unknown viral-human interactions involving E1-KEAP1 and L2-RNF20/40. The KEAP1-NRF2 pathway has been implicated in oncogenesis and chemoresistance of several tumor types, including lung adenocarcinoma, lung squamous cell carcinoma, papillary renal cell carcinoma, and esophageal cancers (62, 63). Individual genes of this pathway have also been shown to be mutated in CESC as well as HNSCC (13, 32). Epigenetic inactivation of this pathway has been reported for HNSCC and linked to decreased survival for affected patients (64). Although analysis of HNSCC tumor genomes has previously indicated an increased mutation rate in the HPV⁺ population of patients (13), we extended this analysis by identifying that the HPV E1 protein directly interacts with KEAP1, thus phenocopying the frequent inactivating mutations in the *CUL3-KEAP1-NFE2L2* pathway.

RNF20, which was found to interact with HPV L2, forms a functional histone ubiquitin ligase complex with RNF40 and monoubiquitinates histone H2B K120. Notably, RNF20/40 loss of function has previously been linked to tumor progression and invasiveness in breast cancer cells (49, 50). We observed large-scale expression changes in response to the overexpression of L2. EMT and cell invasion are the most prominently upregulated functions, reflecting a similar oncogenic phenotype as observed in previous RNF20 loss-of-function experiments. Interestingly, the upregulated genes in L2 overexpression are also significantly enriched for a previously reported gene cluster associated with cell motility in HNSCC, characterizing poorly differentiated tumors (65). These findings suggest that L2 binding can phenocopy the highly invasive phenotype promoted by inactivating mutations in RNF20/40, highlighted by our finding of increased invasiveness in an *in vitro* invasion assay upon overexpression of HPV L2. We further showed a decrease in the invasive phenotype of HPV⁺ HNSCC cells when RNF20 was ablated using CRISPR, suggesting a link between the virus-host PPI and an invasive phenotype. These data are consistent with our finding that CRISPR-mediated removal of L2 in an HPV⁺ cancer cell line reduced the invasive phenotype.

Altogether, these results show that viral phenocopying of recurrent tumor mutations is not a rare evolutionary event restricted to a few classic interactions, but a common scenario. Many HPV proteins, including the previously unappreciated E1 and L2, can play key roles in carcinogenesis. Finally, methodologically, the combination of PPI mapping and tumor genome analysis defines a pipeline that will enable the study of many other virally induced cancers, including those linked to hepatitis, Epstein-Barr virus, and adenoviruses.

METHODS

Plasmids and Cell Lines

Codon-optimized versions of HPV-31 gene sequences were provided by Jacques Archambault (McGill University, Canada). Codon-optimized versions of HPV-16 genes were gifts from Alison McBride (E1, NIAID, NIH), Richard Schlegel (E5, Georgetown University), and John Schiller (p16sheLL, containing L1 and L2 sequences, Addgene plasmid # 37320; ref. 66). PCR-amplified inserts modified using SLIC (67)

to contain a C-terminal 2× Strep-tag were cloned into pcDNA4/TO (Invitrogen) or the lentiviral vector system pLVX-TetOne-Puro (Clontech) and verified by sequencing. pMD2.G and pDR8.91 were gifts from Didier Trono (Addgene plasmid # 12259), FLAG-KEAP1 was a gift from Qing Zhong (Addgene plasmid # 28023; ref. 68), and NC16 pCDNA3.1 FLAG NRF2 was a gift from Randall Moon (Addgene plasmid # 36971; ref. 69). Plasmids encoding truncation mutants of HPV-31 E1 and L2 were created using PCR-based amplification of the truncated sequences using primers introducing restriction sites to allow subcloning into the original pcDNA4/TO plasmid containing a C-terminal 2× Strep-tag. Deletion mutants were created using PCR-based amplification of the complete backbone of the full-length construct using outward-facing primers flanking the desired deletion. All constructs were verified by sequencing. Primer and plasmid sequences will be shared upon request.

The human cervical carcinoma cell line C33A (21), obtained in 2004, as well as the human embryonic kidney cell lines 293 (HEK293; ref. 30) and 293T (70), obtained before 2010, were maintained in high-glucose DMEM supplemented with 10% FBS (Invitrogen), 1 mmol/L sodium pyruvate, and 100 U/mL penicillin/streptomycin (Invitrogen). The human esophagus cell line Het-1A (31), obtained in 2015, was maintained in BEGM (Lonza), consisting of Broncho Epithelial Basal medium with the additives of the Bullet kit except GA-1000 (gentamycin-amphotericin B mix) and supplemented with 100 U/mL penicillin/streptomycin. The HPV⁺ cell line UD-SCC-2 (SCC2; hypopharynx) was a gift from Silvio Gutkind, obtained in 2015. These cells were maintained in high-glucose DMEM supplemented with 10% FBS (Invitrogen), 1× nonessential amino acids (UCSF Cell Culture Facility), and 100 U/mL penicillin/streptomycin (Invitrogen). All cell lines were last validated in 2017 by STR analysis using the GenePrint 10 assay (Promega) at the University of California, Berkeley Cell Culture Facility.

AP, WB, and Silver Stain

APs were performed as previously described (15). Briefly, C33A, HEK293, and 293T cells were transfected using PolyJet transfection reagent (SignaGen Laboratories), and cell pellets were harvested 40 hours after transfection. Het-1A cells were transduced with Tet-One inducible lentiviral particles. These were produced from 293T cells transfected with plasmids pDR8.91, pCMV-VSVg (pMD2.G), and pLVX-TetOne-Puro versions encoding the respective viral proteins using PolyJet, supernatants harvested 2 days after transfection, cleared by passing through 0.22-μm filters, concentrated with 8.5% of PEG-6000 and 0.3 mol/L of NaCl and resuspended in PBS. Polyclonal population of stable Het-1A cells were selected using 2 μg/mL puromycin (Calbiochem) 24 hours after transduction and maintained in 2 μg/mL puromycin. Expression of HPV-31 genes was induced by treatment with 1,000 μg/mL doxycycline for 16 hours before harvest. Clarified cell lysates were incubated with prewashed Strep-Tactin beads (IBA Life Sciences) and allowed to bind for 2 hours. Following purification, complexes bound to beads were washed and then eluted with desthiobiotin (IBA Life Sciences).

Proteins from cell lysates and AP eluates were separated by SDS-PAGE and either directly stained using the Pierce Silver Stain Kit (Thermo Fisher Scientific) or transferred to a PVDF membrane. Membranes were probed with the indicated primary antisera and bound antibodies were detected using Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific). Proteins were detected using the following antibodies: Strep-tag (#34850, Qiagen), KEAP1 (#10503-2-AP, Proteintech), and RNF20 (#A300-714, Bethyl Laboratories and clone D6E10, #11974, Cell Signaling Technology).

Mass Spectrometry

AP eluates were trypsin digested, desalted, and concentrated as previously described (14). Digested peptide mixtures were analyzed by LC/MS-MS on a Thermo Scientific Velos Pro Linear Ion Trap MS

system or a Thermo Scientific Q Exactive Hybrid Quadrupole Orbitrap MS system equipped with a Proxeon EASY nLC II high-pressure liquid chromatography and autosampler system.

Network Scoring and Benchmarking

MS raw data were searched against a database containing SwissProt human protein sequences and the individual viral bait sequences using the Protein Prospector algorithm. Interactions were scored with the MiST (15, 22) algorithm using the standard weights (0.309 for reproducibility, 0.685 for specificity, and 0.006 for abundance). Previously reported interactions supported by evidence from at least two publications were included in a list of “gold standards” (Supplementary Table S1). PPIs that were not detected in our data set at all (4) were excluded from this analysis, but are listed in the Supplementary Table. A ROC curve was plotted using the R package “pROC” and shows the performance of the MiST score for the 27 gold standards. At a cutoff of MiST > 0.75, we found 137 high-confidence interactions, including 18 gold standards, of a total of 4,055 PPIs detected (Supplementary Table S2). Cytoscape (71) was used for visualization of the PPI network (Fig. 2), as well as all other network representations of data (Figs. 3D, 4A, 5A). An interactive version of Fig. 2 is available on the public NDEx server (ref. 72; <https://doi.org/10.18119/N9JS3T>).

Functional Enrichment

Gene ontology (GO) term enrichment was performed using the online DAVID Functional Annotation Tool (<https://david.ncifcrf.gov>; ref. 73) using only high-confidence interacting human proteins as input. Interactors of individual viral proteins were tested using the “functional annotation clustering” tool using standard settings for the annotation category “GOTERM_CC_ALL” (Cellular Components) or the default settings (Processes and Pathways). Enriched GO terms passing a cutoff of Benjamini-Hochberg corrected FDR < 5% were manually curated as shown in Supplementary Table S3.

Data Processing for Tumor Samples

RNA-seq, CNA (SNP 6.0), and somatic mutations of 505 HNSCC samples and 177 cervical cancer samples from TCGA (13, 32) were downloaded from the Broad Institute’s Firehose (<http://gdac.broadinstitute.org/>). Somatic mutations were identified by comparing whole-exome sequences of paired tumor and normal samples using the Firehose pipeline (<http://gdac.broadinstitute.org/>). RNA-seq data were used to exclude genes with low mRNA expression levels from further analysis. First, each gene’s RNA-seq by Expectation Maximization (RSEM; ref. 74) value was normalized by dividing it by the 75th percentile of RSEM values within the tumor sample and then multiplying by 1,000, as per TCGA practice (<https://wiki.nci.nih.gov/display/TCGA/RNAseq+Version+2>). Genes were retained if the normalized RSEM was greater than 0.125 in over 50% of samples in either cancer type, resulting in 17,864 expressed genes. CNA were identified based on the output from GISTIC2, indicating copy-number status of the genome region containing each gene, normalized to the average for the chromosome arm (+1 = increased relative to arm; 0 = same as arm average; -1 = decreased relative to arm).

Somatic mutations and CNA data of 118 HNSCC samples from a University of Chicago study (33) were also analyzed. In this cohort, 617 cancer-associated genes were sequenced in matched tumor-normal pairs. Somatic mutations were identified using MuTect as described previously (33). CNAs were identified from the sequencing data using the CONTRA algorithm (33).

HPV Status of TCGA and Chicago HNSCC Samples

For the HNSCC samples from TCGA, samples were classified as HPV+ if they had more than 1,000 RNA-seq reads primarily aligned to the HPV genes E6 and E7 (13). For the samples without

sequencing-based HPV status, HPV status was indicated from a diagnostic run on the tissues after they had reached TCGA, downloaded on February 12, 2016, from <https://tcga-data.nci.nih.gov/>, and is now available through http://ideker.ucsd.edu/papers/eckhardt2018/nationwidechildrens.org_auxiliary_hnsc.txt. For the cervical cancer samples from TCGA, HPV status was determined using consensus results from MassArray and RNA-seq (32). For the University of Chicago HNSCC samples, their HPV status was previously determined by E6/E7-specific qRT-PCR as described (33).

Identification of Mutated Genes in Each Tumor

Genes were classified as wild-type or altered in each of the 800 tumors with alterations defined as follows. Most oncogenes (e.g., *EGFR*) were considered altered (activated) if affected by a missense mutation, in-frame indel, or copy-number amplification. For the subset of oncogenes typically altered only by amplification (ref. 75; *CCND1*, *LMO1*, *MDM2*, *MDM4*, *MYC*, *MYCL*, *MYCN*, *NCOA3*, *NKX2-1*, and *SKP2*), only copy-number amplifications were considered as alterations and not SNVs or indels. All other genes including tumor suppressors (e.g., *CDKN2A*) were considered altered (inactivated) if there was any type of nonsilent mutation or a copy-number deletion.

Difference in Mutation Rate by HPV Status

For each gene, we fit a logistic regression model of its alteration state g (0 = wild-type; 1 = altered) as a function of v , the HPV status (1 = HPV+; 0 = HPV-), controlling for the impact of the cancer types t (HNSCC vs. cervical cancer) and cohorts c (TCGA vs. University of Chicago) as covariates:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 v + \beta_2 t + \beta_3 c)}} \quad (1)$$

where p is the probability that $g = 1$. The parameters β were estimated from data from 800 tumors with matched exome sequencing, CNA and HPV status data. To assess whether the HPV status is significantly associated with a gene’s alterations, the likelihood of the complete model (Eq. 1) was compared with that of a simple model under null hypothesis of no association (i.e., $\beta_1 = 0$). The deviance D (i.e., log likelihood ratio) between the two nested models was calculated as follows:

$$D = -2 \ln \left(\frac{L_{\text{null}}}{L_{\text{complete}}} \right) \quad (2)$$

where

$$L = \prod_{i=1}^m (p^i)^{g^i} (1 - p^i)^{1 - g^i}$$

and m represents the number of samples. For the genes with an increased mutation rate in HPV- tumors (i.e., $\beta_1 < 0$), D was used as the test statistic for the following network propagation.

Network Propagation

Network propagation (39) was used to identify the clusters of genes within the ReactomeFI (38) that are enriched in both HPV interactors and genes with an increased mutation rate in HPV- tumors. All human proteins that were included in the ReactomeFI reference network and had mRNA expressed in either HNSCC or cervical cancer, as described in the Data Processing for Tumor Samples section, were used for network propagation ($n = 10,225$ proteins). MiST scores (without threshold), which characterize the confidence of HPV-human PPIs, or the deviances D , which characterize the significance of the increased mutation rates in HPV- tumors (Eq. 2), were separately propagated across the ReactomeFI network based on a random walk model (equivalent to heat diffusion) with a restart probability

of 0.5. After convergence, the two scores of each gene represent its network proximity to HPV protein binding or increased mutation rates in HPV⁺. The product of two propagated scores for each gene, representing its network proximity to both the above events, was used as the test statistic. To estimate the expected background of this product controlling for the network topology, 20,000 permutations were performed by randomly reassigning MiST scores or deviances D , performing the network propagation and calculating the product of the two propagated scores. Intuitively, the empirical probability derived from this analysis controls for the impact of network structure on the flow of the propagation (i.e., a protein with many neighbors is likely to acquire a high propagated score regardless of the starting data). The goal was to produce a distribution of this product for each gene under the null hypothesis that the network neighborhood of this gene is not significantly enriched in both HPV binding and increased genetic alterations in HPV⁺, while retaining the network structure. This null distribution was indexed with the observed product to obtain an empirical P value characterizing the significance in both propagations. Using the Storey approach (76), q -values were then calculated and a threshold of FDR $\leq 25\%$ was applied.

Luciferase Reporter Assay

The Cignal Antioxidant Response reporter assay kit (Qiagen) was used to measure NRF2 activity in C33A cells according to the manufacturer's instructions. Briefly, cells were seeded in 96-well plates (Costar) and transfected using PolyJet (SigmaGen Laboratories) with the following plasmids: Cignal ARE Luciferase Reporter, Negative Control or Positive Control mix (25 ng), pcDNA4/TO-based pcHPV31-E1-S, pcGFP-SF or an empty vector as control (50 ng), pCDNA3.1 FLAG NRF2 (12.5 ng; Addgene #36971), and FLAG-KEAP1 (37.5 ng; Addgene #28023). The medium was replaced 6 hours after transfection and cells were harvested 36 to 40 hours after transfection. Luciferase activity (Firefly and Renilla) was measured using the Dual-Glo-Luciferase kit (Promega) according to the manufacturer's instructions, and NRF2 (Firefly) activity was calculated by normalization to the Renilla luciferase readings.

RNA-seq Sample Preparation and Data Analysis

Het-1A cells were induced for L2 or GFP overexpression using 1,000 $\mu\text{g/mL}$ doxycycline for 72 hours before harvest. Cell pellets were flash-frozen and total RNA extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA integrity was checked using the Eukaryote Total RNA Nano BioAnalyzer assay (Agilent), followed by RNA-seq library preparation using the RNA-seq Strand-Specific Library Preparation Kit (NuGen). Sequencing library quality was checked using a BioAnalyzer High-Sensitivity DNA assay (Agilent) and quantified by quantitative PCR using the KAPA qPCR Library Quantification Kit (KAPA Biosystems). The experiment was performed in triplicates, isolating RNA in two batches and constructing the sequencing libraries in one batch. DNA was pooled at 10 nmol/L and run on a single lane of the Illumina HiSeq400 sequencer in single-end mode.

All sequencing reads were aligned to the human reference genome and transcriptome (GRCh37) using TopHat version 2.0.14 (77) with default parameters. Cufflinks version 2.2.1 (78) was used to quantify the mRNA abundance for each gene [Fragments Per Kilobase of exon model per Million mapped fragments (FPKM)]. The following nondefault options were used with Cufflinks: frag-bias-correct and multi-read-correct. RankProd (79) was applied to perform differential expression analysis between L2 overexpression and GFP overexpression samples.

We also performed gene set enrichment analysis (<http://www.broad.mit.edu/gsea/>; refs. 80, 81) on all the genes detected by RNA-seq, which were ranked by their $\log_{10}(P)$ values given by the RankProd package (79). A cutoff of FDR $\leq 10\%$ was used to define the enriched upregulated and downregulated functions.

In Vitro Invasion Assay

In vitro invasion assays were carried out in growth factor-reduced Matrigel-coated Transwell chambers (Corning Life Sciences). Briefly, cells were plated in the Transwell chamber in serum-reduced media (BEGM media minus BPE, hEGF, and GA, or full DMEM growth media with 2.5% BSA). Cells were incubated at 37°C with 5% CO₂ for 72 hours. Cells that invaded the Transwell chamber were fixed and stained using the HEMA3 staining solutions (Fisher Scientific) according to the manufacturer's instructions, and 0.5% crystal violet. The number of cells that invaded the Matrigel was determined by microscopy. A fraction of cells for each condition was seeded in triplicate wells of a 96-well plate in parallel and cultured for 72 hours. Cell viability was determined using the CellTiter-Glo2.0 assay (Promega), and cell counts were normalized accordingly.

Creation of CRISPR-Mediated Polyclonal Gene Disruptions

We used the Zhang lab's GeCKO Lentiviral CRISPR Toolbox, specifically LentiCRISPR v2 (82, 83), to introduce a puromycin-selectable hSpCas9 and the chimeric guide RNA with a single lentiviral vector. For each guide RNA, a pair of annealed oligos designed based on a 20-bp target site sequence was cloned into the single-guide RNA scaffold following the established protocol (https://media.addgene.org/data/plasmids/52/52961/52961-attachment_B3xTwla0bkYD.pdf). Lentiviral particles were produced from 293T cells transfected with plasmids pDR8.91, pCMV-VSVg (pMD2.G), and the LentiCRISPR v2 plasmid versions containing the respective guide RNA sequence using PolyJet, and supernatants were harvested 2 days after transfection, cleared by passing through 0.22- μm filters, concentrated with 8.5% of PEG-6000 and 0.3 mol/L of NaCl, and resuspended in PBS. Polyclonal population of stable cells were selected using 2.5 $\mu\text{g/mL}$ puromycin (Calbiochem) 48 hours after transduction. Four days later, half the cells were harvested for Tracking of Indels by Decomposition (TIDE) analysis, and the remainder maintained at a concentration of 2 $\mu\text{g/mL}$ puromycin. Guide RNA sequences used are listed in the Supplementary Methods.

PCR Amplification of Target Regions and TIDE Analysis

Primer design and sequences are described in detail in the Supplementary Methods. For extraction of genomic DNA, cells were lysed in QuickExtract solution (Epicentre) according to the manufacturer's protocol. PCR was performed using the high-fidelity polymerase Phusion (Thermo Fisher); thermocycler settings are described in the Supplementary Methods. PCR cleanup was either performed following gel purification of the fragments using NucleoSpin Gel and PCR Clean-Up (Macherey Nagel) or ordered through Quintarabio, followed by sequencing using one of the PCR primers. Sequencing traces were then analyzed using the TIDE webtool (<https://tide.nki.nl/>; ref. 84).

Disclosure of Potential Conflicts of Interest

A.M. Gross is a scientist at Illumina. T. Ideker has ownership interest (including stock, patents, etc.) in Data4Cure and Ideaya and is a consultant/advisory board member for Data4Cure and Ideaya. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: M. Eckhardt, W. Zhang, T. Ideker, N.J. Krogan
Development of methodology: M. Eckhardt, W. Zhang, K.E. Franks-Skiba, T. Ideker, N.J. Krogan
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M. Eckhardt, J.R. Johnson, K.E. Franks-Skiba, D.L. Swaney, T.L. Johnson, G.M. Jang, P.S. Shah, T.M. Brand, J. Archambault
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M. Eckhardt, W. Zhang,

A.M. Gross, J. Von Dollen, D.L. Swaney, P.S. Shah, J.F. Kreisberg, J.R. Grandis, T. Ideker, N.J. Krogan

Writing, review, and/or revision of the manuscript: M. Eckhardt, W. Zhang, K.E. Franks-Skiba, D.L. Swaney, J. Archambault, J.F. Kreisberg, J.R. Grandis, T. Ideker, N.J. Krogan

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): K.E. Franks-Skiba, J. Archambault, T. Ideker, N.J. Krogan

Study supervision: T. Ideker, N.J. Krogan

Other (experimental work): G.M. Jang

Acknowledgments

We would like to thank Alex Choi, Stefanie Jäger, Erica Stevenson, Laura Satkamp, Billy Newton, Margaret Soucheray, Adriana Pitea, Natali Gulbahce, Cyrus Maher (Ryan Hernandez lab), Eric Verschueren, Neil Bhola, Michael McGregor, Judd Hultquist, John Gordan, Mike Shales, and David Jimenez-Morales for technical help and many fruitful discussions during various stages of this project. We thank Eric Chow (UCSF CAT) for sequencing; Natasha Carli and Jim McGuire (Gladstone Genomics Core) for RNA-seq library preparation; Alison McBride and Richard Schlegel for providing plasmids; and Silvio Gutkind for providing cell lines and expert advice. This research was funded by grants from NIH (U54 CA209891 to J.R. Grandis, T. Ideker, and N.J. Krogan; P50 GM082250 to N.J. Krogan; and P50 GM085764 to T. Ideker) and TRDRP (25FT-0010 to M. Eckhardt). The results published here are in whole or part based upon data generated by the TCGA Research Network (<http://cancergenome.nih.gov/>).

Received September 11, 2017; revised March 22, 2018; accepted August 17, 2018; published first September 12, 2018.

REFERENCES

- Viens LJ, Henley SJ, Watson M, Markowitz LE, Thomas CC, Thompson TD, et al. Human papillomavirus-associated cancers - United States, 2008-2012. *MMWR Morb Mortal Wkly Rep* 2016;65:661-6.
- Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359-86.
- Benard VB, Castle PE, Jenison SA, Hunt WC, Kim JJ, Cuzick J, et al. Population-based incidence rates of cervical intraepithelial neoplasia in the human papillomavirus vaccine era. *JAMA Oncol* 2016; Available from: <http://dx.doi.org/10.1001/jamaoncol.2016.3609>
- Venuti A, Paolini F, Nasir L, Corteggio A, Roperto S, Campo MS, et al. Papillomavirus E5: the smallest oncoprotein with many functions. *Mol Cancer* 2011;10:140.
- Huibregtse JM, Scheffner M, Howley PM. A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *EMBO J* 1991;10:4129-35.
- Dyson N, Howley PM, Münger K, Harlow E. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science* 1989;243:934-7.
- O'Shea CC. Viruses - seeking and destroying the tumor program. *Oncogene* 2005;24:7640-55.
- Ou HD, May AP, O'Shea CC. The critical protein interactions and structures that elicit growth deregulation in cancer and viral replication. *Wiley Interdiscip Rev Syst Biol Med* 2011;3:48-73.
- Rozenblatt-Rosen O, Deo RC, Padi M, Adelman G, Calderwood MA, Rolland T, et al. Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature* 2012;487:491-5.
- White EA, Sowa ME, Tan MJA, Jeudy S, Hayes SD, Santha S, et al. Systematic identification of interactions between host cell proteins and E7 oncoproteins from diverse human papillomaviruses. *Proc Natl Acad Sci U S A* 2012;109:E260-7.
- White EA, Kramer RE, Tan MJA, Hayes SD, Harper JW, Howley PM. Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity. *J Virol* 2012;86:13174-86.
- Jang MK, Anderson DE, van Doorslaer K, McBride AA. A proteomic approach to discover and compare interacting partners of papillomavirus E2 proteins from diverse phylogenetic groups. *Proteomics* 2015;15:2038-50.
- Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 2015;517:576-82.
- Jäger S, Gulbahce N, Cimermancic P, Kane J, He N, Chou S, et al. Purification and characterization of HIV-human protein complexes. *Methods* 2011;53:13-9.
- Jäger S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, et al. Global landscape of HIV-human protein complexes. *Nature* 2011;481:365-70.
- Jäger S, Kim DY, Hultquist JF, Shindo K, LaRue RS, Kwon E, et al. Vif hijacks CBF- β to degrade APOBEC3G and promote HIV-1 infection. *Nature* 2011;481:371-5.
- Ramage HR, Kumar GR, Verschueren E, Johnson JR, Von Dollen J, Johnson T, et al. A combined proteomics/genomics approach links hepatitis C virus infection with nonsense-mediated mRNA decay. *Mol Cell* 2015;57:329-40.
- Davis ZH, Verschueren E, Jang GM, Kleffman K, Johnson JR, Park J, et al. Global mapping of herpesvirus-host protein complexes reveals a transcription strategy for late genes. *Mol Cell* 2015;57:349-60.
- Mirrashedi KM, Elwell CA, Verschueren E, Johnson JR, Frando A, Von Dollen J, et al. Global mapping of the INC-human interactome reveals that retromer restricts chlamydia infection. *Cell Host Microbe* 2015;18:109-21.
- Shah PS, Wojcechowskyj JA, Eckhardt M, Krogan NJ. Comparative mapping of host-pathogen protein-protein interactions. *Curr Opin Microbiol* 2015;27:62-8.
- Auersperg N. Long-term cultivation of hypodiploid human tumor cells. *J Natl Cancer Inst* 1964;32:135-63.
- Verschueren E, Von Dollen J, Cimermancic P, Gulbahce N, Sali A, Krogan NJ. Scoring large-scale affinity purification mass spectrometry datasets with MiST. *Curr Protoc Bioinformatics* 2015;49:8.19.1-16.
- Scheffner M, Munger K, Byrne JC, Howley PM. The state of the p53 and retinoblastoma genes in human cervical carcinoma cell lines. *Proc Natl Acad Sci* 1991;88:5523-7.
- Antonucci LA, Egger JV, Krucher NA. Phosphorylation of the retinoblastoma protein (Rb) on serine-807 is required for association with Bax. *Cell Cycle* 2014;13:3611-7.
- Conrad M, Bubbs VJ, Schlegel R. The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein. *J Virol* 1993;67:6170-8.
- Lehoux M, Gagnon D, Archambault J. E1-mediated recruitment of a UAF1-USP deubiquitinase complex facilitates human papillomavirus DNA replication. *J Virol* 2014;88:8545-55.
- Cohn MA, Kee Y, Haas W, Gygi SP, D'Andrea AD. UAF1 is a subunit of multiple deubiquitinating enzyme complexes. *J Biol Chem* 2009;284:5343-51.
- Bellacchio E, Paggi MG. Understanding the targeting of the RB family proteins by viral oncoproteins to defeat their oncogenic machinery. *J Cell Physiol* 2013;228:285-91.
- Mortensen F, Schneider D, Barbic T, Sladewska-Marquardt A, Kühnle S, Marx A, et al. Role of ubiquitin and the HPV E6 oncoprotein in E6AP-mediated ubiquitination. *Proc Natl Acad Sci U S A* 2015;112:9872-7.
- Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol* 1977;36:59-74.
- Stoner GD, Kaighn ME, Reddel RR, Resau JH, Bowman D, Naito Z, et al. Establishment and characterization of SV40 T-antigen immortalized human esophageal epithelial cells. *Cancer Res* 1991;51:365-71.
- Cancer Genome Atlas Research Network, Albert Einstein College of Medicine, Analytical Biological Services, Barretos Cancer Hospital,

- Baylor College of Medicine, Beckman Research Institute of City of Hope, et al. Integrated genomic and molecular characterization of cervical cancer. *Nature* 2017;543:378–84.
33. Seiwerth TY, Zuo Z, Keck MK, Khattri A, Pedamallu CS, Stricker T, et al. Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clin Cancer Res* 2015;21:632–41.
 34. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014;505:495–501.
 35. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res* 2012;22:398–406.
 36. Leiserson MDM, Wu H-T, Vandin F, Raphael BJ. CoMet: a statistical approach to identify combinations of mutually exclusive alterations in cancer. *Genome Biol* 2015;16:160.
 37. Sherr CJ, Roberts JM. CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev* 1999;13:1501–12.
 38. Wu G, Feng X, Stein L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol* 2010;11:R53.
 39. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. *Nat Methods* 2013;10:1108–15.
 40. Itoh K, Mimura J, Yamamoto M. Discovery of the negative regulator of Nrf2, Keap1: a historical overview. *Antioxid Redox Signal* 2010;13:1665–78.
 41. Kobayashi A, Kang M-I, Okawa H, Ohtsui M, Zenke Y, Chiba T, et al. Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2. *Mol Cell Biol* 2004;24:7130–9.
 42. Zhang DD, Lo S-C, Cross JV, Templeton DJ, Hannink M. Keap1 is a redox-regulated substrate adaptor protein for a Cul3-dependent ubiquitin ligase complex. *Mol Cell Biol* 2004;24:10941–53.
 43. Jaramillo MC, Zhang DD. The emerging role of the Nrf2–Keap1 signaling pathway in cancer. *Genes Dev* 2013;27:2179–91.
 44. Taguchi K, Motohashi H, Yamamoto M. Molecular mechanisms of the Keap1–Nrf2 pathway in stress response and cancer evolution. *Genes Cells* 2011;16:123–40.
 45. Kansanen E, Kuosmanen SM, Leinonen H, Leinonen A-L. The Keap1–Nrf2 pathway: mechanisms of activation and dysregulation in cancer. *Redox Biol* 2013;1:45–9.
 46. Wood A, Krogan NJ, Dover J, Schneider J, Heidt J, Boateng MA, et al. Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. *Mol Cell* 2003;11:267–74.
 47. Hwang WW, Venkatasubrahmanyam S, Ianculescu AG, Tong A, Boone C, Madhani HD. A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. *Mol Cell* 2003;11:261–6.
 48. Lee J-S, Shukla A, Schneider J, Swanson SK, Washburn MP, Florens L, et al. Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell* 2007;131:1084–96.
 49. Shema E, Tirosh I, Aylon Y, Huang J, Ye C, Moskovits N, et al. The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes Dev* 2008;22:2664–76.
 50. Prenzel T, Begus-Nahrman Y, Kramer F, Hennion M, Hsu C, Gorsler T, et al. Estrogen-dependent gene transcription in human breast cancer cells relies upon proteasome-dependent monoubiquitination of histone H2B. *Cancer Res* 2011;71:5739–53.
 51. Bedi U. Regulation of H2B monoubiquitination pathway in breast cancer. 2015; Available from: <https://ediss.uni-goettingen.de/handle/11858/00-1735-0000-0022-5D8E-C?locale-attribute=de>
 52. Aydin I, Villalonga-Planells R, Greune L, Bronnimann MP, Calton CM, Becker M, et al. A central region in the minor capsid protein of papillomaviruses facilitates viral genome tethering and membrane penetration for mitotic nuclear entry. *PLoS Pathog* 2017;13:e1006308.
 53. Ying Y, Tao Q. Epigenetic disruption of the WNT/beta-catenin signaling pathway in human cancers. *Epigenetics* 2009;4:307–12.
 54. Hu T, Li C. Convergence between Wnt- β -catenin and EGFR signaling in cancer. *Mol Cancer* 2010;9:236.
 55. Basu S, Haase G, Ben-Ze'ev A. Wnt signaling in cancer stem cells and colon cancer metastasis. *F1000Res* 2016;5. Available from: <http://dx.doi.org/10.12688/f1000research.7579.1>
 56. Bello JOM, Nieva LO, Paredes AC, Gonzalez AMF, Zavaleta LR, Lizano M. Regulation of the Wnt/ β -catenin signaling pathway by human papillomavirus E6 and E7 oncoproteins. *Viruses* 2015;7:4734–55.
 57. Ramírez-Salazar E, Centeno F, Nieto K, Valencia-Hernández A, Salcedo M, Garrido E. HPV16 E2 could act as down-regulator in cellular genes implicated in apoptosis, proliferation and cell differentiation. *Virology* 2011;8:247.
 58. Collins SI, Constandinou-Williams C, Wen K, Young LS, Roberts S, Murray PG, et al. Disruption of the E2 gene is a common and early event in the natural history of cervical human papillomavirus infection: a longitudinal cohort study. *Cancer Res* 2009;69:3828–32.
 59. Parfenov M, Pedamallu CS, Gehlenborg N, Freeman SS, Danilova L, Bristow CA, et al. Characterization of HPV and host genome interactions in primary head and neck cancers. *Proc Natl Acad Sci U S A* 2014;111:15544–9.
 60. Speel EJM. HPV integration in head and neck squamous cell carcinomas: cause and consequence. *Recent Results Cancer Res* 2017;206:57–72.
 61. Stott FJ. The alternative product from the human CDKN2A locus, p14ARF, participates in a regulatory feedback loop with p53 and MDM2. *EMBO J* 1998;17:5001–14.
 62. Leinonen HM, Kansanen E, Pölonen P, Heinäniemi M, Leinonen A-L. Dysregulation of the Keap1–Nrf2 pathway in cancer. *Biochem Soc Trans* 2015;43:645–9.
 63. Cancer Genome Atlas Research Network, Analysis Working Group: Asan University, BC Cancer Agency, Brigham and Women's Hospital, Broad Institute, Brown University, et al. Integrated genomic characterization of oesophageal carcinoma. *Nature* 2017;541:169–75.
 64. Martinez VD, Vucic EA, Thu KL, Pikor LA, Lam S, Lam WL. Disruption of KEAP1/CUL3/RBX1 E3-ubiquitin ligase complex components by multiple genetic mechanisms: association with poor prognosis in head and neck cancer. *Head Neck* 2015;37:727–34.
 65. Rickman DS, Millon R, De Reynies A, Thomas E, Wasyluk C, Muller D, et al. Prediction of future metastasis and molecular characterization of head and neck squamous-cell carcinoma based on transcriptome and genome analysis by microarrays. *Oncogene* 2008;27:6607–22.
 66. Buck CB, Thompson CD, Roberts JN, Müller M, Lowy DR, Schiller JT. Carrageenan is a potent inhibitor of papillomavirus infection. *PLoS Pathog* 2006;2:e69.
 67. Li MZ, Eledge SJ. SLIC: a method for sequence- and ligation-independent cloning. In: Peccoud J, editor. *Gene Synthesis*. Humana Press; p. 51–9.
 68. Fan W, Tang Z, Chen D, Moughon D, Ding X, Chen S, et al. Keap1 facilitates p62-mediated ubiquitin aggregate clearance via autophagy. *Autophagy* 2010;6:614–21.
 69. Camp ND, James RG, Dawson DW, Yan F, Davison JM, Houck SA, et al. Wilms tumor gene on X chromosome (WTX) inhibits degradation of NRF2 protein through competitive binding to KEAP1 protein. *J Biol Chem* 2012;287:6539–50.
 70. Pear WS, Nolan GP, Scott ML, Baltimore D. Production of high-titer helper-free retroviruses by transient transfection. *Proc Natl Acad Sci U S A* 1993;90:8392–6.
 71. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.
 72. Pratt D, Chen J, Welker D, Rivas R, Pillich R, Rynkov V, et al. NDEx, the Network Data Exchange. *Cell Syst* 2015;1:302–5.
 73. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane H, et al. 10.1186/gb-2003-4-9-r60. *Genome Biol*. 2003. p. R60. Available from: <http://genomobiology.biomedcentral.com/articles/10.1186/gb-2003-4-9-r60>
 74. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* 2011;12:323.
 75. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science* 2013;339:1546–58.
 76. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;100:9440–5.

77. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;14:R36.
78. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;7:562–78.
79. Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 2006;22:2825–7.
80. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
81. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003;34:267–73.
82. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. 2014. Available from: <http://dx.doi.org/10.1101/006726>
83. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen T, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 2014;343:84–7.
84. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 2014;42:e168.

CANCER DISCOVERY

Multiple Routes to Oncogenesis Are Promoted by the Human Papillomavirus–Host Protein Network

Manon Eckhardt, Wei Zhang, Andrew M. Gross, et al.

Cancer Discov 2018;8:1474-1489. Published OnlineFirst September 12, 2018.

Updated version Access the most recent version of this article at:
doi:[10.1158/2159-8290.CD-17-1018](https://doi.org/10.1158/2159-8290.CD-17-1018)

Supplementary Material Access the most recent supplemental material at:
<http://cancerdiscovery.aacrjournals.org/content/suppl/2018/09/07/2159-8290.CD-17-1018.DC1>

Cited articles This article cites 78 articles, 28 of which you can access for free at:
<http://cancerdiscovery.aacrjournals.org/content/8/11/1474.full#ref-list-1>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

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

Permissions To request permission to re-use all or part of this article, use this link
<http://cancerdiscovery.aacrjournals.org/content/8/11/1474>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.