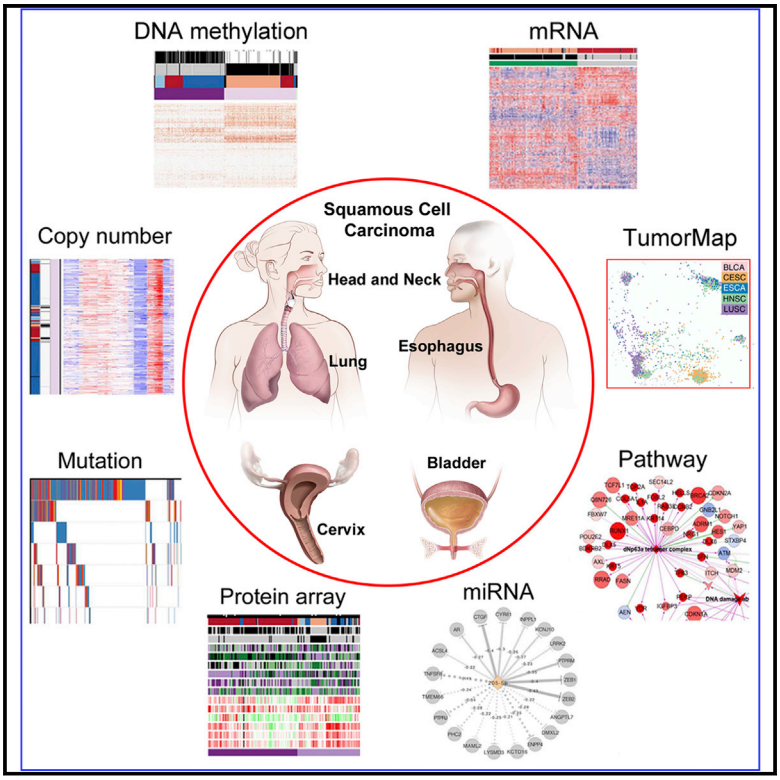
Resource

Cell Reports

Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas

Graphical Abstract



Highlights

- SCCs show chromosome or methylation alterations affecting multiple related genes
- These regulate squamous stemness, differentiation, growth, survival, and inflammation
- Copy-quiet SCCs have hypermethylated (*FANCF*, *TET1*) or mutated (*CASP8*, *MAPK-RAS*) genes
- Potential targets include ΔNp63, WEE1, IAPs, PI3K-mTOR/MAPK, and immune responses

Authors

Joshua D. Campbell, Christina Yau, Reanne Bowlby, ..., Curtis R. Pickering, Zhong Chen, Carter Van Waes

Correspondence

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

In Brief

Campbell et al. reveal that squamous cell cancers from different tissue sites may be distinguished from other cancers and subclassified molecularly by recurrent alterations in chromosomes, DNA methylation, messenger and microRNA expression, or by mutations. These affect squamous cell pathways and programs that provide candidates for therapy.

Campbell et al., 2018, Cell Reports *23*, 194–212
April 3, 2018
https://doi.org/10.1016/j.celrep.2018.03.063

CellPress

OPEN ACCESS
CellPress

Cell Reports
Resource

Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas

Joshua D. Campbell,[1,2,3,36] Christina Yau,[4,5,36] Reanne Bowlby,[6,36] Yuexin Liu,[7,36] Kevin Brennan,[8,36] Huihui Fan,[9,36] Alison M. Taylor,[1,2,36] Chen Wang,[10,36] Vonn Walter,[11,12,36] Rehan Akbani,[7,36] Lauren Averett Byers,[7,13,36] Chad J. Creighton,[7,14,36] Cristian Coarfa,[15,36] Juliann Shih,[1,2] Andrew D. Cherniack,[1,2] Olivier Gevaert,[8] Marcos Prunello,[8] Hui Shen,[9] Pavana Anur,[16] Jianhong Chen,[17] Hui Cheng,[17] D. Neil Hayes,[12] Susan Bullman,[1,2] Chandra Sekhar Pedamallu,[1,2] Akinyemi I. Ojesina,[18,19] Sara Sadeghi,[6] Karen L. Mungall,[6] A. Gordon Robertson,[6] Christopher Benz,[5] Andre Schultz,[7] Rupa S. Kanchi,[7] Carl M. Gay,[13] Apurva Hegde,[7] Lixia Diao,[7] Jing Wang,[7] Wencai Ma,[7] Pavel Sumazin,[20] Hua-Sheng Chiu,[20] Ting-Wen Chen,[20] Preethi Gunaratne,[21,22] Larry Donehower,[23] Janet S. Rader,[24] Rosemary Zuna,[25] Hikmat Al-Ahmadie,[26] Alexander J. Lazar,[27] Elsa R. Flores,[28] Kenneth Y. Tsai,[29] Jane H. Zhou,[30] Anil K. Rustgi,[31] Esther Drill,[32] Ronglei Shen,[32] Christopher K. Wong,[33] The Cancer Genome Atlas Research Network, Joshua M. Stuart,[33] Peter W. Laird,[9] Katherine A. Hoadley,[34] John N. Weinstein,[7] Myron Peto,[16] Curtis R. Pickering,[35] Zhong Chen,[17,*] and Carter Van Waes[17,37,*]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2]The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142, USA
[3]Boston University School of Medicine, Boston, MA 02118, USA
[4]Department of Surgery, University of California, San Francisco, San Francisco, CA 94115, USA
[5]Buck Institute for Research on Aging, 8001 Redwood Boulevard, Novato, CA 94945, USA
[6]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC V5Z 4S6, Canada
[7]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[8]Department of Medicine-Biomedical Informatics Research, Stanford University, Stanford, CA 94305, USA
[9]Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI 49503, USA
[10]Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA
[11]Department of Public Health Sciences, Penn State Milton Hershey Medical Center, Hershey, PA 17033, USA
[12]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[13]Department of Thoracic/Head & Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[14]Department of Medicine and Dan L Duncan Comprehensive Cancer Center Division of Biostatistics, Baylor College of Medicine, Houston, TX 77030, USA
[15]Department of Molecular & Cell Biology, Baylor College of Medicine, Houston, TX 77030, USA
[16]Department of Molecular & Medical Genetics, Oregon Health & Science University, Portland, OR 97201, USA
[17]Head and Neck Surgery Branch, National Institute on Deafness and Other Communication Disorders, NIH, Bethesda, MD 20892, USA
[18]Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA
[19]Hudson Alpha Institute for Biotechnology, Huntsville, AL 35806, USA

(Affiliations continued on next page)

SUMMARY

This integrated, multiplatform PanCancer Atlas study co-mapped and identified distinguishing molecular features of squamous cell carcinomas (SCCs) from five sites associated with smoking and/or human papillomavirus (HPV). SCCs harbor 3q, 5p, and other recurrent chromosomal copy-number alterations (CNAs), DNA mutations, and/or aberrant methylation of genes and microRNAs, which are correlated with the expression of multi-gene programs linked to squamous cell stemness, epithelial-to-mesenchymal differentiation, growth, genomic integrity, oxidative damage, death, and inflammation. Low-CNA SCCs tended to be HPV(+) and display hypermethylation with repression of *TET1* demethylase and *FANCF*, previously linked to predisposition to SCC, or harbor mutations affecting CASP8, RAS-MAPK pathways, chromatin modifiers, and immunoregulatory molecules. We uncovered hypomethylation of the alternative promoter that drives expression of the *ΔNp63* oncogene and embedded *miR944*. Co-expression of immune checkpoint, T-regulatory, and Myeloid suppressor cells signatures may explain reduced efficacy of immune therapy. These findings support possibilities for molecular classification and therapeutic approaches.

INTRODUCTION

Squamous cell carcinomas (SCCs) are common cancers that can arise from the epithelia of the aerodigestive and genitourinary tracts. They share histological characteristics, which are of limited value for predicting site of origin, cause, clinical behavior, prognosis, or optimal therapy. The Cancer Genome Atlas (TCGA) recently completed initial analyses of mutations,

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[20]Department of Medicine-Pediatrics, Texas Children's Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA
[21]Department of Biology & Biochemistry, UH-SeqNEdit Core, University of Houston, Houston, TX 77204, USA
[22]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA
[23]Center for Comparative Medicine, Baylor College of Medicine, Houston, TX 77030, USA
[24]Department of Obstetrics and Gynecology, Medical College of Wisconsin, Milwaukee, WI 53226, USA
[25]University of Oklahoma Health Sciences Center, Department of Pathology, Oklahoma City, OK 73104, USA
[26]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[27]Departments of Pathology, Genomic Medicine, Dermatology, and Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77401, USA
[28]Molecular Oncology, Moffitt Cancer Center, Tampa, FL 33612, USA
[29]Departments of Anatomic Pathology and Tumor Biology, Moffitt Cancer Center, Tampa, FL 33612, USA
[30]Department of Pathology, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA
[31]Division of Gastroenterology, Departments of Medicine and Genetics, Abramson Cancer Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA
[32]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[33]Department of Biomolecular Engineering, Center for Biomolecular Sciences and Engineering University of California, Santa Cruz, Santa Cruz, CA 95064, USA
[34]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[35]Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[36]These authors contributed equally
[37]Lead Contact
*Correspondence: chenz@nidcd.nih.gov (Z.C.), vanwaesc@nidcd.nih.gov (C.V.W.)
https://doi.org/10.1016/j.celrep.2018.03.063

DNA copy-number alterations, DNA methylation, RNA/microRNA, and protein expression for SCCs from 5 individual sites, including lung (LUSC), head and neck (HNSC), esophageal (ESCA), cervical (CESC), and bladder cancers (BLCA) (Cancer Genome Atlas Network, 2015; Cancer Genome Atlas Research Network, 2012, 2014; Cancer Genome Atlas Research Network et al., 2017a, 2017b). Those studies highlighted selected genomic alterations of potential biologic or therapeutic interest in tumors from these sites, and related to tobacco use and human papillomavirus (HPV) infection. Previous comparisons (PanCan-12; Dotto and Rustgi, 2016; Hoadley et al., 2014) suggested that tumors from these different sites share some common molecular signatures. Since then, TCGA datasets have been reanalyzed and nearly doubled with new data for ~1,400 squamous cancers, and they have expanded to include ~10,000 tumors of 33 cancer types. These provide an opportunity to use newer tools to integrate omics data toward a better molecular taxonomy for SCCs and their subtypes and identify features and relationships of biologic and clinical relevance for future investigation.

This pursuit of a molecular taxonomy of SCCs and their subtypes has been aided by the availability of newer analytical tools and computational resources. We used TumorMap (TM) (Newton et al., 2017), an interactive visualization and analysis portal, coupled with integrated Cluster (iCluster [iC]) (Shen et al., 2009), and we found high overlap with original histopathologic classifications of SCC. Further, these tools uncovered broader and subtype-related genetic and epigenetic alterations that distinguish SCCs from other cancers and from one another. We examined the complex recurrent chromosomal alterations and methylation patterns underlying genome-wide mRNA expression observed in SCCs using MVisAGe (for Modeling, Visualizing and Analyzing the Cancer Genome) and MethylMix (Gevaert, 2015). These identified recurrent chromosomal alterations and CpG methylation strongly correlated with the expression of multiple genes that converge on pathways and functions relevant to SCC biology and therapeutics. mRNA clustering viewed using interactive Next-Generation Clustered Heatmaps (NG-CHMs) (Broom et al., 2017), and an updated Pathway Recognition Algorithm using Data Integration on Genomic Models (PARADIGM) tool (Vaske et al., 2010), helped to integrate omics data with pathways related to squamous cell stemness, differentiation, growth, immortalization, proliferation, survival, and inflammation. Clustered mRNA alterations for immune checkpoint PD-L1, cytokines, and cell determinants were deconvoluted using validated gene signatures for immune cell types and CIBERSORT, revealing overlap between effector T cell and immune checkpoint signatures with those of T-regulatory and Myeloid suppressor cells, which are linked to reduced efficacy of immune therapy (Charoentong et al., 2017; Gentles et al., 2015). These analyses and findings have the potential to influence how we classify SCCs into molecular subtypes, with possible implications for diagnosis, prognosis, and therapy. They also provide an atlas of organized datasets for further hypotheses generation and exploration by the large communities of biological and clinical researchers who are investigating squamous malignancies.

RESULTS

TM and iC Identify Significant Features Distinguishing SCCs and Subtypes among PanCancer-33 Tumors

To identify a molecular signature-based classification, we conducted an integrated TM and iC analysis of 9,759 tumor samples from PanCancer-33 cancers for which DNA copy-number

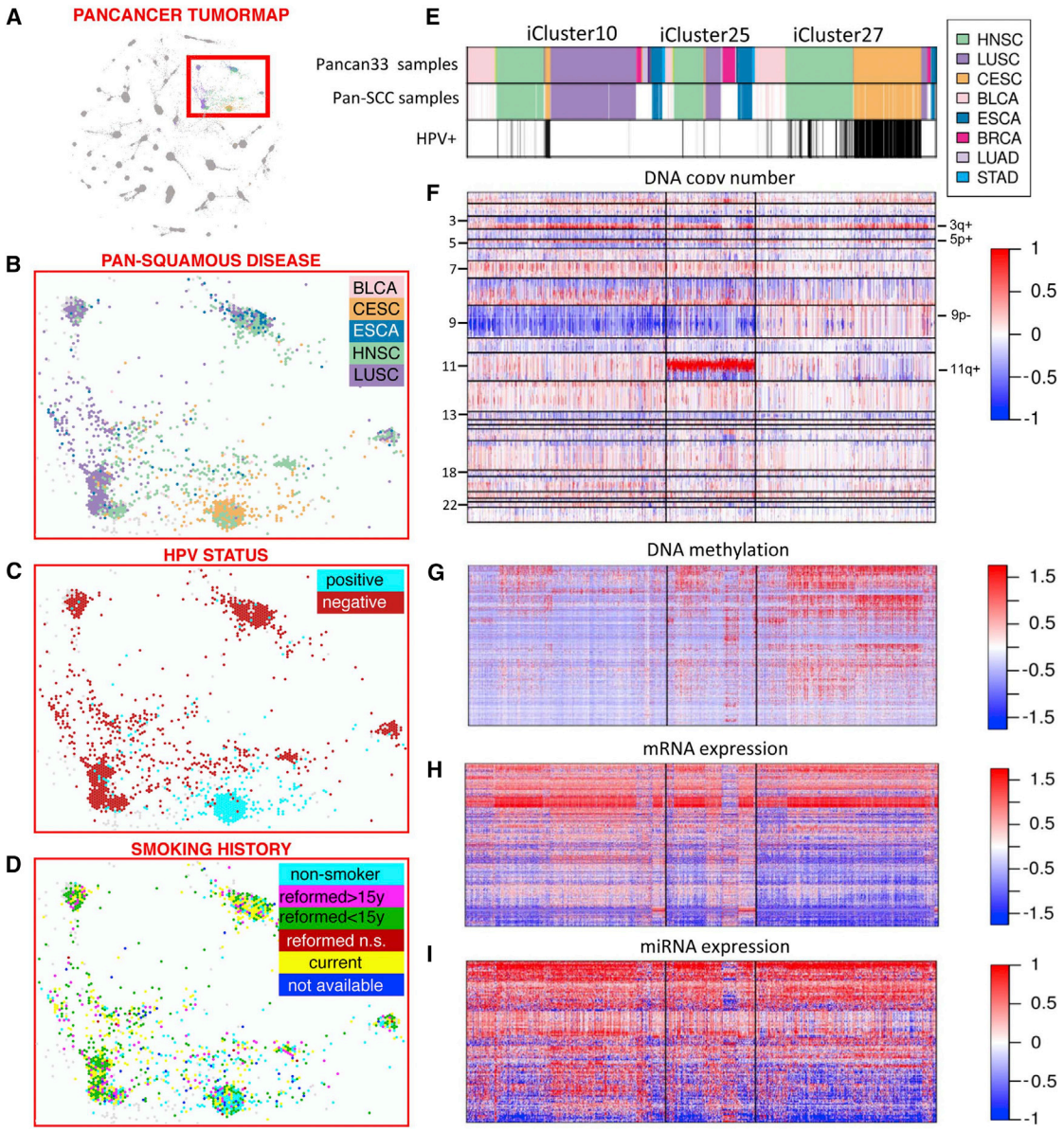


Figure 1. TumorMap and iCluster of Squamous Cancers from PanCancer-33 Analysis

(A) TumorMap analysis visualizing close mapping of LUSC, HNSC, ESCA, CESC, and BLCA among 28 PanCancer-33 islands.

(B) Higher resolution view of TM islands and distribution of SCC from 5 sites.

(C) HPV status showing the majority of HPV(+) CESC and HNSC map around a distinct island.

(D) Smoking history of SCC. Each spot in the map represents a sample. The colors of the sample spots represent attributes as described for each panel.

(E–I) Summary of iCluster analysis (E), DNA copy-number (F), methylation (G), mRNA (H), and miRNA (I) expression. PanCancer-33 SCC and other tumors and Pan-SCC from 5 sites identified by histopathologic diagnosis cluster within iC10, iC25, and iC27. Annotation bars show cancer type and HPV status, and keys show an increase (red) or decrease (blue) in features as indicated: DNA copy number, copy-number log ratio (tumor versus normal); DNA methylation, normalized beta values; miRNA expression, normalized log expression counts; miRNA expression, normalized log expression counts.

alteration (CNA), methylation, mRNA, microRNA (miRNA), and a smaller set of protein expression profiles were available. SCC tumors from 5 sites of origin (LUSC, HNSC, CESC, ESCA, and BLCA) were found to overlap 5/28 islands closely co-localized by TM and 3 major iCs when compared to other cancers (Figures 1A–1I). Most HPV+ CESC and HNSC samples mapped closest to a distinct TM island and iC27 ($p < 0.001$) enriched for lifelong non-smoking individuals, while most HPV(−) cancers mapped to nearby islands and iC10 and iC25, associated with distinct molecular patterns, tissues of origin, and smoking history.

SCCs segregated into major subtypes by CNA, methylation, and RNA/miRNA expression patterns, underpinned by significant molecular features in SCC versus non-SCC, and between SCCs (Figures 1E–1I; Tables S1A–S1L). All three major

SCC-related clusters included significant chromosome 3q and 5p copy gains (Figures 1E and 1F; Table S1A). iC10/25 displayed 9p losses, and iC25 harbored 11q gains. Many iC10/25 HPV(−) SCC tumors were associated with higher DNA CNA and broad hypomethylation, with corresponding patterns of increased mRNA and miRNA expression (Figures 1F–1I). The majority of iC27 HPV(+) CESCs, HNSCs, and some HPV(−) SCCs exhibited lower genomic DNA CNAs and wider hypermethylation, with a broader decrease in mRNAs and miRNAs. These observations suggest that most SCCs are driven by a combination of recurrent CN and other alterations, while HPV, epigenetic, or other alterations may have a greater role in subtypes with fewer CNAs. Overall, mRNA expression in SCC was enriched for 3q genes SOX2, TP63, and TP73, implicated in squamous stemness and differentiation, and immune chemokines, cytochrome, oxidative reduction, and cell adhesion pathway-related genes (Tables S1C, S1F, and S1G).

Strikingly, this multiplatform molecular classification by TM/iC co-mapped together all 1,341 (100%) of 1,409 tumors with squamous histopathologic diagnosis for which data for the 4 platforms were available, among 1,481 tumors from PanCan-33 (Figures 1A, 1B, and 1E; Table S1M). Additional BLCA tumors clustered with BLCA with histopathologic squamous differentiation, suggesting more of these cancers share squamous molecular features than appreciated by pathologic criteria. A fraction of PanCan-33 breast, lung, and esophageal adenocarcinomas shared molecular features and co-clustered with SCC (Figures 1E–1I), similar to the PanCan-12 study (Hoadley et al., 2014). We used 1,409 tumors confirmed to have squamous histology for further Pan-SCC analyses below, for which clinical, individual platform, and HPV classification are included in Tables S1M–S1P. DNA copy-number, mutations, methylation, mRNA, miRNA, and protein expression analyses are aggregated in Tables S2A–S2N, S3A, S3B, S4A–S4F, S5A, and S5B).

DNA CNAs Correlate with Expression of mRNAs in Key Growth, Mitotic DNA Integrity, Chromatin Modifier, and Death Pathways

To explore the relationship of recurrent chromosomal CNAs with mRNA expression genome-wide, Pan-SCC CNAs were correlated with expression for each coding region using MVisAGe. Smoothed Pearson correlation coefficients (ρ values) were plotted to identify chromosomal regions for which CNA was most highly correlated with gene expression, and selected individual genes with $\rho \geq 0.6$ were highlighted (Figure 2A; Table S3). This revealed broad and focal chromosomal regions for which CNAs were highly correlated with the expression of multiple genes, in addition to those within CNA peaks found by genomic identification of significant targets in cancer (GISTIC) analyses (Figures 2B–2G, S1A, and S1B; Table S3). Remarkably, many of these genes on the same or different chromosomes are implicated in related pathways and functions.

Chromosome 3q most significantly associated with SCC by iC (Table S1A) showed the highest correlation of CN gain with expression for multiple genes in a broad peak between 3q24 and 3q29 (Figure 2B; ~160–190 Mb). Strikingly, *ACTL6A* at the top peak in 3q26 was recently associated with worse prognosis, and it was reported to form a novel complex with oncogenic N-terminal-truncated ΔNp63 isoforms of the nearby 3q28 squamous differentiation gene *TP63* in HNSC (Saladi et al., 2017). Unexpectedly, the CN/expression correlation for *TP63* was lower than for other nearby genes, and it was associated with predominant expression of the ΔNp63α isoform for all 5 sites (Figure 2H), consistent with epigenetic regulation of these alternatively transcribed isoforms discovered below. The ACTLA6/ΔNp63α complex can cooperatively drive a transcriptional program that suppresses differentiation and promotes activation of Hippo growth pathway transcriptional co-factor *YAP1*. Intriguingly, we found 11q22 amplification to be highly correlated with *YAP1* expression, and enrichment for this amplicon in mostly HPV(+) SCCs displayed relative mutual exclusivity with higher 3q amplifications harboring *ACTL6A* and *TP63* in the Pan-SCC dataset (Figures 2E and S1C; Fisher's exact test, $p = 0.007$). These observations suggest that 3q26 or 11q22 CNAs could be alternative drivers orchestrating deregulation of *ACTLA6*/*TP63* differentiation and Hippo growth pathway *YAP1* gene expression in SCC subtypes. 3q26 and 11q22 gains also strongly correlated with the expression of additional genes implicated in cell stemness (*SOX2* and *PRKCI*), immortalization (*TERC* and *FXR1*) WNT/β-catenin differentiation (*DVL3*), growth (*PIK3CA* and *ZNF639*), and survival (*BIRC2*).

Chromosome 5p gains that distinguished Pan-SCC tumors by iC correlated with the expression of genes linked to chromosomal instability and mitosis (Figure 2C). *TRIP13* can promote error-prone non-homologous end joining, cell proliferation, survival, and cisplatin chemoresistance in HNSC (Banerjee et al., 2014), and it can cooperate with chaperonin *CCT* in regulating the mitotic assembly and checkpoint system (Kaisari et al., 2017). 5p gene *TERT* and 3q gene *TERC* form telomerase subunits important in stability of chromosomal tips, and they are associated with syndromes at increased risk of HNSC and genito-urinary (GU) tract SCC (Alter et al., 2013). Together, alteration of 5p genes with these functions is consistent with the generation of increased CNAs found in most SCCs.

Chr 8p11 CNAs correlate with the expression of chromosomal modifier *WHSC1L1/NSD3* in a subset enriched for HPV(−) SCC (Figure 2D). This encodes a novel methyltransferase recently found to promote monomethylation of histones and signal activation of membrane and nuclear epidermal growth factor receptor (EGFR) (Saloura et al., 2016, 2017). Chr 11q gene *KDM2A* is a histone demethylase implicated in the activation of genes involved in stemness, differentiation, and inflammation (Chen et al., 2017).

Chromosome 11q13/22, 5p15, and 14q32 CNAs correlate with expression of multiple components of the nuclear factor κB (NF-κB)/REL- and ATM-signaling axes involved in cell survival or death (Derakhshan et al., 2017) (Figures 2C–2F). These include tumor necrosis factor receptor (TNFR)-associated Fas-associated death domain (*FADD*) and inhibitor of apoptosis proteins (IAPs) encoded by *BIRC2/3*, which can complex to promote NF-κB survival over cell death signaling. This complex can recruit IKKβ encoded by *IKBKB* to enhance the activation of NF-κB *RELA*, which is a transcriptional enhancer of cyclin *CCND1* and prosurvival genes. These alterations in the extrinsic death pathway may be complemented by loss of *ATM* and gain of *FASTKD3* expression, which are implicated in inhibiting the

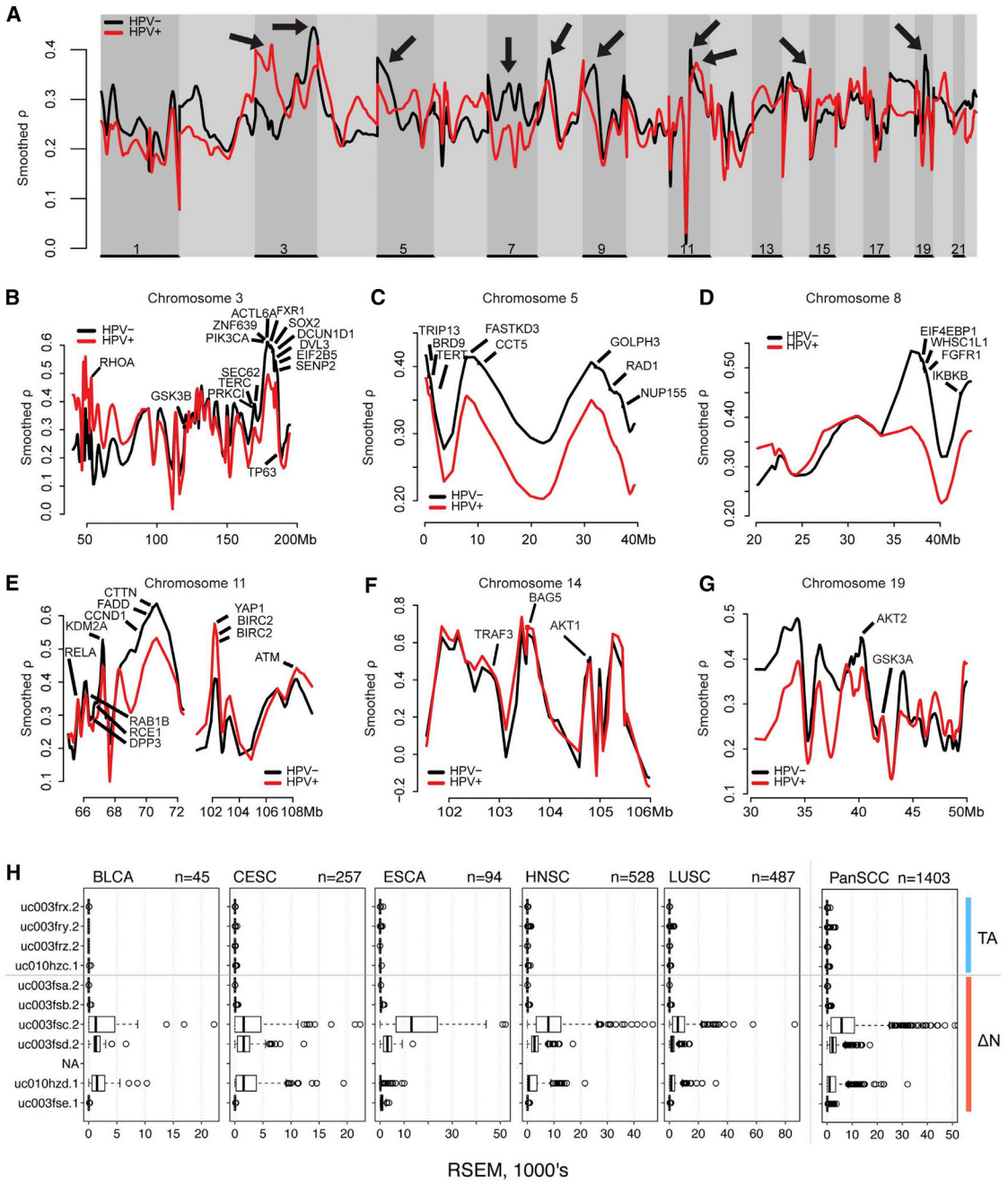


Figure 2. Correlation between DNA Copy Number of Chromosomal Regions and Expression of Multiple Genes, and Predominant Expression of ΔNp63 Isoforms of TP63 Gene for 5 Pan-SCC Tumor Sites

The MVisAGe R-package was used to compute and plot gene-level Pearson correlation coefficients (ρ values) based on quantitative measurements of DNA copy number (CN) and log2(RSEM + 1) gene expression measurements for Pan-SCC data.

(A) Smoothed ρ values plotted for all chromosomes, with arrows highlighting regions of peak correlation between CN and expression for HPV(−) (black) and (+) (red) SCC.

(B–G) Smoothed ρ values and selected genes with individual unsmoothed ρ > 0.6 plotted based on genomic positions in selected regions of chr3q (B), 5p (C), 8p (D), 11q13/q22 (E), 14q (F), and 19 (G).

(H) TP63 isoform mRNA abundance (RSEM) for full transactivating (TA) domain or alternatively transcribed N-terminally truncated (ΔN) isoforms in Pan-SCC tumors. ΔNp63α (uc003fsc.2) and other ΔN isoforms are preferentially expressed compared to TA isoforms. Boxplots show median values and the 25th to 75th percentile range in the data, i.e. the interquartile range (IQR). Whisker bars extend 1.5 times the IQR.

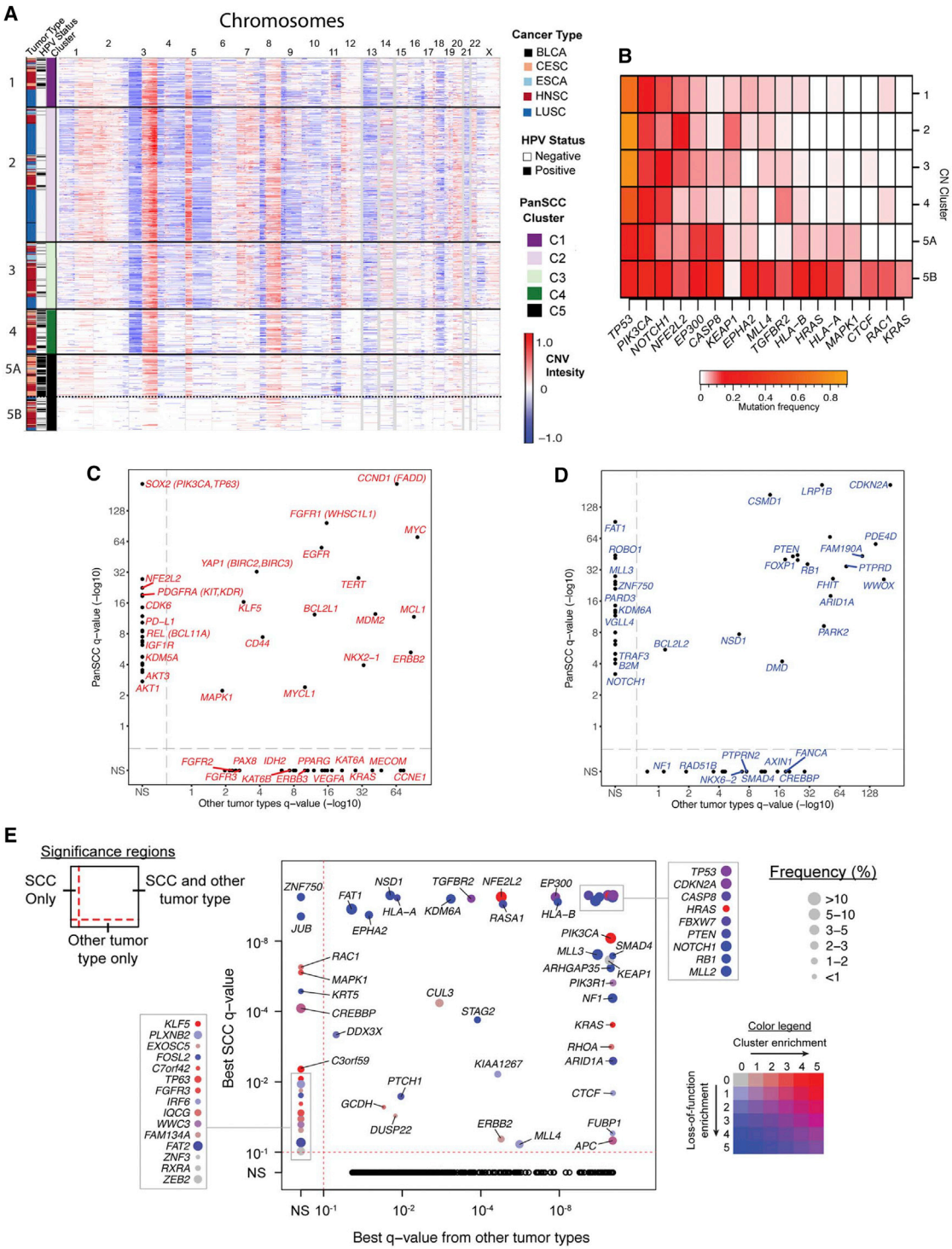


Figure 3. Common, Unique, and Heterogeneous Genomic Alterations in Squamous Cell Carcinomas

(A) Unsupervised clustering of CNAs in 1,386 squamous cell carcinomas revealed five distinct clusters, with higher recurrent amplifications or deletions or with few focal alterations. Color bars at the left indicate the 5 tumor types (HNSCs, LUSCs, ESCAs, CESCs, and BLCAs), HPV status, and CNA cluster. Red indicates copy gain, blue indicates copy loss, and white indicates copy-number neutrality.

(B) 63 genes were significantly mutated in one or more of 5 tumors in the Pan-SCC cohort (MutSig2CV analysis; FDR q-value < 0.1), and the mutation frequencies of 17 of the genes that were correlated with CNA cluster are indicated.

(legend continued on next page)

intrinsic mitochondrial cytochrome-mediated cell death pathway (Simarro et al., 2010). Copy loss of TNFR-associated factor *TRAF3* has recently been implicated as a tumor suppressor of NF-κB gene expression and HPV infection, and it is a marker for HPV(+) HNSC tumors with better prognosis (Hajek et al., 2017).

This analysis also reveals CN-driven expression across several chromosomes of multiple components of the PI3K-AKT-mTOR-eIF pathway important in cell metabolism, protein expression, growth, and survival (Figures 2B–2D, 2F, and 2G). These include 3q amplicon genes *PIK3CA* and *EIF2B2*, 5p gene *GOLPH3*, 8p gene *EIF4EBP1*, and chromosome (chr) 14 or 19 genes *AKT1/2*. PI3K-AKT signaling has been implicated in the activation of 3q transcription factor *SOX2* and stemness, alternative transcription of ΔNp63, and phosphorylation and function of YAP1 in complex with ΔNp63 (Barbieri et al., 2003; Ehsanian et al., 2010). Together, the significance of these CN alterations, distinguishing major subsets of SCC by iC (Figure 1B; Table S1A), and strongly correlated expression by MVisAGe (Figure 2), support their roles as important drivers of SCC.

Relationships among DNA CNAs, HPV Status, and Mutations Affecting Genes Involved in Genomic Integrity, Mitogen and Death Pathways, and Chromatin Modification

Integration of unsupervised hierarchical clustering of significant CNAs, available for 1,386 samples of squamous histology, HPV status, and significant mutations, helped resolve different candidate drivers of high- and low-copy-number variation (CNV) subtypes (Figures 3A, 3B, and S2A–S2C). We resolved 5 major clusters, including higher to lower CNA C1–4, and a copy-quiet C5 with a sub-cluster C5A enriched for HPV(+) tumors (Figure 3A). C1–4 with higher CNAs displayed 5p amplification and the highest frequency of deleterious mutations of *TP53*, consistent with their function in maintaining genomic integrity. Mutations in *NFE2L2* and *KEAP1*, important in oxidative damage, were also enriched in C1–3. Low-CNA C5A and B tumors were enriched for mutations in (1) epigenetic modifiers *EP300*, *MLL4*, and *CTCF*; (2) mitogen pathway components *EPHA2*, *HRAS*, *MAPK1*, and *RAC1*; and (3) cell death mediator caspase *CASP8* (Figures 1A, 1B, and S2B). Intriguingly, EP300 is a chromatin modifier recently linked to the enhancement of target gene activation by stemness transcription factor *SOX2*, which is amplified on 3q in higher CNA SCCs (Kim et al., 2017), and these alterations tended toward mutual exclusivity in CNA versus quiet subtypes ($p = 0.004$). Mutations in *EPHA2*, *HRAS*, *MAPK1*, and *RAC1* cumulatively affected ~27% and 46% of C5 and C5A tumors, with EPHA2 and HRAS mutations tending toward mutual exclusivity across all C5 samples (Figure S2B; $p = 0.037$). EPHA2 mutations were enriched for truncating alterations, consistent with evidence that it serves as a negative regulator of RAS pathway signaling (Macrae et al., 2005). Conversely, HRAS, MAPK1, and RAC1 showed missense hotspot mutations (Figure S2C), implicated in signal activation. HRAS and CASP8 significantly co-occured (Figure S2B; $p = 0.001$), suggesting CASP8 inactivation may be linked to escape from HRAS-induced senescence. C5A SCC displayed mutations of *HLA-A* and *-B* and deletions of *B2M*, implicated in immune escape (Figure 3B).

We examined if significant CNA or mutated genes are more significantly altered in SCCs, other cancers, or both (Figures 3C–3E, S1A, and S1B). SCC-related alterations underscore the importance of those implicated in stemness (*SOX2*), oxidative DNA damage response (*NFE2L2*), mitogenic growth and cell cycle (*PDGFRA*, *IGF1R*, *CDK6*, *RAC1*, *MAPK1*, *EPHA2*, and *CREBBP*), PI3K signaling (*AKT1/3*), NF-κB signaling (*REL* and *TRAF3*), squamous differentiation (*FAT1/2*, *ROBO1*, *ZNF750*, *JUB*, *NOTCH1*, and *TP63*), chromatin modifiers (*KDM5A/6A*, *MLL3*, and *NSD1*), and immune escape (*PD-L1* and *B2M*). CN and mutations inactivating *FAT1* trended toward mutual exclusivity with amplification of *YAP1* ($p = 0.08$), consistent with a role of FAT1 as a negative regulator of Hippo growth pathway (Gao et al., 2014). Interestingly, these were exclusive of amplifications of 3q gene *PIK3CA* ($p = 0.005$) or mutations of *PTEN* ($p = 0.002$), which could potentially enhance AKT signaling implicated in YAP1 inactivation via cytoplasmic sequestration (Ehsanian et al., 2010). Inactivating deletions or mutations of *TP63* and *ZNF750* support possible alternative mechanisms for deregulation of the TP63-ZNF750 differentiation pathways (Figures 2D, 2E, S2D, and S2E) (Okuyama et al., 2007; Sen et al., 2012). JUB has been linked as a negative regulator of the WNT pathway (Haraguchi et al., 2008).

Integration of DNA Methylation, mRNA Expression, and Mutations Uncovers Chromatin Modifier, Fanconi DNA Repair, and SRC Kinase Family Signatures

To identify significant alterations in CpG island methylation between tumor and normal and inverse correlations with expression of their corresponding mRNAs, we used the recently developed MethylMix program (Gevaert, 2015). 905 differentially methylated and expressed genes were identified and assorted by consensus clustering into 5 groups (Figure 4A; Table S2K). Notably, hypermethylated C2 enriched for HPV(+) CESC and HNSC ($p < 2.2E-16$) predominantly overlapped the low-CNA cluster C5A (Figures 4A, 4B, and S3A; Fisher's exact test for CNV-MethylMix Clusters, $p = 1E-5$). Hypermethylated C4 overlapped copy-quiet CNA C5B and C3 and C4 with mostly HNSC. Hypomethylated C1, C3, and C5 overlapped with higher CNA C1–3 enriched for HPV(−) LUSC, HNSC, ESCA, and BLCA. Among 28/51 genes significantly mutated and differentially distributed among the methylation clusters in SCC (Table S2L),

(C and D) The q-values for (C) recurrent amplifications and (D) deletions in the Pan-SCC cohort (y axis) are plotted against q-values for the same gene in the cohort of 27 non-SCC tumor types (x axis). Genes in the top left and bottom right quadrants are significantly altered exclusively in the Pan-SCC and non-SCC cohorts, respectively; genes in the top right are significantly altered in both.

(E) The best q-value for each significantly mutated gene across all SCC types (x axis) is plotted against the best q-value for the same gene in the 27 other tumor types (y axis). Point size is proportional to the frequency of mutations in the gene in the Pan-SCC cohort. Point color indicates enrichment for mutation clustering defined by MutSig2CV (–log10 pCL) and/or enrichment for gain- or loss-of-function mutations (–log10 p value; Fisher's exact test) in the Pan-SCC cohort. Black circles in the lower quadrant indicate genes more significant in another cancer type, compared to SCC tumor types.

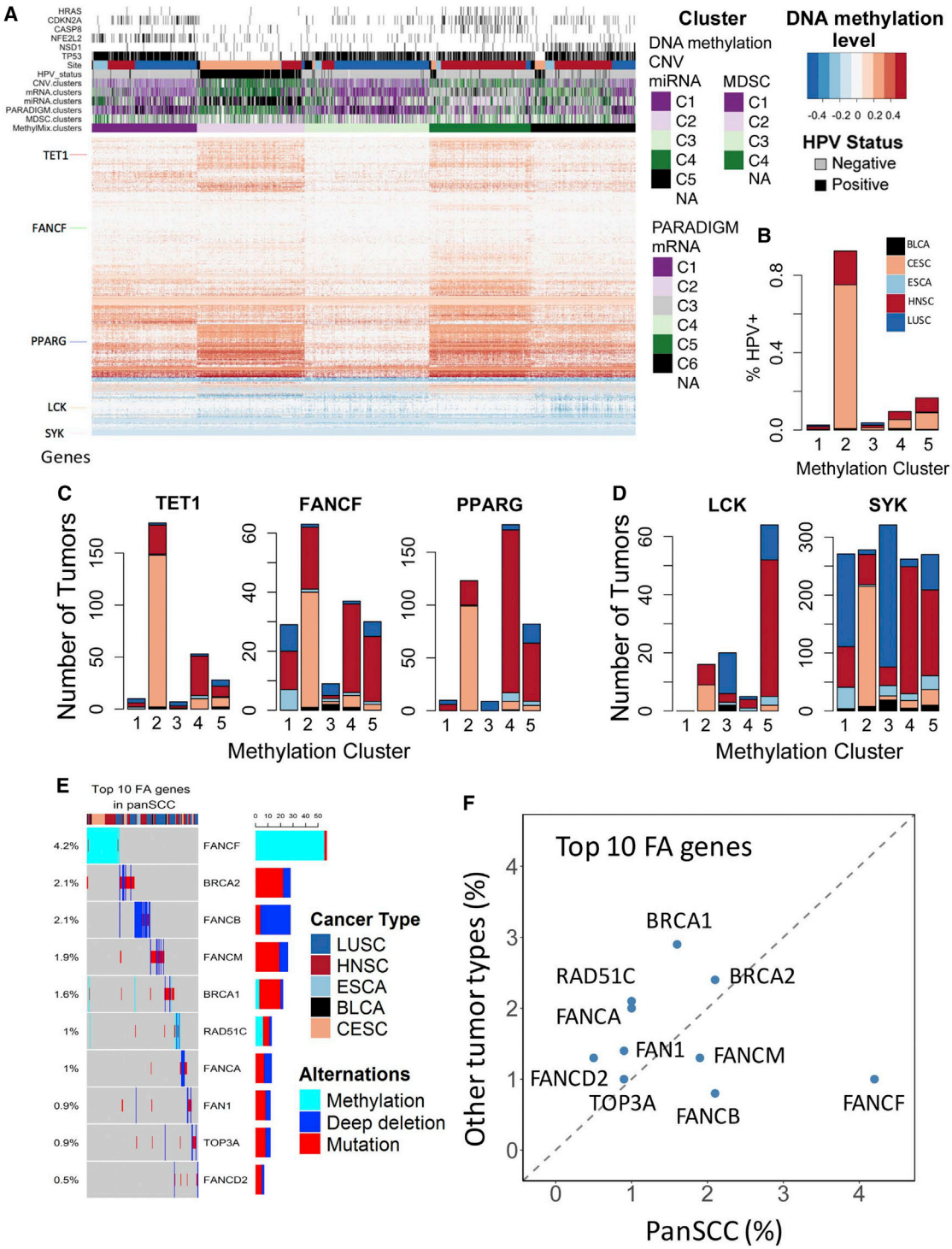


Figure 4. DNA Methylation Consensus Clusters with Distinct Mutation and HPV Profiles, and Unique DNA Damage and Repair Genes in Squamous Cell Carcinomas

(A) MethylMix identified 905 abnormally methylated genes inversely associated mRNA expression, and that formed five DNA methylation consensus clusters presented in the heatmap. Top bars indicate DNA methylation clusters, cancer types, HPV status, mutations in genes, and other platform clusters that are significantly differentially distributed between DNA methylation clusters. Brown, hypermethylation; blue, hypomethylation.

(B) Variability in the percentage of patients within each DNA methylation cluster that are HPV positive. Bar colors indicate the portion of different cancer types among HPV-positive patients within each methylation cluster.

(legend continued on next page)

hypermethylated HPV-enriched C2 also harbored fewer mutations in *HRAS*, *CDKN2A(p16)*, *CASP8*, *NFE2L2*, *NSD1*, and *TP53* than did clusters with predominantly HPV(−) SCC (Figures 4A and S3B). Strikingly, hypomethylation in C5 was linked to inactivating mutations in the H3K36 histone methyltransferase NSD1, defining a distinct subtype across SCC tissue sites previously observed in HNSC (Cancer Genome Atlas Network, 2015; Papillon-Cavanagh et al., 2017).

Several new differentially methylated and expressed genes in SCC clusters have been causally implicated in cancer development in Catalogue of Somatic Mutations in Cancer (COSMIC) (Figures 4C and 4D; Table S2K). These include hypermethylated, repressed genes *TET1*, *FANCF*, and *PPARG*, enriched in C2 HPV(+) CESC and HNSC and C4 with HPV(−)HNSC (Figure 4C). *TET1* is a demethylase whose inactivation is implicated in sustaining CpG hypermethylation in cancer (Li et al., 2016), consistent with hypermethylation found in C2 and C4. *FANCF* is a component of the Fanconi-BRCA pathway essential for DNA repair by non-homologous recombination (Ceccaldi et al., 2016). A broader analysis of FANC and DNA damage repair pathway genes revealed an unexpectedly high frequency (~12%) of somatic methylation, CNAs, and mutations affecting *FANC-BRCA* genes in SCC (Figure 4E), suggesting that acquired as well as germline alterations in this pathway may contribute to the development of a subset of SCC (Alter et al., 2013; Ceccaldi et al., 2016). Of these, *FANCF* methylation is more often observed in Pan-SCC than other PanCan-33 tumors (Figures 4F and S3C; chi-square, $p < 2.2E{-}16$). *PPARG* encodes a nuclear receptor and transcriptional modulator of squamous differentiation of interest as a target for chemoprevention (McCormick et al., 2015). Hypomethylated, overexpressed genes included *LCK* in C5 and *SYK* (Figure 4D; Tables S2K and S2L). These are SRC family kinases implicated in signal activation of STAT transcription factors in SCC and in activated immune cells expressing immunoregulatory checkpoint molecules (Lund et al., 1999; Ma et al., 2015; Sen et al., 2015).

mRNA Analyses Identify SCC Subtypes Differentially Expressing 3q/11q, Oxidative DNA Damage, EMT, Transcription Factor, and Immune Signatures

To determine how genomic, epigenetic, and transcriptional alterations may relate to wider mRNA expression in SCC subtypes, we performed unsupervised consensus cluster analysis for 1,867 annotated cancer-related genes (Sadelain et al., 2011). K-means discriminated 6 mRNA expression clusters that included mRNAs linked to significant CN, methylation, and miRNA-related alterations found via other platforms in this study (Figures 5A, 5B, and S4). Broadly, mRNA C1 with LUSC and other SCCs displayed higher expression of lymphocyte kinase (LCK), immune checkpoint PD-L1(CD274), T-regulatory (FOX3P), and Myeloid-Derived Suppressor Cell (IDO1) immunoregulatory mRNA markers. Supporting the alternative CNAs in 3q or 11q22 observed above, C2 tumors displayed a significantly higher expression of 3q (SOX2 and PIK3CA) mRNAs and lower 11q22 (BIRC2/YAP1) mRNAs. Conversely, C3 and C6 showed lower expression of those, and they more highly expressed 11q22-encoded YAP1/BIRC2 mRNAs. C5 enriched for HPV(+) CESC and HNSC showed a lower expression of mRNAs for 11q22 (YAP1 and BIRC2); 14q (TRAF3); and hypermethylated genes FANCF, TET1, and PPARG (Figures 5A, 5B, and S4). C6 and C1 were enriched for HNSC and LUSC with mRNAs for ZEB2, IL-6, TWIST, SNAI1, CTGF, and CYR61 (Figures 5A, 5B, and S4), found below to be associated with miRNA clusters related to the epithelial-mesenchymal transition.

The increased expression of LCK overlapped those of immune checkpoint CD274/PDL1, Treg marker FOXP3, and myeloid derived suppressor cells (MDSCs) IDO1 mRNAs in C1, C5, and C6 subclusters (Figures 5A and S4), suggesting their expression could be linked to cellular immune responses. Another immune signature seen in C1, C3, C5, and C6 includes transcription factors NFKB1, STAT3, EGR1, and JUN/FOS, as well as TNF and chemokines CXCL1–3 mRNAs implicated in recruiting such cellular immune responses (Figures 4A and S4) (Davis et al., 2016). We explored if expression of PD-L1 overlaps signatures that were recently developed and validated in other cancers for MDSCs, CD8+ CTL, Tregs, and other immune cells (Charoentong et al., 2017; Gentles et al., 2015). Consensus clustering using an MDSC-related signature sorted 4 clusters with very high to low expression of 49 MDSC-related genes, including PD-L1 (Figure S5A). MDSC-inflamed C1 and C2 most significantly overlapped mRNA C1, 5, and 6 with increased immune checkpoint CD274/PDL1, Treg marker FOXP3, and MDSC IDO1 mRNAs (Figures 5A, S4A, and S5A, mRNA cluster tracks; Table S4A; $p = 1E{-}07$). CIBERSORT profiling for other immune cell types (Figure S5B) revealed a parallel pattern of expression for CD8 CTL, natural killer, CD4+ (resting > activated) T helper (Th), and Treg signatures. Additionally, these tumors showed a higher ratio of M2 > M1 macrophage signatures, which are linked to the suppression of Th1 and CTL tumor immunity. These observations indicate that SCC with increased CD8 CTL, natural killer (NK), and CD4 Th responses co-occur with opposing PD-L1, MDSC, and Treg signatures, providing a possible explanation and other targets for improving the limited efficacy of immune checkpoint therapy observed in SCC.

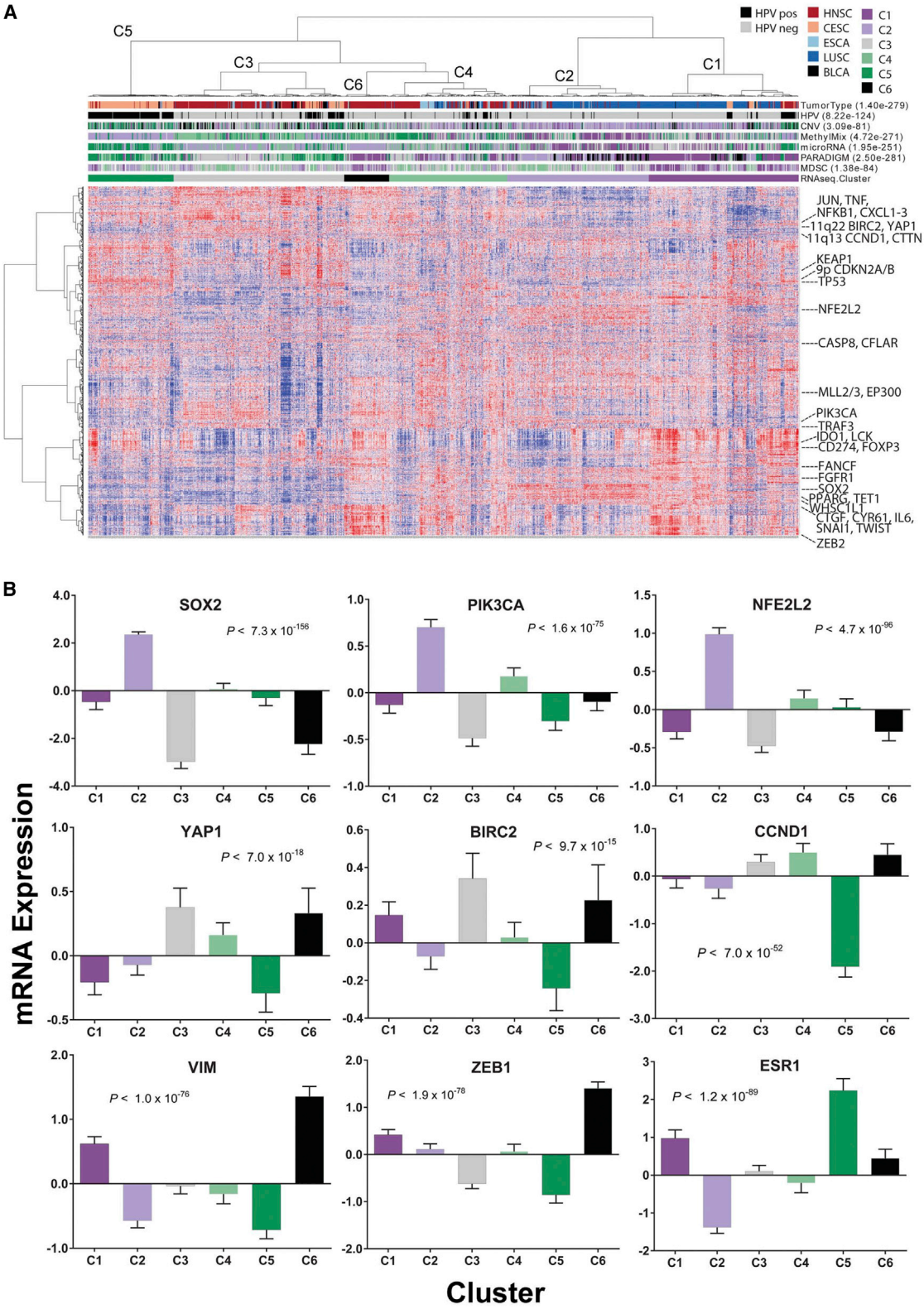
PARADIGM Pathway Analysis Distinguishes SCC Subtypes with Signaling, Transcription Factor, Immune, and Cell Cycle Signatures

To better understand the relationship between these complex patterns of mRNA expression to underlying alterations and

(C and D) Genes that are hypermethylated (C) or hypomethylated (D) and anti-correlated with mRNA expression in SCC, and annotated in COSMIC. The number and portion of tumors from 5 SCC sites displaying abnormal methylation and expression within each DNA methylation cluster are shown on the Y axes.

(E) Dysregulations of Fanconi Anemia (FA) and DNA repair pathways across squamous cell carcinomas. Oncoprint representation of frequency of mutation, deep deletion, and methylation for FA and DNA damage response pathway genes.

(F) The percentage of cancer samples with altered FA and DNA damage response genes in the Pan-SCC cohort (x axis) are plotted against for the same genes in the PanCan-33 tumor cohort (y axis). *FANCF* in the right lower region is significantly altered more frequently in the Pan-SCC cohort (2-sample test for equality of proportions, chi-square = 84.5, $p < 2.2E{-}16$).



(legend on next page)

pathways of biologic and clinical relevance, we used PARADIGM (Vaske et al., 2010). This analysis inferred the activities of ~19,000 pathway features based on expression, copy-number, and pathway interaction data for 9,829 tumor samples, including 1,373 SCCs. The analysis distinguished SCCs from other cancer types, and 6 SCC clusters were defined by hierarchical cluster analysis (Figure 6A). Several cluster pathways were significantly aligned with genomic and transcriptomic alterations defined above, when compared using Benjamini-Hochberg false discovery rate (FDR) corrections. C1, which includes predominantly LUSC and HNSC, supports relatively high inferred activation of MAPK-JUN/FOS, RELA/p50(NFKB1) complex, p53/63/73, and immune-related/STAT pathways. C1 was enriched for amplification of *MAPK1* (p = 0.001) and deletion of NF-κB negative regulator *TRAF3* (p = 3E−05), relative to other clusters. In contrast, C2, with predominantly LUSC and ESCA, showed higher inferred activation of proliferation-related cell cycle components, with enrichment for *CDK6* amplification (p = 1.3E−08), *CDKN2A* deletion (3.6E−07), a decreased immune signature, and a lower proportion of cases with amplification of immune checkpoint *PDL1* (p = 0.0003). C3 with HNSC showed MAPK-JUN-FOS, TP53/63/73, and proliferation signatures and lower immune signatures, associated with amplifications of *EGFR*, *IGF1R*, and *PDGFA* ($p \leq 0.005$). C4 and C5, with HPV+ CESC and some HPV(−) tumors, shared high proliferation-related features, but they had a lower proportion of cases with amplifications of *MAPK1* ($p \leq 6.4E{-}0.05$) and *FGFR1* (p = 0.0006). C4, which contains higher MYB/MYC negative regulator *FBXW7* mutations (p = 0.04), displayed low inferred activation of immune features, while C5 was enriched for *PDL1* (*CD274*) amplification (p = 0.0009), differentiating these HPV(+) SCC subsets. LUSC enriched cluster C6, which contained a higher proportion of cases with *CDK6* amplifications (p = 1.9E−05) and exhibited higher proliferation-related signature but lower JUN/FOS and TP53/63/73 pathway activation.

Display of underlying components of MAPK-JUN-FOS, immune-related, TP53/63/73, and proliferation-related pathways highlight the activation of important regulatory nodes in SCC (Figures 6B–6E). Consistent with overlapping expression patterns for transcription factors observed with mRNA profiling above (Figures 5A and S4A) PARADIGM revealed that JUN-FOS, RELA/p50, and STAT3 form a network of co-activated transcription factors that regulate diverse cancer and immune-related mRNAs, such as TNF, CXCL1, PTGS2, and LCK (Figures 6B and 6C). Strikingly, PARADIGM C1, C5, and C6 with increased immune signatures also appeared to closely overlap the increased MDSC C1, C2, C3, and related immune signatures (Figures S5A and S5B, PARADIGM track; Table S4B; Fisher's exact test, p = 1E−5), suggesting these pathways are linked to the co-occurring effector and deleterious immune responses observed in SCC.

Pan-SCC Protein Expression

Reverse-phase protein array (RPPA) data were obtained for 748 SCCs using a set of 189 antibodies to assess expression and phosphorylation of proteins in multiple cancer-related pathways. Unsupervised clustering as described in the STAR Methods identified 6 clusters that revealed distinguishing patterns of protein expression and pathway activity (Figure S6A; Table S2M). Notably, a C2 arm and C3 with mostly HPV(+) CESC and C4 and C5 with LUSC, ESCA, CESC, and BLCA were enriched for growth factor and rapamycin-sensitive mTORC1 target P70S6KpT389 and RAD51 DNA damage factor (Dibble et al., 2009). The HPV(+) CESC C2 arm and C5 were also enriched for the mTORC2 target RICTORpT1135. C1, a C2 HNSC-enriched branch, and C6 with mostly LUSC lacked this RICTOR signature. However, C1 was enriched for activated EGFRpY1068/1173 and HER2pY1248, as potential therapeutic targets for this subset. C2 and C6 showed increased MAPKpT202Y204. AKTp473/T308 and GSK3p21S9 were enriched in C4 arm 1 and C6.

We found positive Pearson's correlations between upstream MAPKpT202Y204 and JUN phospho-proteins, between AKT and mTOR, and among GSK3αβ, GSK3p21S9, and NF-κBpS536 (Figure 6B). These are consistent with the genomic, mRNA, and inferred pathway alterations found above and co-activation of these pathways observed in functional and preclinical studies from HNSC (Mohan et al., 2015). Subsets of C1, C2, C5, and C6 tumors with increased CAVEOLIN1, MYH11, and YAPpS127 and decreased βCATENIN correlated with higher EMT and reactive tissue scores (Figure 6A), reported in breast and other cancers characterized by profuse stromal invasion and tumor fibroblast signaling. As RPPA-robust antibodies for immune checkpoint determinants were not available at the time of these analyses, we integrated RPPA data with mRNA expression data to identify protein correlates of CTLA4 and PD-L1 mRNA expression. Increased LCK protein expression, which was found to co-cluster with PD-L1 in mRNA analyses above, was also found to correlate with CTLA4 mRNA expression across most tumor types, except ESCA (Figure 6C). Taken together, our methylation, mRNA, and RPPA profiling data highlight LCK/PDL1/CTLA expression signatures that could also be investigated as predictors of response to immune therapies.

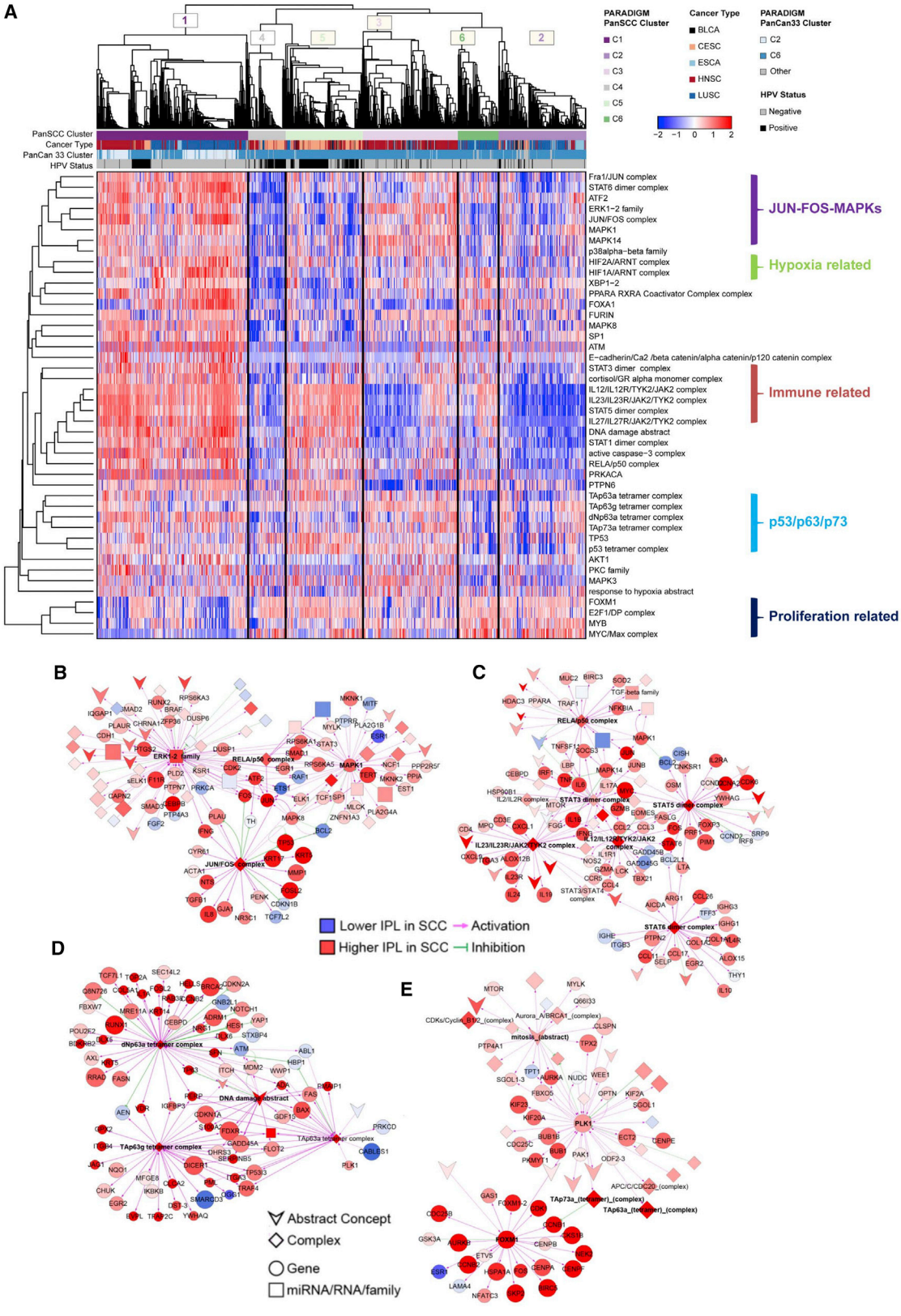
miRNAs Linked to Expression of EMT and Transcription FactorΔNp63 mRNAs and Hypomethylation in SCC

We performed unsupervised consensus clustering for 1,381 Pan-SCC samples using 270 expressed miRNA mature strands (≥25 reads per million [RPM] in at least 10% of samples), and we selected a five-cluster solution, as described in the STAR Methods (Figure S7A). This segregated HPV(−) tumors into C1–4 and most HPV(+) CESC and HNSC in C5.

Figure 5. mRNA Expression Subtypes in Squamous Cell Carcinomas.

(A) Consensus unsupervised clustering analysis of 1,867 functionally defined cancer genes resulted in the identification of six gene expression-based clusters/subtypes from the five types of squamous cell carcinomas, visualized via clustered heatmap. The cancer types, HPV status, and clusters are indicated by the annotation bars on the top. Differentially clustered oncogenes, tumor suppressor, and immune gene signatures are highlighted on the right side.

(B) The relative mRNA expression levels of genes significantly differentially expressed across Pan-SCC mRNA subtypes. Mean mRNA expression with bars representing 95% confidence intervals are shown.



(legend on next page)

Additionally, we identified miRNAs that were differentially abundant in SCC (n = 1,381) versus non-SCC (n = 9,436) tumors (Figures S7A [bold] and S7B). Of these, we highlight the two with the largest positive fold changes in SCC, miR-205-5p and miR-944, and a set that included miRs-200a-c-5/3p, 141-5/3p, and 429, which we observed to exhibit decreased expression linked with an increased EMT score in miRNA C2 and C3 (Figures 7A and S7A, EMT score track). For these miRNAs, we identified significantly anti-correlated mRNAs (FDR < 0.05, Spearman rho < −0.2) for which there was also functional evidence annotated in miRTarBase version (v.)6.0 (Figures 7B and S7C). Notably, miR-205-5p as well as miR-200/141 and 429 were anti-correlated (rho ≤ −0.4) to the EMT-related transcription factors ZEB1 and ZEB2 (Figures 7C and S7C). Other anti-correlated miR-205-5p targets potentially related to EMT included connective tissue growth factor (CTGF), cysteine-rich protein 61 (CYR61) (Lau, 2016; Thakur and Mishra, 2016; Yeger and Perbal, 2016), and the inositol phosphatase INPPL1 (SHIP2), which is involved in extracellular matrix (ECM) degradation and carcinoma invasiveness (Rajadurai et al., 2016). The EMT-related mRNAs ZEB2, CTGF, and CYR61 were observed to cluster together above in a branch of mRNA C1 with LUSC and C6 with HNSC that overlap miRNA C2 and C3 with decreased expression of these miRs (Figure S7A, mRNA track). These observations support a role for miR-205 and miR-200 family members in regulating the expression of ZEB transcription factors and EMT differentiation gene signatures in these SCC subtypes. miR-944 targets include S100PBP, implicated in adhesion; SPRY1, a modulator of EGFR signaling; and NPR1, an Inhibitor-κB homolog that attenuates NF-κB signaling (Figure 7B) (He et al., 2016; Subramanian et al., 2016).

We examined the possibility that overexpression of miR-205 and miR-944 in SCC could be related to hypomethylation of CpG sites in the transcriptional start sites (TSSs) of these miRs and their host genes. Decreased methylation of the CpG TSSs predicted for *MIR205* and other CpGs in the region of host gene *MIR205HG* was strongly anti-correlated with miR205 expression (Spearman rho > 0.5), supporting a role for regional hypomethylation in the regulation of miR205 and its EMT target genes (Table S5A). Intriguingly, *MIR944* resides within the *TP63* gene, within an intron beyond the alternative TSS for ΔNp63 isoforms, which we found to be preferentially expressed in the Pan-SCC dataset (Figures 7D and 7E). Expression of miR-944 is most strongly and significantly correlated with expression of TP63 mRNAs among all miRs across the Pan-SCC dataset (Figure 7F; r = 0.51, p < 5E−90). As the correlation of expression of TP63 with copy gain was lower than expected, we explored how both CN and methylation of TSSs and other CpG site probes for the TA and ΔNp63 isoforms affect the expression of TP63 (Figure 7D). We discovered that two CpG sites that were nearest the TSS for *ΔNp63* and an experimentally determined TSS for *MIR944* (Budach et al., 2016) were associated with lower CN coefficients and negative methylation coefficients when compared to other TP63-associated sites, reflecting selective hypomethylation of these relative to other sites in the *TP63* gene (Figures 7G and 7H). The cg06520450 site with lowest methylation was most significantly correlated with overall expression of TP63 and miR-944 (Tables S5A and S5B). These findings support a role for differential methylation as well as CN in the preferential expression of ΔNp63 and miR-944 observed in SCC.

DISCUSSION

Here, integrated analyses of genetic, epigenetic, and expression alterations of the PanCan-33 and the Pan-SCC datasets reveal that SCCs from 5 sites have overlapping and distinguishing molecular features that collectively set them apart from other cancers. Several SCC subtypes distinguished by genomic and epigenetic alterations were corroborated by independent analyses, demonstrating overlap with corresponding mRNA and miRNA expression and pathway activation inferred by PARADIGM and RPPA. Although some of these features may occur individually in other cancers, TM and iC multi-omic molecular classification closely overlapped classifications by histopathologic diagnosis, clinical site, and etiology, while identifying molecular alterations underlying these subtypes of biologic and clinical significance.

We uncovered a significant mutually exclusive relationship between gains in 3q or 11q22 affecting the majority of SCCs (Figure S1C; Table S2A). This finding supports these as possible alternative drivers for a recently described mechanism by which 3q genes *ACTLA6* and *ΔNp63* were found to repress squamous differentiation and promote activation of Hippo growth pathway transcriptional factor *YAP1* (Saladi et al., 2017). This inverse relationship in 11q22 and 3q CN gain is independently supported by a reciprocal pattern of YAP1 and p63 protein immunostaining observed previously in HNSC tissue arrays (Ehsanian et al., 2010). In that study, ΔNp63 and AKT inhibition were shown to modulate YAP1. Recent studies indicate that the function or stability of ΔNp63 and YAP1 can be disrupted by natural isothiocynates such as sulforaphane, and by digitoxin, indicating potential as targets for chemoprevention or therapy (Fisher et al., 2017; Huang et al., 2017). We discovered that predominant expression of ΔNp63 isoforms and embedded miR-944 by SCC is correlated

Figure 6. PARADIGM Analysis Revealed Specific Signatures Enriched in Squamous Cell Carcinomas

(A) Consensus clustering of SCC based on top varying PARADIGM inferred pathway levels (IPLs). The heatmap shows scaled PARADIGM IPLs of key regulatory nodes with >15 downstream targets also showing differential inferred activation. Column color annotation shows consensus cluster membership, tumor type, PanCancer-33 cluster membership, and HPV status. Row color annotation on the right side highlights groups of regulatory nodes potentially implicating the same pathway categories or biological processes.

(B–E) Cytoscape plot of pathway features with differential PARADIGM IPLs connected by regulatory interactions through nodes with >15 differential downstream targets. Subnetwork neighborhoods centered around (B) ERK/MAPK1/JUN/FOS, (C) RELA/p50 and STAT Immune related, (D) p63/DNA damage, and (E) proliferation/mitosis. IPL level (red, higher in SCC; blue, lower in SCC) and node shape reflect feature type (circle, genes; diamond, complexes; V, abstract processes; square, protein family or miRNA). Edge color and type represent interaction type (activating, purple arrow; green T, inhibitory). Proteins and selected complexes are labeled, and regulatory nodes with >15 downstream targets are highlighted in bold.

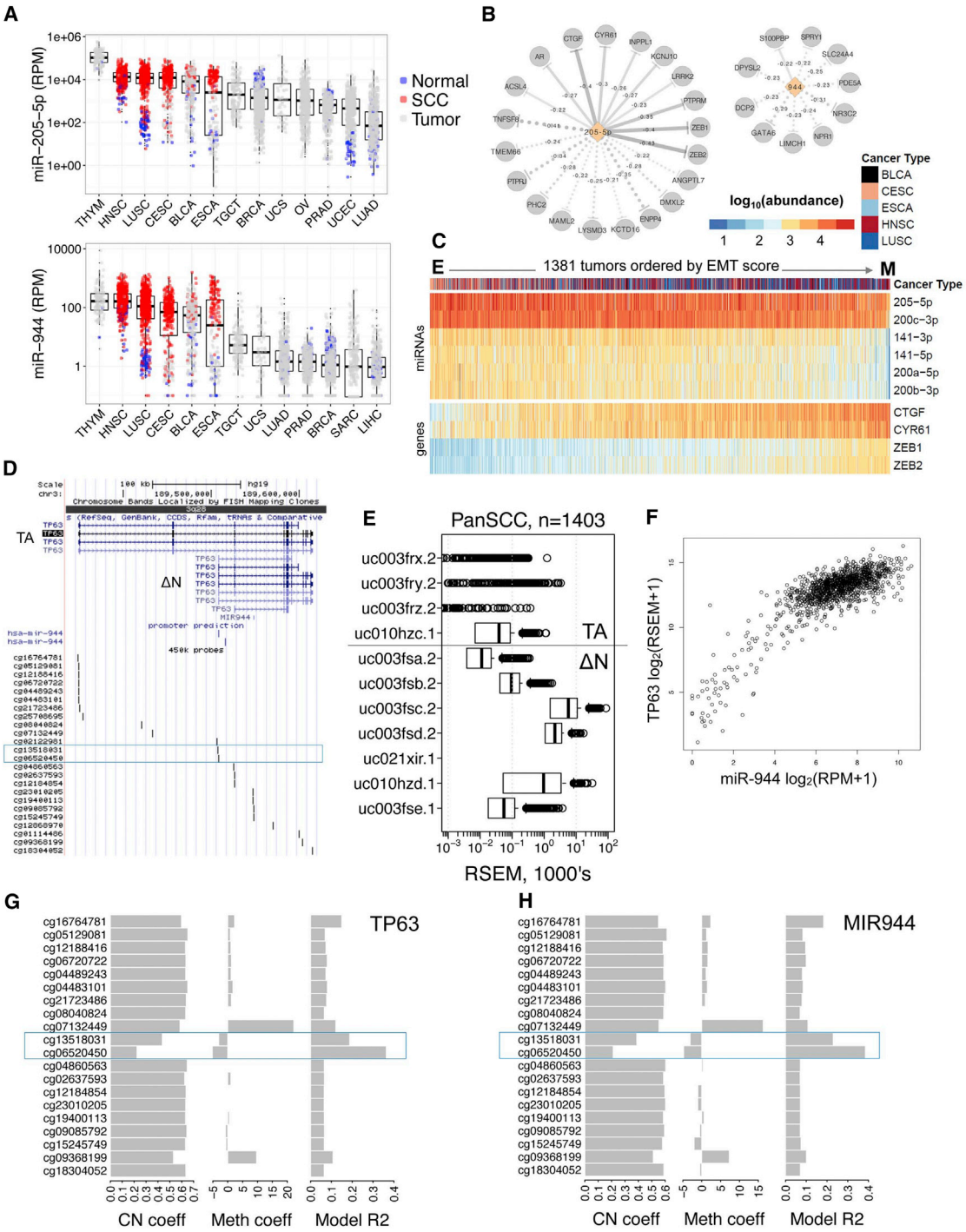


Figure 7. miRNAs Associated with EMT and Hypomethylation and Expression of ΔNp63 Isoforms of TP63 in SCC

(A) Abundance of the most differentially expressed miRNAs miR-205-5p and miR-944 with the highest median expression across the TCGA cancer types (Figure S7). Dots represent Pan-SCC tumors (red), non-squamous TCGA tumors (gray), and normal tissues (blue). Boxplots show median values and the 25th to 75th percentile range in the data, i.e. the interquartile range (IQR). Whisker bars extend 1.5 times the IQR.

(B) Potential gene targets that are significantly anti-correlated to miR-205-5p and miR-944 (Spearman <−0.2, FDR < 0.05) and that have functional validation evidence for direct targeting in miRTarBase v.6.0. Solid versus dotted lines indicate strong versus weaker functional evidence. Numbers on network edges show Spearman correlations between a miRNA and gene.

(C) Heatmap of log10 abundance of miRNAs associated with EMT mRNAs across squamous tumors (n = 1,381). Samples are ordered by the sum of the *Z* scores across the EMT-associated miRNAs.

(legend continued on next page)

with decreased methylation of CpGs at the alternative TSS compared to those of the TSS for the TAp63 isoforms. A correlation between overall TP63 expression and miR-944 due to hypomethylation of the same TSS CpG island is supported by a recent genome-wide analysis (Doecke et al., 2016), but the link with the differential methylation of the alternative TSSs for TA/ΔN isoforms was unrecognized. Repression of TAp63 relative to ΔNp63 was reported to be reversed by 5-Aza-2′-deoxycytidine in BLCA lines (Park et al., 2000). The preferential transcription of ΔNp63 in SCC was also previously reported to be enhanced by PI3K signaling (Barbieri et al., 2003), consistent with the frequent alterations in PI3K-AKT found. These observations suggest that methylation and PI3K inhibitors could modulate TA/ΔNp63 to inhibit SCC.

Indeed, PI3K-AKT-mTOR-eIF signaling appears to be a common pathway in which recurrent 3q26 CNAs (69%; Table S2) and PIK3CA mutations (11%–27%; Figure S2A) are observed. MVisAGe revealed a wider variety of CNAs strongly correlated with the expression of multiple components downstream of PI3K than previously appreciated. Consistent with this, we observed increases in a variety of PI3K, AKT, mTOR, eIF components, and phospho-proteins and greater correlation scores for signaling downstream of mTOR than PI3K detected by RPPA. These observations may help explain the relatively lower sensitivity to PI3K inhibitors of tumors with 3q and other CNAs than those with hotspot mutations of PIK3CA (Mazumdar et al., 2014). SCCs are enriched for P70S6KpT389, RICTORpT1135, and RAD51 DNA damage proteins, which are associated with growth factor and rapamycin sensitivity (Dibble et al., 2009). Recent preclinical studies demonstrate that sensitivity of HNSC lines and xenografts with PIK3CA gains to dual PI3K-mTOR inhibitors and irradiation is correlated with p-AKT and DNA damage responses, supporting investigation of agents targeting PI3K and mTOR in tumors in conjunction with irradiation and pharmacodynamic markers of functional activation (Leiker et al., 2015; Mohan et al., 2015). CNAs or mutations that enhance expression and activation of receptors and kinases activating PI3K-AKT and MAPK signal axes were observed and supported by RPPA. PI3K-mTOR and MEK inhibitors have demonstrated combinatorial inhibitory activity in preclinical studies and in subsets or selected patients in clinical studies (Grilley-Olson et al., 2016; Herzog et al., 2013; Hou et al., 2014; Mohan et al., 2015). Co-activated MAPK-JUN-FOS, RELA/p50, and STAT3 inferred by PARADIGM in major SCC subsets (Figure 6), may be targeted simultaneously by HSP90 inhibitors (Friedman et al., 2013).

HPV(+) and (−) subsets harbored distinct alterations in cell death and survival pathways, which have potential biologic and therapeutic implications. Previously, rare germline genomic alterations in FANC-BRCA pathways have been shown to convey extreme risk for the development of HNSC and GU tract SCCs and susceptibility to HPV infection, but the association with HPV(+) SCC is controversial (Alter et al., 2013). FANC-BRCA defects are associated with increased sensitivity to standard DNA-damaging therapies, potentially helping explain the relative sensitivity of some HPV+ tumors to chemoradiotherapy and potential for their de-escalation. Targeted agents, such as WEE1 inhibitors that prevent G2 checkpoint arrest and DNA repair, may warrant investigation in SCCs with these defects (Aarts et al., 2015) or those with TP53 mutations (Kao et al., 2017). PARADIGM supported increased inferred activity of a network including WEE1, PLK1, AURKA/B, and mTOR linked to SCC displaying the proliferation signature, and activity targeting WEE1 and others is supported by published genome-wide functional RNAi screens and preclinical studies targeting these kinases in HNSC (Hu et al., 2016; Kao et al., 2017). Lastly, the prevalence of 11q13/22 with *FADD*/*IAP* alterations in >30% of HPV(−) HNSC, LUSC, and ESCA subtypes and their sensitivity to IAP inhibitors plus radiotherapy in recent preclinical studies support the investigation of IAP antagonists in those tumors (Eytan et al., 2016).

HPV(+) and (−) subtypes display signatures for LCK, checkpoint PD-L1, Tregs, and MDSCs that overlap protective immune CD4, CD8, and NK responses, possibly helping to explain immune escape of these tumors and limited response rates to immune checkpoint therapies. Small molecules, antibodies, or miRNA mimetics targeting these chemokines or their receptors could be of interest in targeting MDSCs and Tregs in conjunction with checkpoint inhibitors.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Human Subjects
  - Clinical Samples, Data Types, and Genomic Platforms
- METHODS DETAILS
  - iC and TM analysis of PanCan 33
  - CNV analysis and clustering
  - Mutation
  - Global Methylation

(D) Top, Genome view of TAp63, ΔNp63 isoforms, and MIR-944, with PROmiRNA experimentally supported transcriptional start sites (TSSs) for MIR944 (Marsico et al., 2013) that overlap the TSS of alternatively transcribed ΔNp63 isoforms. Bottom, Illumina 450k probes for CpG sites in region of TP63 corresponding to TSSs and coding portion of TAp63, ΔNp63, and MIR944 (blue box).

(E) TP63 isoform mRNA abundance (RSEM) for full transactivating (TA) domain or alternatively transcribed N-terminally truncated (ΔN) isoforms in Pan-SCC tumors (n = 1,403). The ΔN/TAp63 median ratio difference is 212.8-fold. Boxplots show median values and the 25th to 75th percentile range in the data, i.e. the interquartile range (IQR). Whisker bars extend 1.5 times the IQR.

(F) Across Pan-SCC data, miR-944 has largest positive Spearman correlation coefficient for expressed TP63 isoforms.

(G and H) Comparison of coefficients of correlation for copy number (CN), methylation (Meth), and rho-squared (R2) for Illumina 450k probes for CpG sites from (D), with expression of TP63 (G), and MIR944 (H). The blue box corresponds to probes at TSS for ΔNp63 and the TSS for MIR944, which show relatively lower CN and negative Meth coefficients, that most highly correlate with expression of TP63 (for cg06520450 R2 = 0.36, p = 7.4E−106) and MIR944 (cg06520450 R2 = 0.39, p = 1.2E−112).

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and five tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.03.063.

ACKNOWLEDGMENTS

We are grateful to the patients who contributed to this study, and the support of the TCGA Steering Committee and Project Team, especially Samantha Caesar-Johnson and Ina Felau. This work was supported by the following grants from the NIH: U54 HG003273, U54 HG003067, U54 HG003079, U24 CA143799, U24 CA143835, U24 CA143840, U24 CA143843, U24 CA143845, U24 CA143848, U24 CA143858, U24 CA143866, U24 CA143867, U24 CA143882, U24 CA143883, U24 CA144025, and P30 CA016672, and NIDCD intramural project ZIA-DC-000074.

AUTHOR CONTRIBUTIONS

The TCGA Network, Pan-Cancer Atlas, and Pan-Squamous Cell Carcinoma Working Groups contributed collectively to this study. Initial guidance in the project design was provided by the PanCancer and Tissue of Origin leaders J.M.S., K.A.H., and P.W.L. We acknowledge the following TCGA investigators of the Pan Squamous Analysis Working Group, who contributed substantially to the analysis and writing of this manuscript. Project leaders and graphic abstract, C.V.W. and Z.C.; data coordinator, M. Peto; manuscript coordinator, summary of clinical and pathological data, C.R.P.; editor, J.N.W.; TM and iC, C.K.W., E.D., R.S., and J.S.; CNAs, A.M.T., J.S., and A.D.C.; mutations, J.D.C.; CNA-mRNA analysis, V.W. and D.N.H.; DNA methylation, K.B., O.G., H.F., H.S., and M. Prunello; mRNA expression, Y.L., P.A., J.C., and H.C.; miRNA expression, R.B., A.G.R., P.S., H.-S.C., and T.-W.C.; PARADIGM super-pathway, C.Y. and C.B.; Reverse Phase Protein Array analysis, R.A., A.S., L.A.B., R.S.K., C.M.G., A.H., L.D., J.W., W.M., and C.J.C.; HPV calls, R.B., S.S., K.L.M., A.I.O., S.B., and C.S.P.; clinical and pathological characteristics, disease experts, J.S.R., R.Z., H.A.-A., A.J.L., J.H.Z., E.R.F., K.T.T., and A.K.R.; DNA damage repair pathway, C.W.; integrated analysis of p63 and miRNAs, R.B., A.G.R., P.G., P.S., C.C., and L.D.

DECLARATION OF INTERESTS

Michael Seiler, Peter G. Smith, Ping Zhu, Silvia Buonamici, and Lihua Yu are employees of H3 Biomedicine, Inc. Parts of this work are the subject of a patent application, WO2017040526 titled "Splice variants associated with neomorphic sf3b1 mutants." Shouyoung Peng, Anant A. Agrawal, James Palacino, and Teng are employees of H3 Biomedicine, Inc. Andrew D. Cherniack, Ashton C. Berger, and Galen F. Gao receive research support from Bayer Pharmaceuticals. Gordon B. Mills serves on the External Scientific Review Board of Astrazeneca. Anil Sood is on the Scientific Advisory Board for Kiyatec and is a shareholder in BioPath. Jonathan S. Serody receives funding from Merck, Inc. Kyle R. Covington is an employee of Castle Biosciences, Inc. Preethi H. Gunaratne is founder, CSO, and shareholder of NextmiRNA Therapeutics. Christina Yau is a part-time employee/consultant at NantOmics. Franz X. Schaub is an employee and shareholder of SEngine Precision Medicine, Inc. Carla Grandori is an employee, founder, and shareholder of SEngine Precision Medicine, Inc. Robert N. Eisenman is a member of the Scientific Advisory Boards and shareholder of Shenogen Pharma and Kronos Bio. Daniel J. Weisenberger is a consultant for Zymo Research Corporation. Joshua M. Stuart is the founder of Five3 Genomics and shareholder of NantOmics. Marc T. Goodman receives research support from Merck, Inc. Andrew J. Gentles is a consultant for Cibermed. Charles M. Perou is an equity stock holder, consultant, and Board of Directors member of BioClassifier and GeneCentric Diagnostics and is also listed as an inventor on patent applications on the Breast PAM50 and Lung Cancer Subtyping assays. Matthew Meyerson receives research support from Bayer Pharmaceuticals; is an equity holder in, consultant for, and Scientific Advisory Board chair for OrigiMed; and is an inventor of a patent for EGFR mutation diagnosis in lung cancer, licensed to LabCorp. Eduard Porta-Pardo is an inventor of a patent for domainXplorer. Han Liang is a shareholder and scientific advisor of Precision Scientific and Eagle Nebula. Da Yang is an inventor on a pending patent application describing the use of antisense oligonucleotides against specific lncRNA sequence as diagnostic and therapeutic tools. Yonghong Xiao was an employee and shareholder of TESARO, Inc. Bin Feng is an employee and shareholder of TESARO, Inc. Carter Van Waes received research funding for the study of IAP inhibitor ASTX660 through a Cooperative Agreement among NIDCD, NIH, and Astex Pharmaceuticals. Raunaq Malhotra is an employee and shareholder of Seven Bridges, Inc. Peter W. Laird serves on the Scientific Advisory Board for AnchorDx. Joel Tepper is a consultant at EMD Serono. Kenneth Wang serves on the Advisory Board for Boston Scientific, Microtech, and Olympus. Andrea Califano is a founder, shareholder, and advisory board member of DarwinHealth, Inc. and a shareholder and advisory board member of Tempus, Inc. Toni K. Choueiri serves as needed on advisory boards for Bristol-Myers Squibb, Merck, and Roche. Lawrence Kwong receives research support from Array BioPharma. Sharon E. Plon is a member of the Scientific Advisory Board for Baylor Genetics Laboratory. Beth Y. Karlan serves on the Advisory Board of Invitae.

Received: August 1, 2017
Revised: February 26, 2018
Accepted: March 15, 2018
Published: April 3, 2018

REFERENCES

Aarts, M., Bajrami, I., Herrera-Abreu, M.T., Elliott, R., Brough, R., Ashworth, A., Lord, C.J., and Turner, N.C. (2015). Functional Genetic Screen Identifies Increased Sensitivity to WEE1 Inhibition in Cells with Defects in Fanconi Anemia and HR Pathways. Mol. Cancer Ther. *14*, 865–876.

Alter, B.P., Giri, N., Savage, S.A., Quint, W.G., de Koning, M.N., and Schiffman, M. (2013). Squamous cell carcinomas in patients with Fanconi anemia and dyskeratosis congenita: a search for human papillomavirus. Int. J. Cancer *133*, 1513–1515.

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

Barbieri, C.E., Barton, C.E., and Pietenpol, J.A. (2003). Delta Np63 alpha expression is regulated by the phosphoinositide 3-kinase pathway. J. Biol. Chem. *278*, 51408–51414.

Broom, B.M., Ryan, M.C., Brown, R.E., Ikeda, F., Stucky, M., Kane, D.W., Melott, J., Wakefield, C., Casasent, T.D., Akbani, R., and Weinstein, J.N. (2017). A Galaxy Implementation of Next-Generation Clustered Heatmaps for Interactive Exploration of Molecular Profiling Data. Cancer Res. *77*, e23–e26.

Budach, S., Heinig, M., and Marsico, A. (2016). Principles of microRNA Regulation Revealed Through Modeling microRNA Expression Quantitative Trait Loci. Genetics *203*, 1629–1640.

Campbell, J.D., Alexandrov, A., Kim, J., Wala, J., Berger, A.H., Pedamallu, C.S., Shukla, S.A., Guo, G., Brooks, A.N., Murray, B.A., et al.; Cancer Genome Atlas Research Network (2016). Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat. Genet. *48*, 607–616.

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

Cancer Genome Atlas Research Network (2012). Comprehensive genomic characterization of squamous cell lung cancers. Nature *489*, 519–525.

Cancer Genome Atlas Research Network (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. Nature *507*, 315–322.

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; Buck Institute for Research on Aging; Canada's Michael Smith Genome Sciences Centre; Harvard Medical School; Helen F. Graham Cancer Center &Research Institute at Christiana Care Health Services; et al. (2017a). Integrated genomic and molecular characterization of cervical cancer. Nature *543*, 378–384.

Cancer Genome Atlas Research Network; Analysis Working Group: Asan University; BC Cancer Agency; Brigham and Women's Hospital; Broad Institute; Brown University; Case Western Reserve University; Dana-Farber Cancer Institute; Duke University; Greater Poland Cancer Centre; et al. (2017b). Integrated genomic characterization of oesophageal carcinoma. Nature *541*, 169–175.

Ceccaldi, R., Sarangi, P., and D'Andrea, A.D. (2016). The Fanconi anaemia pathway: new players and new functions. Nat. Rev. Mol. Cell Biol. *17*, 337–349.

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., Hackl, H., and Trajanoski, Z. (2017). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. *18*, 248–262.

Chen, F., Zhang, Y., Şenbabaoğlu, Y., Ciriello, G., Yang, L., Reznik, E., Shuch, B., Micevic, G., De Velasco, G., Shinbrot, E., et al. (2016). Multilevel Genomics-Based Taxonomy of Renal Cell Carcinoma. Cell Rep. *14*, 2476–2489.

Chen, J.Y., Luo, C.W., Lai, Y.S., Wu, C.C., and Hung, W.C. (2017). Lysine demethylase KDM2A inhibits TET2 to promote DNA methylation and silencing of tumor suppressor genes in breast cancer. Oncogenesis *6*, e369.

Chou, C.H., Chang, N.W., Shrestha, S., Hsu, S.D., Lin, Y.L., Lee, W.H., Yang, C.D., Hong, H.C., Wei, T.Y., Tu, S.J., et al. (2016). miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res. *44* (*D1*), D239–D247.

Davis, R.J., Van Waes, C., and Allen, C.T. (2016). Overcoming barriers to effective immunotherapy: MDSCs, TAMs, and Tregs as mediators of the immunosuppressive microenvironment in head and neck cancer. Oral Oncol. *58*, 59–70.

Derakhshan, A., Chen, Z., and Van Waes, C. (2017). Therapeutic Small Molecules Target Inhibitor of Apoptosis Proteins in Cancers with Deregulation of Extrinsic and Intrinsic Cell Death Pathways. Clin. Cancer Res. *23*, 1379–1387.

Dibble, C.C., Asara, J.M., and Manning, B.D. (2009). Characterization of Rictor phosphorylation sites reveals direct regulation of mTOR complex 2 by S6K1. Mol. Cell. Biol. *29*, 5657–5670.

Doecke, J.D., Wang, Y., and Baggerly, K. (2016). Co-localized genomic regulation of miRNA and mRNA via DNA methylation affects survival in multiple tumor types. Cancer Genet. *209*, 463–473.

Dotto, G.P., and Rustgi, A.K. (2016). Squamous Cell Cancers: A Unified Perspective on Biology and Genetics. Cancer Cell *29*, 622–637.

Ehsanian, R., Brown, M., Lu, H., Yang, X.P., Pattatheyil, A., Yan, B., Duggal, P., Chuang, R., Doondeea, J., Feller, S., et al. (2010). YAP dysregulation by phosphorylation or ΔNp63-mediated gene repression promotes proliferation, survival and migration in head and neck cancer subsets. Oncogene *29*, 6160–6171.

Eytan, D.F., Snow, G.E., Carlson, S., Derakhshan, A., Saleh, A., Schiltz, S., Cheng, H., Mohan, S., Cornelius, S., Coupar, J.F., 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.

Fisher, M.L., Ciavattone, N., Grun, D., Adhikary, G., and Eckert, R.L. (2017). Sulforaphane reduces YAP/ΔNp63α signaling to reduce cancer stem cell survival and tumor formation. Oncotarget *8*, 73407–73418.

Friedman, J.A., Wise, S.C., Hu, M., Gouveia, C., Vander Broek, R., Freudlsperger, C., Kannabiran, V.R., Arun, P., Mitchell, J.B., Chen, Z., and Van Waes, C. (2013). HSP90 Inhibitor SNX5422/2112 Targets the Dysregulated Signal and Transcription Factor Network and Malignant Phenotype of Head and Neck Squamous Cell Carcinoma. Transl. Oncol. *6*, 429–441.

Gao, Y.B., Chen, Z.L., Li, J.G., Hu, X.D., Shi, X.J., Sun, Z.M., Zhang, F., Zhao, Z.R., Li, Z.T., Liu, Z.Y., et al. (2014). Genetic landscape of esophageal squamous cell carcinoma. Nat. Genet. *46*, 1097–1102.

Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S.V., Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D., et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat. Med. *21*, 938–945.

Gevaert, O. (2015). MethylMix: an R package for identifying DNA methylation-driven genes. Bioinformatics *31*, 1839–1841.

Gonzalez-Angulo, A.M., Hennessy, B.T., Meric-Bernstam, F., Sahin, A., Liu, W., Ju, Z., Carey, M.S., Myhre, S., Speers, C., Deng, L., et al. (2011). Functional proteomics can define prognosis and predict pathologic complete response in patients with breast cancer. Clin. Proteomics *8*, 11.

Grilley-Olson, J.E., Bedard, P.L., Fasolo, A., Cornfeld, M., Cartee, L., Razak, A.R., Stayner, L.A., Wu, Y., Greenwood, R., Singh, R., et al. (2016). A phase Ib dose-escalation study of the MEK inhibitor trametinib in combination with the PI3K/mTOR inhibitor GSK2126458 in patients with advanced solid tumors. Invest. New Drugs *34*, 740–749.

Hajek, M., Sewell, A., Kaech, S., Burtness, B., Yarbrough, W.G., 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.

Haraguchi, K., Ohsugi, M., Abe, Y., Semba, K., Akiyama, T., and Yamamoto, T. (2008). Ajuba negatively regulates the Wnt signaling pathway by promoting GSK-3beta-mediated phosphorylation of beta-catenin. Oncogene *27*, 274–284.

He, Q., Jing, H., Liaw, L., Gower, L., Vary, C., Hua, S., and Yang, X. (2016). Suppression of Spry1 inhibits triple-negative breast cancer malignancy by decreasing EGF/EGFR mediated mesenchymal phenotype. Sci. Rep. *6*, 23216.

Hennessy, B.T., Lu, Y., Poradosu, E., Yu, Q., Yu, S., Hall, H., Carey, M.S., Ravoori, M., Gonzalez-Angulo, A.M., Birch, R., et al. (2007). Pharmacodynamic markers of perifosine efficacy. Clin. Cancer Res. *13*, 7421–7431.

Hennessy, B.T., Lu, Y., Gonzalez-Angulo, A.M., Carey, M.S., Myhre, S., Ju, Z., Davies, M.A., Liu, W., Coombes, K., Meric-Bernstam, F., et al. (2010). A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. Clin. Proteomics *6*, 129–151.

Herzog, A., Bian, Y., Vander Broek, R., Hall, B., Coupar, J., Cheng, H., Sowers, A.L., Cook, J.D., Mitchell, J.B., 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.

Hoadley, K.A., Yau, C., Wolf, D.M., Cherniack, A.D., Tamborero, D., Ng, S., Leiserson, M.D.M., Niu, B., McLellan, M.D., Uzunangelov, V., et al.; Cancer Genome Atlas Research Network (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell *158*, 929–944.

Hou, M.M., Liu, X., Wheler, J., Naing, A., Hong, D., Coleman, R.L., Tsimberidou, A., Janku, F., Zinner, R., Lu, K., et al. (2014). Targeted PI3K/AKT/mTOR therapy for metastatic carcinomas of the cervix: A phase I clinical experience. Oncotarget *5*, 11168–11179.

Hu, J., He, X., Baggerly, K.A., Coombes, K.R., Hennessy, B.T., and Mills, G.B. (2007). Non-parametric quantification of protein lysate arrays. Bioinformatics *23*, 1986–1994.

Hu, L., Liu, J., Li, Z., Wang, C., and Nawshad, A. (2016). Transforming growth factor-β1 activates ΔNp63/c-Myc to promote oral squamous cell carcinoma. Oral Surg. Oral Med. Oral Pathol. Oral Radiol. *122*, 460–482.e4.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huang, H., Zhang, W., Pan, Y., Gao, Y., Deng, L., Li, F., Li, F., Ma, X., Hou, S., Xu, J., et al. (2017). YAP Suppresses Lung Squamous Cell Carcinoma Progression via Deregulation of the DNp63-GPX2 Axis and ROS Accumulation. Cancer Res. *77*, 5769–5781.

Kaisari, S., Sitry-Shevah, D., Miniowitz-Shemtov, S., Teichner, A., and Hershko, A. (2017). Role of CCT chaperonin in the disassembly of mitotic checkpoint complexes. Proc. Natl. Acad. Sci. USA *114*, 956–961.

Kao, M., Green, C., Sidorova, J., and Méndez, E. (2017). Strategies for Targeted Therapy in Head and Neck Squamous Cell Carcinoma Using WEE1 Inhibitor AZD1775. JAMA Otolaryngol. Head Neck Surg. *143*, 631–633.

Kim, B.R., Coyaud, E., Laurent, E.M.N., St-Germain, J., Van de Laar, E., Tsao, M.S., Raught, B., and Moghal, N. (2017). Identification of the SOX2 Interactome by BioID Reveals EP300 as a Mediator of SOX2-dependent Squamous Differentiation and Lung Squamous Cell Carcinoma Growth. Mol. Cell. Proteomics *16*, 1864–1888.

Kostic, A.D., Ojesina, A.I., Pedamallu, C.S., Jung, J., Verhaak, R.G., Getz, G., and Meyerson, M. (2011). PathSeq: software to identify or discover microbes by deep sequencing of human tissue. Nat. Biotechnol. *29*, 393–396.

Lau, L.F. (2016). Cell surface receptors for CCN proteins. J. Cell Commun. Signal. *10*, 121–127.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501.

Leiker, A.J., DeGraff, W., Choudhuri, R., Sowers, A.L., Thetford, A., Cook, J.A., Van Waes, C., and Mitchell, J.B. (2015). Radiation Enhancement of Head and Neck Squamous Cell Carcinoma by the Dual PI3K/mTOR Inhibitor PF-05212384. Clin. Cancer Res. *21*, 2792–2801.

Li, L., Li, C., Mao, H., Du, Z., Chan, W.Y., Murray, P., Luo, B., Chan, A.T., Mok, T.S., Chan, F.K., et al. (2016). Epigenetic inactivation of the CpG demethylase TET1 as a DNA methylation feedback loop in human cancers. Sci. Rep. *6*, 26591.

Liang, J., Shao, S.H., Xu, Z.X., Hennessy, B., Ding, Z., Larrea, M., Kondo, S., Dumont, D.J., Gutterman, J.U., Walker, C.L., et al. (2007). The energy sensing LKB1-AMPK pathway regulates p27(kip1) phosphorylation mediating the decision to enter autophagy or apoptosis. Nat. Cell Biol. *9*, 218–224.

Lund, T.C., Coleman, C., Horvath, E., Sefton, B.M., Jove, R., Medveczky, M.M., and Medveczky, P.G. (1999). The Src-family kinase Lck can induce STAT3 phosphorylation and DNA binding activity. Cell. Signal. *11*, 789–796.

Ma, J., Xing, W., Coffey, G., Dresser, K., Lu, K., Guo, A., Raca, G., Pandey, A., Conley, P., Yu, H., and Wang, Y.L. (2015). Cerdulatinib, a novel dual SYK/JAK kinase inhibitor, has broad anti-tumor activity in both ABC and GCB types of diffuse large B cell lymphoma. Oncotarget *6*, 43881–43896.

Macrae, M., Neve, R.M., Rodriguez-Viciana, P., Haqq, C., Yeh, J., Chen, C., Gray, J.W., and McCormick, F. (2005). A conditional feedback loop regulates Ras activity through EphA2. Cancer Cell *8*, 111–118.

Marsico, A., Huska, M.R., Lasserre, J., Hu, H., Vucicevic, D., Musahl, A., Orom, U., and Vingron, M. (2013). PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. Genome Biol. *14*, R84.

Mazumdar, T., Byers, L.A., Ng, P.K., Mills, G.B., Peng, S., Diao, L., Fan, Y.H., Stemke-Hale, K., Heymach, J.V., Myers, J.N., 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.

McCormick, D.L., Horn, T.L., Johnson, W.D., Peng, X., Lubet, R.A., and Steele, V.E. (2015). Suppression of Rat Oral Carcinogenesis by Agonists of Peroxisome Proliferator Activated Receptor γ. PLoS ONE *10*, e0141849.

Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., 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.

Mohan, S., Vander Broek, R., Shah, S., Eytan, D.F., Pierce, M.L., Carlson, S.G., Coupar, J.F., 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.

Neeley, E.S., Kornblau, S.M., Coombes, K.R., and Baggerly, K.A. (2009). Variable slope normalization of reverse phase protein arrays. Bioinformatics *25*, 1384–1389.

Newton, Y., Novak, A.M., Swatloski, T., McColl, D.C., Chopra, S., Graim, K., Weinstein, A.S., Baertsch, R., Salama, S.R., Ellrott, K., et al. (2017). TumorMap: Exploring the Molecular Similarities of Cancer Samples in an Interactive Portal. Cancer Res. *77*, e111–e114.

Okuyama, R., Ogawa, E., Nagoshi, H., Yabuki, M., Kurihara, A., Terui, T., Aiba, S., Obinata, M., Tagami, H., and Ikawa, S. (2007). p53 homologue, p51/p63, maintains the immaturity of keratinocyte stem cells by inhibiting Notch1 activity. Oncogene *26*, 4478–4488.

Papillon-Cavanagh, S., Lu, C., Gayden, T., Mikael, L.G., Bechet, D., Karamboulas, C., Ailles, L., Karamchandani, J., Marchione, D.M., Garcia, B.A., et al. (2017). Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. Nat. Genet. *49*, 180–185.

Park, B.J., Lee, S.J., Kim, J.I., Lee, S.J., Lee, C.H., Chang, S.G., Park, J.H., and Chi, S.G. (2000). Frequent alteration of p63 expression in human primary bladder carcinomas. Cancer Res. *60*, 3370–3374.

Rajadurai, C.V., Havrylov, S., Coelho, P.P., Ratcliffe, C.D., Zaoui, K., Huang, B.H., Monast, A., Chughtai, N., Sangwan, V., Gertler, F.B., et al. (2016). 5′-Inositol phosphatase SHIP2 recruits Mena to stabilize invadopodia for cancer cell invasion. J. Cell Biol. *214*, 719–734.

Sadelain, M., Papapetrou, E.P., and Bushman, F.D. (2011). Safe harbours for the integration of new DNA in the human genome. Nat. Rev. Cancer *12*, 51–58.

Saladi, S.V., Ross, K., Karaayvaz, M., Tata, P.R., Mou, H., Rajagopal, J., Ramaswamy, S., and Ellisen, L.W. (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.

Saloura, V., Vougiouklakis, T., Zewde, M., Kiyotani, K., Park, J.H., Gao, G., Karrison, T., Lingen, M., Nakamura, Y., and Hamamoto, R. (2016). WHSC1L1 drives cell cycle progression through transcriptional regulation of CDC6 and CDK2 in squamous cell carcinoma of the head and neck. Oncotarget *7*, 42527–42538.

Saloura, V., Vougiouklakis, T., Zewde, M., Deng, X., Kiyotani, K., Park, J.H., Matsuo, Y., Lingen, M., Suzuki, T., Dohmae, N., et al. (2017). WHSC1L1-mediated EGFR mono-methylation enhances the cytoplasmic and nuclear oncogenic activity of EGFR in head and neck cancer. Sci. Rep. *7*, 40664.

Sen, G.L., Boxer, L.D., Webster, D.E., Bussat, R.T., Qu, K., Zarnegar, B.J., Johnston, D., Siprashvili, Z., and Khavari, P.A. (2012). ZNF750 is a p63 target gene that induces KLF4 to drive terminal epidermal differentiation. Dev. Cell *22*, 669–677.

Sen, M., Pollock, N.I., Black, J., DeGrave, K.A., Wheeler, S., Freilino, M.L., Joyce, S., Lui, V.W., Zeng, Y., Chiosea, S.I., and Grandis, J.R. (2015). JAK kinase inhibition abrogates STAT3 activation and head and neck squamous cell carcinoma tumor growth. Neoplasia *17*, 256–264.

Shen, R., Olshen, A.B., and Ladanyi, M. (2009). Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics *25*, 2906–2912.

Simarro, M., Gimenez-Cassina, A., Kedersha, N., Lazaro, J.B., Adelmant, G.O., Marto, J.A., Rhee, K., Tisdale, S., Danial, N., Benarafa, C., et al. (2010). Fast kinase domain-containing protein 3 is a mitochondrial protein essential for cellular respiration. Biochem. Biophys. Res. Commun. *401*, 440–446.

Subramanian, U., Kumar, P., Mani, I., Chen, D., Kessler, I., Periyasamy, R., Raghavaraju, G., and Pandey, K.N. (2016). Retinoic acid and sodium butyrate suppress the cardiac expression of hypertrophic markers and proinflammatory

mediators in Npr1 gene-disrupted haplotype mice. Physiol. Genomics *48*, 477–490.

Thakur, R., and Mishra, D.P. (2016). Matrix reloaded: CCN, tenascin and SIBLING group of matricellular proteins in orchestrating cancer hallmark capabilities. Pharmacol. Ther. *168*, 61–74.

Tibes, R., Qiu, Y., Lu, Y., Hennessy, B., Andreeff, M., Mills, G.B., and Kornblau, S.M. (2006). Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. Mol. Cancer Ther. *5*, 2512–2521.

Vaske, C.J., Benz, S.C., Sanborn, J.Z., Earl, D., Szeto, C., Zhu, J., Haussler, D., and Stuart, J.M. (2010). Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics *26*, i237–i245.

Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

Yeger, H., and Perbal, B. (2016). CCN family of proteins: critical modulators of the tumor cell microenvironment. J. Cell Commun. Signal. *10*, 229–240.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
RPPA antibodies	RPPA Core Facility, MD Anderson Cancer Center	https://www.mdanderson.org/research/research-resources/core-facilities/functional-proteomics-rppa-core/antibody-information-and-protocols.html
Biological Samples		
Raw, processed and clinical data	TCGA Network	https://portal.gdc.cancer.gov/legacy-archive/search/f
PancanAtlas publication page	TCGA Network	https://gdc.cancer.gov/about-data/publications/pancanatlas
Mutation data	TCGA Network	https://gdc.cancer.gov/about-data/publications/mc3-2017
FireBrowse portal	Broad Institute	http://gdac.broadinstitute.org
cBioPortal	Memorial Sloan Kettering Cancer Center	http://www.cbioportal.org
Critical Commercial Assays		
Genome-Wide Human SNP Array 6.0	Affymetrix/ThermoFisher	901182
HumanMethylation450 Platform	Illumina	WG-314-1003
HumanMethylation27 Platform	Illumina	WG-311-2201
Deposited Data		
Raw genomic and clinical data	NCI Genomic Data Commons	https://gdc.cancer.gov
COSMIC Census	PMID: 14993899	http://cancer.sanger.ac.uk/census
Software and Algorithms		
TumorMap	PMID: 29092953	https://tumormap.ucsc.edu/
iCluster	PMID: 19759197	https://bioconductor.org/packages/release/bioc/html/iClusterPlus.html
GISTIC2	PMID: 21527027	http://software.broadinstitute.org/cancer/software/genepattern/modules/docs/GISTIC_2.0
MutSig2CV	PMID: 24390350	N/A
DAVID Bioinformatics	PMID: 19131956	https://david.ncifcrf.gov
MVisAGe software	Comprehensive R Archive Network	https://cran.r-project.org/web/packages/MVisAGe/index.html
MethylMix R Package	PMID: 25609794	https://www.bioconductor.org/packages/release/bioc/html/MethylMix.html
Next-Generation Clustered Heatmaps (NG-CHMs)	PMID:29092932	http://bioinformatics.mdanderson.org/TCGA/NGCHMPortal/
ConsensusClusterPlus R package	PMID: 20427518	https://www.bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html
Next-Generation Clustered Heatmap (NG-CHM)	PMID: 29092932	http://bioinformatics.mdanderson.org/TCGA/NGCHMPortal/
PARADIGM	PMID: 20529912	http://sbenz.github.io/Paradigm/
miRTarBase v6.0	PMID: 26590260	http://mirtarbase.mbc.nctu.edu.tw
SuperCurveGUI R package	PMID: 17599930	http://bioinformatics.mdanderson.org/Software/supercurve/
NumPy Python library	PMID: 27362647	http://www.numpy.org/
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 11,188 patients were analyzed in TCGA with at least on molecular-profiling platform. This study contained both males and females with inclusions of genders dependent on tumor types. There were 5,769 females, 5,282 males and 137 missing information about gender. TCGA's goal was to characterize adult human tumors; therefore, the vast majority are over the age of 18. However, there are 20 samples that are under the age of 18 that had tissue submitted prior to clinical data. Age was missing for 188 patients. The range of ages was 10 – 90 (maxed 90 for protection of human subjects) with a median age of diagnosis of 60 years of age.

Clinical Samples, Data Types, and Genomic Platforms

Details about sample collection, tissue-specific sample selection criteria, clinical annotations, and the genomic data pipelines the PanCan-33 atlas can be found via the TCGA publication page of the Genome Data Commons (https://gdc.cancer.gov/about-data/publications/pancanatlas) and the original TCGA marker paper for each tissue site (Cancer Genome Atlas Network, 2015; Cancer Genome Atlas Research Network, 2012, 2014; Cancer Genome Atlas Research Network et al., 2017a, 2017b). Appropriate consent was obtained for all subjects by the local committee as required by TCGA. Data for molecular features from iC analysis for PanCan 33 tumors are found in Tables S1A–S1L. We conducted a comprehensive study of 1409 mostly untreated primary TCGA tumors with clinical-pathologic diagnosis of SCC or squamous differentiation (Tables S1M and S1N). The tissues of origin included 522 HNSC, 489 LUSC, 95 ESCA, 256 CESC, and 47 BLCA cases with squamous differentiation (Table S1O). Histological diagnoses by pathologists of the originating institution were used and were independently verified with high concordance by specialty pathologists for a majority of samples from each site (Cancer Genome Atlas Network, 2015; Cancer Genome Atlas Research Network, 2012, 2014; Cancer Genome Atlas Research Network et al., 2017a, 2017b). Human papilloma virus (HPV) status and viral subtypes were assessed independently at the Broad Institute by DNA sequencing and PathSeq algorithm (Kostic et al., 2011) and at BC Cancer Center by RNA-seq expression levels (Cancer Genome Atlas Research Network et al., 2017a). The concordance between the two was >99% (Table S1P). Data on DNA copy number, mutations, methylation, mRNA and miRNA sequencing and expression, and protein expression by reverse-phase protein arrays (RPPA) are aggregated in Tables S2A–S2N. TCGA clinical and platform data are available on the GDC website (https://gdc.cancer.gov).

METHODS DETAILS

iC and TM analysis of PanCan 33

The iC clustering algorithm formulates the problem of subgroup discovery as a joint multivariate regression of multiple data types with reference to a set of common latent variables, which represent the underlying 28 tumor subtypes (Shen et al., 2009). The Bayesian information criteria was used to guide the choice of number of clusters. Four datatypes for 9759 samples were used in this analysis as input: mRNA expression, somatic CNA, DNA methylation, and miRNA expression. Data were pre-processed using the following procedures: For mRNA, and mature strand miRNA sequence data, poorly expressed genes were excluded based on median-normalized counts, and variance filtering led to a list of reduced features for clustering. mRNA and miRNA expression features were log2 transformed, normalized and scaled before using as an input to iC. Pre-processing led to 3217 mRNA and 382 miRNA features. Pre-processed methylation data was obtained from the methylation single platform analysis group and included 3139 methylation features. CBS-segmented SCNA data was further reduced to a set of 3105 non-redundant regions. iC was optimized using k-means for 28 major clusters, and visualized by TM. The latent variables were used to generate a TM layout of the samples (Newton et al.,

2017). The TM layout was computed using the Euclidean similarity between each pair of samples in the iC latent space. From the sample similarity matrix, TM uses the DrL layout engine to position the samples in a 2-dimensional map. Samples that were similar in the high-dimensional latent space of the input data to iC are positioned in close proximity to each other in the 2-dimensional space of the map.

Copy number/mRNA expression correlations by MVisAGe

MVisAGe software (https://cran.r-project.org/web/packages/MVisAGe/index.html) was used to compute and visualize gene-level Pearson correlation coefficients computed from quantitative measurements of DNA copy number and gene expression. Briefly, quantitative DNA copy number measurements for the five tumor types were obtained after downloading the GISTIC2 output from the Broad Institute's Firehose GDAC (https://gdac.broadinstitute.org/). Gene expression was quantified using log2(RSEM + 1) values from the same cohort. A total of n = 1370 samples had both DNA copy number and gene expression data, while a total of n = 16,872 genes had measurements with non-zero variance in both datatypes. Gene-level Pearson correlation coefficients were computed (i) across all 1370 subjects, (ii) separately within each tumor type (BLCA (n = 44), CESC (n = 242), ESCA (n = 93), HNSC (n = 508), and LUSC (n = (483)), and (iii) separately within each group defined by HPV status (HPV- (n = 1068) and HPV+ (n = 302)). The smoothing parameters used to create plots of smoothed Pearson correlation coefficients over larger genomic regions were chosen based on manual review.

CNV analysis and clustering

Regions of significant CNA were identified using GISTIC2 (Mermel et al., 2011) on PanCanAtlas SNP6 segmentation files for the entire PanSCC cohort or the non-SCC tumors from the same tissue sites. Copy number values or bins for individual genes and regions were output from the GISTIC2 analysis. It should be noted that the regions identified by the GISTIC2 analysis may be smaller than the region of copy number change in many samples. Therefore, all true driver genes may not always be included in the GISTIC-defined peak region. Similarly, annotation of candidate driver gene(s) in the figures and tables is an interpretation of the most likely candidates in or around that peak. We used hierarchical clustering for copy number (Euclidean distance, wardD2 on R), as hierarchical clustering is more stable for copy number segment data because there are fewer copy number data points to cluster.

Mutation

Mutations were obtained from the MC3 maf file (v0.2.8). Significantly mutated genes were identified for each tumor type and for the combined PanSCC cohort using MutSig2CV (Lawrence et al., 2014), which combines p values from tests for high mutational frequency relative to the background mutation rate (pCV), clustering of mutations within the gene (pCL), and enrichment of mutations within evolutionarily conserved sites (pFN). These p values are combined using the Fisher's method. In order to reduce the number of hypotheses tested in the MutSig2CV analysis, we excluded genes that exhibited low expression across tumors (median < 5 FPKM) as previously described (Campbell et al., 2016). Only the genes with higher expression were considered in the Benjamini-Hochberg correction for multiple hypothesis testing. A one-sided Fisher's exact test was used to determine if the proportion of loss-of-function mutations (including nonsense, frameshift, and *de novo* start out-of-frame mutations) to other mutations for a given gene was significantly higher compared to the proportion of loss-of-function mutations to other mutations across all other genes. To determine if mutational frequency for each gene was associated with CNA cluster status, a Fisher's exact test with 10,000 simulations was used followed by a Bonferroni adjustment for multiple hypothesis testing.

Global Methylation

Two generations of Illumina infinium DNA methylation beadarrays, including HumanMethylation27 (HM27) and HumanMethylation450 (HM450), were used to assay 1,406 pan-squamous tumor samples and 156 normal samples in total. Data from HM27 and HM450 were combined and further normalized by using a probe-by-probe proportional rescaling method to yield a common set of 22,601 probes with comparative methylation levels between two platforms, as described in detail in Syn7073804 on Synapse. Briefly, we rescaled the HM27 data based on between-platform difference measured by technical replicates.

Epigenetic silencing genes

Epigenetic silencing calls at the gene level were processed on probes located within (−1500, +1500) bp of all TSS defined by UCSC database and unmethylated in the normal tissues as well as sorted blood cells, with a median beta value of less than 0.2 required for each normal cell type. In order to get rid of the impact of tissue specificity on gene expression, we z-score transformed gene expression data first into scaled data within each tumor type. The Z scores were derived using the mean and standard deviation calculated with the unmethylated tumors only, defined as those with a beta value of (0, 0.1). Samples across all the tumor types were then pooled together. For each probe/gene pair, we chose the probes that exhibited epigenetic silencing with the following criteria: 1) at least 5 samples were observed with a beta value of 0.3 or above (defined as the methylated group); 2) mean Z score of the methylated group was lower than −2; 3) FDR-corrected p value according to one-side t test on Z scores was lower than 1e-5 between unmethylated and methylated group; 4) the maximum beta value of the methylated group was higher than 0.75. Probes survived this step were retained to call epigenetic silencing events based on DNA methylation profiles for each sample. For genes with only one probe retained, a beta value cutoff of 0.3 was applied to call silencing events, while genes with multiple probes that show evidence of silencing, the beta value cutoff was relaxed to 0.2 with the requirement that greater than half of the probes consistently silenced for that gene.

Due to the presence of multiple transcripts in the CDKN2A region, HM27 did not have a correct probe for the p16 promoter. Therefore, the silencing status for p16 was called with probe cg13601799 on HM450 as previously described (Chen et al., 2016), with a beta value of 0.2 or above considered as epigenetic silencing.

Functional analysis of those epigenetic silencing genes was conducted using DAVID bioinformatics resources (Huang et al., 2009).

MethylMix

Clustering of DNA methylation data

Methylation of neighboring CpG sites tends to be highly correlated. To reduce multiple testing of highly correlated CpG probes, and to reduce the dimensionality of the methylation array data, probes for each gene were clustered using hierarchical clustering with complete linkage. Average methylation values of these CpG clusters, were used as input for MethylMix.

Classification of abnormally methylated genes

MethylMix was applied to CpG cluster data available for 1408 SCC tumors to identify CpG clusters (hereafter referred to as 'genes') that are abnormally methylated in all or a subset of cancers compared with adjacent normal tissue, where this abnormal methylation state is associated with decreased RNA expression of the same gene, as previously described (Gevaert, 2015). To maximize the number of patients for which methylation data was available, both 27k and 450k methylation array data was used. Methylation analysis was therefore restricted to probes represented on both arrays.

We aimed to identify genes that were aberrantly methylated in cancer versus 125 normal adjacent tissue samples available across multiple SCC types, i.e., pan-cancer abnormally methylated genes.

Methylation of many genes differs between tissues under normal (non-cancerous) conditions. It is difficult to distinguish between genes that display normal tissue-specific methylation differences, and those that are abnormally methylated in some cancers but not others. Therefore, we restricted our analysis to genes whose methylation state was consistent across normal adjacent tissues associated with each SCC cancer site, i.e., genes that are 'unimodal' across normal tissues. We applied MethylMix separately to normal tissue data for all SCCs sites, and then to tumor and normal tissue data combined, to identify genes that were unimodal across normal tissues, but abnormally methylated in all or a subgroup of tumors versus normal tissue. MethylMix identifies CpG clusters that are inversely linearly associated with mRNA expression of the corresponding gene, using linear regression, with an r-squared value > 0.1, and a p value < 0.001. These CpG clusters are termed 'functional genes'. Overall, MethylMix identified 905 genes that were unimodal in normal tissue but abnormally methylated in cancer, and where methylation was inversely associated with RNA expression in cancer (Table S2K).

Consensus clustering to identify SCC DNA methylation subgroups

Unsupervised Consensus clustering was applied to DM values data for these 905 genes in all SCC patients, to identify robust methylation clusters (Putative subtypes). Consensus clustering was performed using the ConsensusClusterPlus R package (Wilkerson and Hayes, 2010), using 1000 rounds of k-means clustering, with a maximum of $k = 10$ clusters. We identified five methylation clusters, with selection of the optimal number of clusters based on inspection of plots, dendrograms and features provided by the ConsensusClusterPlus output (Figure 4A). All of the clusters included more than one cancer type that displayed similar methylation patterns (Figure 4A).

DNA methylation profiles of DNA methylation clusters

The SCC clusters differed greatly in their average numbers hypermethylated and hypomethylated genes. Clusters 2 and 4 displayed the highest numbers of hypermethylated genes, while cluster 3 had relatively few hypermethylated genes. Clusters 3 and 5 had the highest number of hypomethylated genes.

Differential distribution of significantly mutated genes between DNA methylation clusters

Of 51 genes that are significantly mutated in SCC overall, 28 were significantly differentially distributed between clusters (Table S2L; Figures 4A and S3B). These abnormally methylated genes included genes that have been causally implicated in cancer development (listed in the COSMIC census http://cancer.sanger.ac.uk/census). Some of these genes were abnormally methylated across multiple SCC types, and therefore represent pan-cancer abnormally methylated genes, while other genes were abnormally methylated within specific subtypes. For example, TET1 and FANCF are specifically hypermethylated in HPV+ subtype 2, while SYK is hypomethylated in HNSC, LUSC, and CESC within all subtypes. For each of the 905 abnormally methylated genes in SCC, MethylMix ascribed differential methylation (DM) values, a categorical variable indicating the methylation state for that gene (normal, hypomethylated or hypermethylated, relative to normal tissue) in each cancer.

mRNA

Consensus hierarchical clustering was performed in R on the PanSCC cohort using 1867 previously defined cancer-related genes, as described in the text. The 1,867 gene compilation we used is based on evidence based compilation of cancer related genes, from major data bases such as the Cancer Atlas (999), Sanger Cancer Gene Census (452), CAN genes (192) and Waldman gene locus (455). These data bases include genes curated from SCC and other sites. The list was compiled by generating a dereplicated gene list from "Table 1: Databases of genes implicated in cancer" in Sadelain et al. (2011). Review of this gene list confirms that it includes many of the most significant and novel cancer related genes and signatures associated with iC, CNAs, mutations, methylation, miRNAs, paradigm analysis, and RPPA analyses found in the present study. Data was visualized with the next-generation clustered heatmap tool mRNA clustering viewed using interactive Next-Generation Clustered Heatmaps (NG-CHMs)(Broom et al., 2017),

(http://bioinformatics.mdanderson.org/TCGA/NGCHMPortal/). A k-means solution was found that discriminated 6 mRNA expression clusters that included mRNAs linked to significant CN, methylation, and miRNA-related alterations found via other platforms (Figures 5A, 5B, and S4).

PARADIGM

The PARADIGM algorithm (Vaske et al., 2010) was used to infer the activities of ~19K pathway features based on expression, copy number and pathway interaction data for 9829 tumor samples, including 1373 squamous cancers. When we compared the PARADIGM integrated pathway levels (IPLs) between squamous and other cancer types. Median-centered IPLs were used to compute the squared Euclidean distance between samples; and this metric was used as the input to the ConsensusClusterPlus algorithm. Hierarchical clustering using the Ward's minimum variance method (i.e., ward inner linkage option) with 80% subsampling was performed over 1000 iterations; and the final consensus matrix was clustered using average linkage. The Ward's minimum variance method is specifically chosen for the hierarchical clustering within each iteration because this method tends to form compact spherical clusters and is less prone to yielding clusters with very few members. The final clustering of the consensus matrix uses average linkage, which is the default option. Applying consensus clustering to subset the 1373 squamous cancers based on ~4000 pathway features with the highest (25%) variance, we observed 6 sub-groups with characteristics patterns of PARADIGM inferred pathway activation pattern. These are shown in Figure 5A. With the exception of the predominantly HNSC cluster C3, and the predominantly LUSC cluster C6, the other clusters are mixed in their tumor type composition. Amplifications, deletions and mutation frequencies between clusters are compared in "one vs. all others" comparisons using the Fisher Exact test with Benjamin-Hochberg False Discovery Rate correction.

miRNA

Unsupervised consensus clustering of miRNA abundance was performed using miRNAs with RPM > 25 in at least 10% of the samples, after assessing k-means clustering metrics, heatmaps and dendrograms, and covariate tracks for clustering results from other platforms. The heatmap displays row-scaled, $\log_{10}$ abundance. MiRNAs with differential expression between PanSCC and all other TCGA samples were identified. Gene-target associations were identified in miRTarBase V6.0 (Chou et al., 2016) and anti-correlations were identified (Spearman < −0.2, FDR < 0.05).

DNA Repair

Fanconi anemia (FA) pathway analysis was performed by the TCGA PanCanAtlas DNA Damage Repair Pathway working group. The FA pathway genes were manually curated. Gene alterations were called based on mutation, methylation, and deep copy number deletions from PanCan 33 dataset for 1409 Pan-SCC, and the top 10 are presented as an oncoprint, and compared with 8350 other cancers. The significance of the difference in FANCF observed in PanSCC versus PanCan 33 data was determined by 2-sample test of the equality of proportions, and two-sided chi-square test p < P < 2.2e-16.

RPPA

RPPA experiments and data processing

Protein was extracted using RPPA lysis buffer (1% Triton X-100, 50 mmol/L HEPES (pH 7.4), 150 mmol/L NaCl, 1.5 mmol/L MgCl2, 1 mmol/L EGTA, 100 mmol/L NaF, 10 mmol/L NaPPi, 10% glycerol, 1 mmol/L phenylmethylsulfonyl fluoride, 1 mmol/L Na3VO4, and aprotinin 10 ug/mL) from human tumors and RPPA was performed as described previously (Hennessy et al., 2007; Hu et al., 2007; Liang et al., 2007; Tibes et al., 2006). Lysis buffer was used to lyse frozen tumors by Precellys' homogenization. Tumor lysates were adjusted to 1 μg/μL concentration as assessed by bicinchoninic acid assay (BCA) and boiled with 1% SDS, and manually serial diluted in two-fold of 5 dilutions with lysis buffer. An Aushon Biosystems 2470 arrayer (Burlington, MA) printed 1,056 samples on nitrocellulose-coated slides (Grace Bio-Labs). Slides were probed with 217 validated primary antibodies followed by corresponding secondary antibodies (Goat anti-Rabbit IgG, Goat anti-Mouse IgG or Rabbit anti-Goat IgG). Signal was captured using a DakoCytomation-catalyzed system and DAB colorimetric reaction. Slides were scanned in a CanoScan 9000F. Spot intensities were analyzed and quantified using Array-Pro Analyzer (Media Cybernetics Washington DC) to generate spot signal intensities (Level 1 data). The software SuperCurveGUI (Hu et al., 2007) available at http://bioinformatics.mdanderson.org/Software/supercurve/, was used to estimate the EC_{50} values of the proteins in each dilution series (in log2 scale). Briefly, a fitted curve ("supercurve") was plotted with the signal intensities on the y axis and the relative log2 concentration of each protein on the x axis using the non-parametric, monotone increasing B-spline model (Tibes et al., 2006). During the process, the raw spot intensity data were adjusted to correct spatial bias before model fitting. A QC metric was returned for each slide to help determine the quality of the slide: if the score was less than 0.8 on a 0-1 scale, the slide was dropped. In most cases, the staining was repeated to obtain a high quality score. If more than one slide was stained for an antibody, the slide with the highest QC score was used for analysis (Level 2 data). Protein measurements were corrected for loading as described (Gonzalez-Angulo et al., 2011; Hu et al., 2007) using median centering across antibodies (level 3 data). Final selection of antibodies was also driven by the availability of high quality antibodies that consistently passed a strict validation process as previously described (Hennessy et al., 2010). These antibodies were assessed for specificity, quantification and sensitivity (dynamic range) in their application for protein extracts from cultured cells or tumor tissue. Antibodies were labeled as *validated* and *use with caution* based on degree of validation by criteria previously described (Hennessy et al., 2010).

Data normalization

Median centering was performed across all the antibodies for each sample to correct for sample loading differences. These differences could arise because protein concentrations are not uniformly distributed per unit volume of lysate due to several factors such as differences in protein concentrations of large and small cells, differences in the amount of proteins per cell, or heterogeneity of the cells comprising the samples. The expression levels across many different proteins in a sample could be used to estimate differences in the total amount of protein in that sample versus other samples. Further, subtracting the median protein expression level forces the median value to become zero, allowing for a comparison of protein expressions across samples. These median-centered data were used for the analysis of all samples. Following this normalization, nine antibodies were removed from the dataset since they showed a large number of incomplete values (over 20%), for a total of 189 antibodies.

Surprisingly, processing similar sets of samples on different slides of the same antibody may result in datasets that have very different means and variances. Neely et al. (Neeley et al., 2009) processed clinically similar ALL samples in two batches and observed differences in their protein data distributions. There were additive and multiplicative effects in the data that could not be accounted by biological or sample loading differences. We observed similar effects in our batches of data as well. A new algorithm, replicates-based normalization (RBN), was therefore developed using replicate samples run across multiple batches to adjust the data for batch effects. The underlying hypothesis is that any observed variation between replicates in different batches is primarily due to linear batch effects plus a component due to random noise. Given a sufficiently large number of replicates, the random noise is expected to cancel out (mean = zero by definition). Remaining differences are treated as systematic batch effects. We can compute those effects for each antibody and subtract them out. Many samples were run in both batches. One batch was arbitrarily designated the ''anchor'' batch and was to remain unchanged. We then computed the means and standard deviations of the common samples in the anchor batch, as well as the other batch. The difference between the means of each antibody in the two batches and the ratio of the standard deviations provided an estimate of the systematic effects between the batches for that antibody (both location-wise and scale-wise). Each data point in the non-anchor batch was adjusted by subtracting the difference in means and multiplying by the inverse ratio of the standard deviations to cancel out those systematic differences. Our normalization procedure significantly reduced technical effects, thereby allowing us to merge the datasets from different batches.

Cluster Analysis

The proteins versus samples data matrix was bi-directionally median centered, then hierarchical clustering was performed with 1-Pearson's correlation coefficient as the dissimilarity measure and Ward's linkage. Selection of cluster number was made after considering k-means clustering metrics, including elbow, Silhouette width, and Gap statistic. Elbow and gap both gave monotonically increasing curves that were noninformative. Silhouette width provided two peaks, one at k = 2 which was too broad, and a smaller one at k = 7 that split the second cluster from the left into two. Based on the dendrograms of the heatmaps, the height of the tree for the second and fourth clusters from the left were almost the same, so there was a limited difference between splitting the dendrogram at k = 7 versus k = 8. Therefore, we selected k = 6 because (i) it is close to the silhouette width result of k = 7, (ii) it produces clusters that aren't too small or too large, (iii) there was a noticeable difference in the dendrogram split point between k = 6 and k = 7, but negligible difference between k = 7 and k = 8. The 6 clusters obtained showed clusters enriched for distinct and shared molecular signatures highlighted in Figure S6A and the Results.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification methods and statistical analysis methods for each of the various data platforms and for integrated analyses are described and referenced in their respective STAR Methods subsections.

DATA AND SOFTWARE AVAILABILITY

The raw data for TCGA PanCanSCC samples, including clinical data, DNA exome sequence, RNA expression sequence, miRNA expression sequence, DNA methylation beta values, SNP Array (copy number data), and RPPA proteomics data are archived and publicly available in the Genomic Data Commons (https://gdc.cancer.gov). Analysis results from other data platforms are provided in the supplemental tables.

Software used for the analyses for each of the data platforms and integrated analyses are described and referenced in the individual Method Details subsections and listed in the Key Resources Table.