Direct reprogramming of oral epithelial progenitor cells to cancer stem cells at single cell resolution *in vivo*

AUTHORS

Farhoud Faraji[1,2,*], Sydney I. Ramirez[3,4], Lauren M. Clubb[5], Kuniaki Sato[2], Paola Y. Anguiano Quiroz[2], William M.G. Galloway[6], Zbigniew Mikulski[4], Thomas S. Hoang[5], Kate Medetgul-Ernar[2], Pauline Marangoni[7], Kyle B. Jones[7], Adam Officer[8], Alfredo A. Molinolo[2], Kenneth Kim[4], Kanako Sakaguchi[9], Joseph A. Califano III[1,2], Quinton Smith[6,10], Ophir D. Klein[7,11], Pablo Tamayo[2,12,13], J. Silvio Gutkind[2,14,*]

AFFILIATIONS

1. Department of Otolaryngology-Head and Neck Surgery, University of California San Diego Health; La Jolla, California, 92037; United States.
2. Gleiberman Head and Neck Cancer Center, Moores Cancer Center, University of California San Diego Health; La Jolla, California, 92037; United States.
3. Division of Infectious Diseases and Global Public Health, Department of Medicine, University of California San Diego Health; La Jolla, California, 92093; United States.
4. La Jolla Institute for Immunology; La Jolla, California, 92037; United States.
5. University of California San Diego, Biomedical Sciences Graduate Program; La Jolla, California, 92093; USA.
6. Department of Chemical and Biomolecular Engineering, University of California Irvine; Irvine, California, 92697; United States.
7. Department of Orofacial Sciences and Program in Craniofacial Biology, University of California San Francisco; San Francisco, California, 94143; United States.
8. University of California San Diego, Bioinformatics and Systems Biology Graduate Program; La Jolla, California, 92093; USA.
9. IDEXX Laboratories KK, Tokyo, 168-0063; Japan.
10. Sue and Bill Gross Stem Cell Research Center, University of California Irvine, Irvine, California, 92697; United States.
11. Department of Pediatrics, Cedars-Sinai Guerin Children's, Los Angeles, California, 90048; United States.
12. Division of Medical Genetics, Department of Medicine, University of California San Diego, California, 92093; United States.
13. Center for Novel Therapeutics, University of California San Diego, La Jolla, California, 92037; United States.
14. Department of Pharmacology, University of California San Diego, School of Medicine, La Jolla, California, 92093; United States.

Lead contact: sgutkind@health.ucsd.edu (J.S.G)
*Correspondence: f1faraji@health.ucsd.edu (F.F.), sgutkind@health.ucsd.edu (J.S.G)

HIGHLIGHTS

- YAP and HPV$^{E6\text{-}E7}$ reprogram oral epithelial progenitor cells into cancer stem cells.
- Single cell analyses reveal the transcriptional architecture of tumor initiation.
- CSC transcriptional programs distinguish normal tissue from carcinoma.
- CSC signatures are associated with poor head and neck cancer survival.

SUMMARY

Tumor initiation represents the initial step in tumorigenesis during which normal progenitor cells undergo cell fate transition to cancer. Most studies investigating cancer-driving mechanisms in solid tumors rely on analyses of established malignant lesions, and thus cannot directly capture processes underlying the reprogramming of normal progenitor cells into cancer cells. Here, using spatiotemporally controlled oncogene expression in a genetically engineered system we demonstrate that concomitant YAP activation and $HPV^{E6\text{-}E7}$-mediated inhibition of tumor suppressive pathways is sufficient to rapidly reprogram oral epithelial progenitor cells (OEPCs) into cancer stem cells (CSCs). Single cell analyses of these nascent CSCs revealed hallmark transcriptional programs driving tumor initiation. Importantly, these CSC-enriched expression signatures distinguish normal tissue from malignant head and neck tumors and are associated with poor patient survival. Elucidating mechanisms underlying OEPC to CSC reprogramming may offer new insights to halt the conversion of premalignant cells into invasive carcinoma.

KEY WORDS

Hippo pathway, YAP, HPV, epithelial stem cells, cancer stem cell, tumor initiating cell, squamous cell carcinoma, oral cancer, HNSCC

INTRODUCTION

Adult stem cells play a central role in tissue homeostasis by balancing self-renewal and differentiation.[1] However, cells with self-renewal capacity may accumulate and propagate oncogenic genomic alterations, ultimately leading to carcinogenesis.[2,3] Stem cells thus possess intrinsic tumor suppressive mechanisms to guard against inappropriate oncogene activation, including terminal differentiation,[4] oncogene-induced senescence,[5] and apoptosis.[6] Current models of carcinogenesis posit that tumor initiation requires the inactivation of intrinsic tumor suppressive mechanisms in addition to oncogene activation.[7] In oral epithelia, stem cells reside in the basal layer of the stratified squamous epithelium and consist of a single pool of self-renewing oral epithelial progenitor cells (OEPCs).[8] This single progenitor cell population renders the oral epithelium an ideal system to elucidate mechanisms underlying the conversion of stem cells into cancer-initiating cells, often referred to as cancer stem cells (CSCs).[9] This represents an advantage over other tissue systems, such as the skin, which are composed of multiple stem cell pools each of which can contribute to cancer initiation in a distinct fashion.[10]

Head and neck squamous cell carcinoma (HNSC) represents the most common malignancy arising from the oral cavity and oropharynx.[11,12] Its recent genomic characterization has revealed that while each individual tumor exhibits alterations in a large number of genes, these genes converge on a finite set of oncogenic molecular pathways.[13–15] Notably, HNSC is characterized by near universal loss-of-function of *TP53* and *CDKN2A* tumor suppressors by either genomic alteration or human papillomavirus (HPV)-mediated inhibition.[13] In addition, gain-of-function mutations in *PIK3CA* and other pathway-specific genes lead to mTOR activation, a widespread feature of HNSC.[16] Of interest, the gene product of *FAT1* is altered in nearly 30% of HNSC.[13] We have recently documented that inactivating mutations in *FAT1* disrupt Hippo pathway signaling, which in turn, results in unrestrained activation of the transcriptional co-activator YAP.[17] These findings have begun to define the landscape of oncogenic pathway alterations in HNSC. However, the minimum complement of pathway alterations sufficient to promote the malignant reprograming of OEPCs into HNSC CSCs remains unclear.

Transcriptome profiling of human HNSC at single cell resolution has identified striking intratumoral cellular heterogeneity and has begun to define candidate transcriptional programs underlying tumor cell phenotypes, including proliferation, invasion, epithelial plasticity, and interactions with the tumor microenvironment.[18–21] Furthermore, single cell analyses have defined recurring cancer cell states shared across diverse solid tumors, suggesting that a common set of transcriptional programs is necessary for the maintenance of malignant cell states.[22] These studies offer important perspectives into the transcriptional landscape of established human malignancies, but offer limited insight into the process of malignant conversion of stem cells as part of tumor initiation programs.

Here, by employing spatiotemporally-controlled oncogene expression in a genetically engineered murine system, we have discovered that, in the context of HPV oncogene-mediated *TP53* and *CDKN2A* suppression, YAP activation is sufficient to rapidly reprogram OEPCs into HNSC CSCs. Histological and transcriptomic analyses reveal that these CSCs display a mitotically active, progenitor-like state that diverges from normal basal layer OEPCs. Single cell analyses of these emerging CSCs defined transcriptional hallmarks of tumor initiation. Importantly, CSC-enriched gene expression programs distinguish normal human tissue from malignant HNSC tissues and are associated with poor patient survival.

RESULTS

YAP activation and HPV$^{E6\text{-}E7}$ expression is sufficient to induce rapid tumor initiation in oral epithelial progenitor cells

TP53, *CDKN2A*, and *FAT1* represent commonly altered genes in HNSC, suggesting that these alterations may underlie squamous carcinogenesis in oral epithelia (**Figure 1A**).[13] The HPV E6 and E7 (E6-E7) oncoproteins induce cell cycle dysregulation by inactivating tumor suppressive pathways mediated by the gene products of *TP53* and *CDKN2A*, respectively (**Figure S1A**).[23–25] Prior observations from our laboratory and others support that E6-E7-mediated cell cycle dysregulation is insufficient to induce oral carcinogenesis *in vivo*[26–28], suggesting that oral epithelial tumor initiation requires alterations beyond *TP53* and *CDKN2A* inactivation.[29–32] We previously demonstrated that *FAT1* inactivation causes Hippo signaling dysregulation leading to unopposed YAP activation (**Figure S1A**).[33] However, the role of YAP activation in tumor initiation remains unclear. We therefore asked if dysregulation of cell cycle progression via E6-E7 combined with constitutive activation of YAP may be sufficient to induce tumor initiation in OEPCs.

To test this hypothesis, we developed a genetically engineered murine system to investigate the effects of local induction of YAP activation by the expression of the *YAP1*^{S127A} allele concomitant with E6-E7-expression in OEPCs. Keratin 14 (KRT14) is expressed almost exclusively in the basal layer of oral epithelia and marks OEPCs, which may represent the cells of origin for oral squamous cell carcinoma.[34,35] We thus employed a tamoxifen-inducible Cre-recombinase (CreERT) driven by the *Krt14* promoter to target genetic manipulations to KRT14^{+} OEPCs.[36] In this system, local administration of tamoxifen activates CreERT-mediated recombination of a floxed STOP cassette (LSL) and enables transcription of the reverse tetracycline-controlled transactivator *(rtTA)* in KRT14^{+} OEPCs (**Figures 1B & S1B-E**). Dietary administration of doxycycline then induces expression of the tetracycline responsive element-regulated (tetON) *HPV16*$^{E6\text{-}E7}$, *YAP1*^{S127A}, or the *H2B-GFP* reporter transgene.[28,37,38]

We bred mice expressing E6-E7 (*Krt14-CreERT/LSL-rtTA/tetON_H2B-GFP/tetON_E6-E7*, "**E**"), *YAP1*^{S127A} (*Krt14-CreERT/LSL-rtTA/tetON_H2B-GFP/tetON_YAP1*^{S127A}, "**Y**"), or both transgenes (*Krt14-CreERT/LSL-rtTA/tetON_H2B-GFP/tetON_E6-E7/tetON_YAP1*^{S127A}, "**EY**"). Littermates that bore neither *tetON_E6-E7* nor *tetON_YAP1*^{S127A} effector transgenes but possessed the *Krt14-CreERT* and *LSL-rtTA* regulatory transgenes were used as negative controls (*Krt14-CreERT/LSL-rtTA/tetON_H2B-GFP*, "**N**"). Intralingual injection of tamoxifen was performed to achieve reliable transgene induction. Mice were started on a doxycycline-containing diet on the first day of tamoxifen treatment (**Figure S1B)**. This treatment regimen resulted in consistent CreERT-mediated excision of the floxed STOP cassette (**Figure S1C**) and expression of effector (**Figure S1D**) and reporter transgenes in KRT14^{+} basal cells (**Figure S1E**).

To determine if E6-E7 expression or YAP activation resulted in epithelial changes, mice were monitored longitudinally by serial examination following local transgene induction. Macroscopic tongue epithelial lesions were noted as early as 8 days after transgene induction. Median lesion-free survival was 110 days in E ($p=0.004$, compared to N) mice, 103 days in Y ($p=0.012$, compared to N), and 16 days in EY ($p<0.0001$, compared to Y, E, or N, **Figures 1C-D**). These macroscopic lesions were readily observed in every EY mouse within 30 days. In fact, by 20 days post-induction, the majority (20/31, 65%) of EY mice bore at least two lesions, while few E or Y and E mice bore any lesions (**Figures S1F**).

Histological analysis by hematoxylin and eosin (H&E) staining of mouse tongues 20 days after transgene induction was performed. Epithelial cell invasion through the basement membrane was confirmed with immunohistochemical (IHC) staining for pan-cytokeratin (**Figure S1G**). At least one invasive HNSC lesion was present in 81% (13/16) of EY mice, while there were few (2/11) in Y and none in E mice (**Figure 1E-F**). The remainder of lesions observed in EY mice consisted of moderate to severe dysplasia, indicating that even mice without epithelial lesions on gross examination displayed dysplastic changes. Most (77%, 10/13) carcinoma-bearing EY mice bore 2 or more independent carcinomas, while the few Y mice with carcinoma each had a single invasive lesion (**Figure 1G**). Carcinoma in EY mice were also larger in size and more deeply invasive (**Figure 1H-I**).

Given the high penetrance of carcinoma in EY, we investigated epithelial histological changes and tumor initiation at higher temporal resolution using a pulse-chase strategy consisting of transgene induction followed by histological analysis of tongue epithelia at 6, 10, and 15 days (**Figure 1J**) post-induction. At day 10, we observed a marked increase in average epithelial thickness of EY compared to N, E, and Y epithelia (**Figure 1K**). Invasive carcinoma occurred in EY by 10 days and was observed in 44% (4/9) of mice, while no carcinoma in Y epithelia was observed until day 20 (**Figure 1L**). Taken together, these findings show that unrestrained YAP activation in KRT14^{+} OEPCs in the context of E6-E7 expression is sufficient to induce oral carcinoma lesions with high penetrance and rapid kinetics.

Concomitant YAP activation and *E6-E7* expression induces the ectopic expansion of a mitotically active stem-like population with invasive potential

Employing the *tetON_H2B-GFP* allele as a reporter for transgene-activated cells, we investigated the early effects of transgene expression that might lead to tumor initiation. In epithelial whole mounts 36 hours after a single dose of intralingual tamoxifen, we observed a similar fraction of H2B-GFP^{+} basal cells (~14-24%) across transgenic conditions, indicating that *tetON_E6-E7* or *tetON_YAP1^{S127A}* expression did not impact the efficiency of transgene induction at an early time point (**Figures S2A-B**). By 6 days post-transgene induction, cross-sectional histology showed marked expansion of H2B-GFP^{+} cells in YAP-expressing epithelia. Remarkably, the full thickness of EY epithelia showed near exclusive occupation by H2B-GFP^{+} cells, raising the possibility that YAP activation combined with E6-E7 expression drives the expansion of a self-renewing progenitor cell compartment (**Figures S2C-D**). Staining for the proliferation marker KI67[39] showed that while isolated mitotically active cells were restricted to the basal layer in control (N) epithelia, the expression of E, Y, or EY resulted in the extension of KI67^{+} cells beyond the basal layer (**Figures 2A, S2E, S2F**). Consistent with the role of E6-E7 in promoting cell cycle progression and hyperplasia, E-expression resulted in dramatic increases in the mean fraction of mitotically active cells (**Figure 2C**).

We next evaluated the expression of epithelial stem cell markers expressed in basal epidermal cells under physiologic conditions. P63 and SOX2 are transcription factors that regulate epithelial morphogenesis and promote stem cell maintenance in the oral epithelial stem cell niche.[40–42] The transmembrane protein ITGA6 is a component of the hemidesmosome that anchors basal epithelial cells to the basement membrane.[43,44] In negative control (N) oral epithelia, expression of P63 and SOX2 was restricted to the basal layer, and ITGA6 localized to membranes of basal cells and to the basement membrane. Expression of E, Y, or EY led to an increase in the abundance of P63^{+} epithelial cells (**Figures 2B & S2G**). However, only Y and EY displayed ectopic extension of P63^{+} cells into suprabasal epithelial strata (**Figures 2B & S2G-H**) and, notably, only EY-expressing cells showed an expansion of the P63^{+} compartment throughout the full thickness of the oral epithelia (**Figure 2B & D**). Similarly, Y expression

resulted in focal extension of SOX2$^+$ cells into suprabasal strata, but the combined EY transgenes drove the expression of SOX2$^+$ cells in all suprabasal and parabasal epithelial layers (**Figures S2I-J**). While E6-E7 expression alone did not significantly alter ITGA6 compartmentalization, Y and EY expression diminished ITGA6 basement membrane localization and resulted in diffuse low (Y) and high (EY) ITGA6 abundance throughout suprabasal strata (**Figures 2B & E**). Importantly, in EY epithelia, nearly every cell at the invasive front was positive for KI67, TP63, and SOX2 (**Figures S2K-M**). Taken together, these findings suggest that YAP expression together with E6-E7-mediated inactivation of tumor suppressor pathways results in the rapid, ectopic expansion of a non-physiologic OEPC-like cell population with invasive potential.

YAP-mediated transcriptional programs induce mTOR activation

We next performed bulk RNA sequencing (RNAseq) of microdissected tongue epithelia to evaluate transcriptional changes occurring upon tumor initiation (**Figure 3A**). We selected 15 days post-transgene induction for transcriptome profiling because half of EY tongues bore invasive carcinoma at this time point. Consistent with our histological observations, principal component analysis showed clear distinctions between EY, Y, and E or N samples. E samples showed more subtle separation from N (**Figures 3B & S3A**).

We next evaluated the expression of 10 canonical YAP target genes[45] to determine YAP activity across transgenic conditions. As expected, YAP target genes were upregulated in YAP-expressing conditions. EY mice had significantly higher expression of the majority of evaluated YAP target genes than Y mice, possibly related to the marked epithelial cell expansion observed upon combined E6-E7 and YAP activation (**Figures 3C & S3B**). To gain a more global view of pathway activity, we performed a single sample gene set enrichment analysis (ssGSEA).[46] As expected, the YAP-expressing conditions (EY and Y) showed enrichment for gene sets indicative of YAP activation, with EY consistently showing greater enrichment than Y (**Figures 3C & S3C, Table S1**). Surprisingly, YAP-expressing conditions also showed enrichment for gene sets indicative of mTOR pathway activation, a commonly activated pathway in oral carcinoma.[47] (**Figures 3D-E & S3C)**. Consistent with ssGSEA results, IHC of Y and EY tongue epithelia demonstrated a pronounced increase in phospho-S6 ribosomal protein abundance, a downstream marker of mTORC1 activity,[48] in basal epithelial cells (**Figures 3F-G**).

These findings raise the possibility that YAP may drive mTOR activity in OEPCs. While a direct link between YAP and mTOR has been elusive, we noted that *Axl* transcription, a YAP-regulated target, is upregulated in EY and Y epithelia (**Figures 3H**). Thus, we hypothesized that *Axl* may represent a mechanistic candidate linking YAP and mTOR activity. We performed siRNA-mediated knockdown of YAP and/or its paralog TAZ (WWTR1) in the human HNSC cell lines Cal27 and Cal33 to test if YAP regulates mTOR activity, as assayed by S6 phosphorylation (**Figures 3I-J**). Consistent with reports that YAP and TAZ serve mutually compensatory functions, knockdown of either YAP or TAZ alone showed minimal effects on expression of the YAP target CYR61 or pS6 abundance.[49] Combined knockdown of YAP and TAZ, however, showed marked depletion of CYR61 and pS6. Intriguingly, siYAP+siTAZ knockdown also resulted in diminished pEGFR but not total EGFR (**Figures 3I-J**), which may be related to YAP-mediated upregulation of the EGFR ligands *Epgn*, *Nrg1*, and *Nrg4* (**Figures S3D**). These findings collectively support a model in which YAP-mediated transcriptional upregulation of *Axl* and EGFR ligands induces an AXL:EGFR-mediated signaling cascade that results in downstream activation of mTOR (**Figure 3K**).

Combined YAP and E6-E7 activation induces oncogenic transcriptional programs promoting cell cycle progression, epithelial cell plasticity, and inflammatory responses

To gain a global view of transcriptional programs modified by transgene expression in OEPCs, differentially expressed genes (DEGs) were identified at thresholds $p_{adj}<0.01$ and $\log_2 FC>1$ by comparing each transgenic condition to control (N). From E to Y to EY, a progressively larger set of DEGs was observed. *E6-E7* expression resulted in 397 DEGs (322 upregulated, 75 downregulated), YAP activation resulted in 1464 (628 upregulated, 836 downregulated) DEGs, and combined YAP activation and *E6-E7* expression resulted in 3736 (1612 upregulated, 2124 downregulated) DEGs (**Table S2**).

Venn analysis was performed to define shared and unique DEGs across EY, Y, and E conditions, revealing 2318 genes differentially expressed solely in EY (termed 'EY unique' DEGs, **Figure 4A, Table S3**). Given that EY represented the only condition in which carcinoma was identified at 15 days post-transgene induction, we hypothesized that EY unique DEGs might illuminate biological processes driving tumor initiation. Gene ontology (GO) analysis identified enrichment among EY unique DEGs for processes underlying cell proliferation (DNA replication), epithelial cell development and identity, and acute phase inflammatory responses (IL-6 production) (**Figure 4B**).[50,51] In line with GO results, upregulated EY unique DEGs showed enrichment in Molecular Signatures Database (MSigDB) Hallmark pathways driving cell proliferation (E2F, G2/M, DNA repair), cell identity (Myc), and inflammatory responses (IFN□, IFNγ, TNFα via NFκB) (**Figure S4A**).[52–54]

Consistent with the possibility of EY activation mediating an acute inflammatory response, *Csf3* encoding the granulocyte cytokine and growth factor G-CSF, represented the most highly upregulated (285-fold) EY unique gene. Further suggestive of granulocyte infiltration into EY epithelia, the granulocyte marker genes *Retnlg*, *Cd177*, and *Ly6g* were identified among the top 5% of the most highly upregulated EY unique genes (**Figure S4B**). Other immune factors represented among the most highly upregulated genes in the EY unique set included the chemokines *Il23a* and *Il17f*. Interestingly, in oral epithelia, TNFα, G-CSF, IL-23, and IL-17 constitute a chemokine feedback mechanism that induces sustained granulocyte recruitment during inflammation.[55] In line with this model, EY epithelial lysates showed increased abundances of TNFα, G-CSF, and IL-17, as well as increases in the granulocyte-specific chemokines CXCL1 and CXCL2 (**Figure S4C**). Collectively, these data suggest that granulocyte infiltration into epithelia constitutes a distinctive feature of concomitant YAP activation and *E6-E7* expression.

The murine oral epithelial stem cell population was recently defined at single cell resolution and consists of stem, cycling progenitor, and differentiating cells in the basal epithelial layer (**Figure 4C**).[8] We used genes enriched in each physiologic cell type defined by this study to construct transcriptional signatures of oral epithelial 'stem cell', 'cycling progenitor', and 'differentiating cell' states. To evaluate the effects of EY expression on components of the oral epithelial stem cell niche, we performed GSEA of these signatures in EY vs N DEGs. We observed pronounced enrichment for the cycling progenitor signature and a dramatic depletion of the differentiation signature, suggesting that EY expression promotes oral epithelial cell cycle progression and inhibits differentiation. Interestingly, the EY transcriptome also displayed depletion of the stem cell program (**Figures 4D-F & S4D**). At the gene level, EY unique DEGs showed upregulation of some OEPC stemness factors (e.g., *Ybx1*, *Procr*, *Ezh2*, *Suz12*, *Eed*) and downregulation of mediators and markers of differentiation (e.g., *Grhl1*, *Ovol1*, *Krt4*, *Krt13*, *Ivl*, *Lor*, *Krtdap*, late cornified envelope genes) and apicobasal polarity (e.g., *Pard3*, *Pard3b*, *Vangl2*, *Camsap3*). Notably, several mediators and markers of the basal progenitor state (e.g., *Hoxc13*, *Pax9*, *Gli2*,

Krt15) were paradoxically downregulated (**Figures S4E**), supporting the possibility that EY-driven transcriptional changes do not reflect a physiologic OEPC cell state, but rather may define a unique cellular state. Indeed, our findings suggest that beyond promoting differentiation arrest, EY expression induces a loss of normal epithelial cell identity and the expression of CSC programs.

To further elucidate this cell state, we employed ssGSEA, and systematically analyzed transcriptomes across all transgenic conditions using transcriptional signatures of human biological processes in MSigDB.[52,54] We identified 24 well-defined gene sets, grouped broadly into 6 categories (cell cycle regulators, cell cycle programs, cell identity regulators, differentiation and development programs, embryonic stem cell (ESC) stemness, and epithelial to mesenchymal transition (EMT) programs), that differentiated N, E, Y, and EY transcriptomes by unsupervised clustering (**Figure 4G**). Drivers of cell cycle progression that were highly active in EY included E2F, AP1, EPC1, and MYC. Cell fate determining factors active in EY included EZH2, YBX1, P63, and SOX2. Notably, cancer cell EMT gene signatures were also enriched in EY, consistent with the possibility that EY expression endows OEPCs with cancer initiating cell transcriptional programs.

To identify potential transcriptional effectors driving gene expression in the EY vs N condition, we performed transcription factor (TF) target enrichment analysis using ChEA3[56] on upregulated (EY_up, n=1612) and downregulated (EY_dn, n=2124) DEGs. We identified 34 unique TFs regulating EY_up transcripts and 16 unique TFs regulating EY_dn transcripts (**Figure S4F & Table S4**). We considered TF expression a potential indicator of TF activity and thus filtered enriched TFs by differential expression in EY vs N. This resulted in 8 candidate TFs regulating EY_up DEGs, including the cancer-associated transcription factors *Foxm1*, *Myc*, *E2f1*, and *Relb* (**Figure S4G**). An additional 8 candidate TFs were identified regulating EY_dn DEGs (**Figure S4H**) including subunits of polycomb repressive complexes PRC1 (*Rnf2*) and PRC2 (*Ezh2*, *Suz12*, *Mtf2*), raising the possibility that EY_dn genes are downregulated via PRC1 and PRC2 epigenetic repression. Protein-protein interaction (PPI) prediction using STRING indicated that EY-regulatory TFs constituted a single physically associating protein network. This PPI network is comprised of 14 of the 16 EY-regulatory TF proteins mediating cell cycle progression, stemness, differentiation, and survival (**Figure S4I**).

Single cell analysis reveals that YAP and E6-E7 induce the emergence of cancer stem cells

To determine the effects of EY-mediated transcriptional changes at single cell resolution, we performed transcriptome profiling of single cells (scRNAseq) at the same time point as bulk RNAseq. Fifteen days after transgene induction, single cell suspensions were generated from tongue epithelia, sorted for viability, and subjected to droplet-based single cell cDNA library preparation and sequencing (**Figure 5A**, n=2 mouse tongue epithelia per group). Global analysis after data filtering identified a total of 24,057 cells. Epithelial cell clusters were identified by expression of epithelial-specific transcripts and lack of expression of non-epithelial cell transcripts. In total, 12,771 epithelial cells were identified across 8 clusters, broadly dividing into clusters reflecting previously defined physiologic cell states of oral epithelial cells[8] (see above, **Figure 4C**) and novel transgene-associated cell states (**Figure 5B**).

We performed gene enrichment analysis as a first step in annotating epithelial cell cluster phenotypes. Quiescent Progenitors ('QP', 2773 cells) constituted the largest cluster and displayed enrichment of basal/progenitor markers (*Krt15, Col17a1, Itgb4, Itga6*), factors required for stem cell maintenance (*Trp63, Sox2, Zfp36, Procr, Snai2*), enrichment of Notch

pathway members (*Hes1, Jag2*), and antiproliferative factors (*Btg2*, *Gas1*). Cycling Progenitors ('CP', 1376 cells) showed enrichment for factors required for G1/S and G2/M transition (*Foxm1, Pcna, Ccna2, Ccnb1, Cdc25b, Cdk1*), mitosis (*Top2a, Prc1, Mki67, Cenpf, Tpx2*), and showed depletion for markers of epithelial cell differentiation (*Krtdap, Krt4, Krt13*). Differentiating 1 ('D1', 1066 cells), Differentiating 2 ('D2', 1674 cells), and Differentiating 3 ('D3', 1422 cells) represented cells on a continuum toward terminal differentiation. Differentiating 1 cells were enriched for markers of lineage commitment (*Krt24*),[57] factors required for transit amplifying cell proliferation (*Foxq1),*[58,59] and factors required for epidermal cell fate determination *(Hoxc13, Pax9*).[60,61] Differentiating 2 to Differentiating 3 cells showed progressively increasing expression of terminal differentiation markers (*Krt4, Krt13, Krtdap, Fabp5*) (**Figures 5C & S5A, Table S5**).

We next evaluated cluster phenotypes using GSEA[62] to compare cluster-defining genes to transcriptional signatures from three studies defining physiologic murine stem cell states.[8,59,63] The QP cluster was enriched for adult tissue stem cell and oral epithelial stem cell signatures, and depleted for oral epithelial cycling and differentiating cell signatures. The CP cluster showed enrichment for the oral epithelial cycling signatures and depletion for the differentiating cell signature. D1 and D2 cells showed enrichment for the transit amplifying cell signature. D1 through D3 clusters displayed a continuum of cell maturation states with progressively increasing magnitude of enrichment for the differentiating cell signature (**Figures 5F & S5D**). Overall, these findings demonstrate that the physiologic cellular identities we assigned to our scRNAseq clusters are consistent with previously defined physiologic cell states observed in oral epithelial cells.

E, Y, and EY expression resulted in the emergence of three new clusters not observed in control (N) epithelia, each corresponding to a transgenic condition (**Figures 5D-E & S5B-C**). The large majority of E-enriched cluster cells (1737/1924 cells, 90.3%) were observed in E epithelia. This E-enriched cluster displayed high expression of *E6-E7* and interferon-stimulated genes known to be upregulated by *E6-E7* (e.g., *Cxcl10, Ifit1, Isg15, Ifi27l2a, Ifitm3*).[64,65] Interestingly, while the E-enriched cluster arose specifically in E-expressing epithelial cells, the *E6-E7* transgene was detected in all E epithelium-derived clusters except for QP cells, suggesting that *E6-E7* expression did not strongly influence differentiation states. QP cluster cells lacking *E6-E7* expression, thus, likely represent progenitor cells in which the E6-E7 transgene was not induced. The Y-enriched cluster represented the smallest epithelial cell cluster (525 cells), of which 98.5% were identified in the Y epithelia, and showed high expression of the *YAP1*^{S127A} transgene and YAP target genes (e.g., *Axl, Amotl2, Lats2, Cyr61*). Consistent with H2B-GFP reporter lineage tracing of transgene-induced cells (**Figures S2C-D**), the tumor initiating (TI) cell cluster (2324 cells) constituted 92.2% of all cells in EY epithelia. These cells expressed both *E6-E7* and *YAP1*^{S127A} transgenes, and displayed high expression of partial epithelial to mesenchymal transition (pEMT) (e.g., *Pdpn, Cnn3, Lgals1, Anxa3, Ndrg1*) and hypoxia (e.g., *Pgk1, Eno1, Ldha, Aldoa, Tpi1*) markers (**Figure 5C**).[18,22]

Reflecting findings observed in bulk transcriptome profiles of EY epithelia, the TI cluster was markedly enriched for YAP, mTORC1, E2F, and MYC-driven transcriptional programs (**Figure 5G & S5E**), suggesting that key oncogenic pathways were activated specifically in the TI cluster cells and raising the possibility that the TI cluster represents the elusive tumor initiating cell population. Indeed, the TI cluster was strongly enriched for transcriptional modules defining pan-cancer cell states[22] including cell cycle, hypoxia, interferon response, squamous differentiation, and pEMT (**Figure 5G & S5E**). Consistent with this possibility, expression of PDPN, a pEMT marker that promotes invadopodia formation, cell migration, and invasion,[66,67] was increased in EY epithelia, and high PDPN expression corresponded with the invasion of EY-expressing through the basement membrane (**Figure S5F**).

To formally test if transgene-induced CSCs possess tumor initiating capacity, we implanted cell suspensions generated from transgene induced epithelia into the tongues of NOD-SCID-gamma (NSG) immune compromised mice (**Figure 5H**). Tumors formed in 100% of mice implanted with EY-derived cells, and a fraction of mice implanted with Y-induced, and no mice implanted with E-induced cell suspensions (**Figure 5I-K**). Tumors derived from EY cells were larger in size than those derived from Y-induced cells (**Figures 5L**). These findings collectively demonstrate that EY-expression endows OEPCs with oncogenic transcriptional phenotypes, and results in OEPC reprogramming into CSCs.

Tumor cell recruitment of MDSCs with the capacity to modify ECM

The basement membrane represents an anatomic barrier that transformed cells must breach to establish invasive carcinoma. Second harmonic generation microscopy[68] identified a dramatic reduction in fibrillar collagen abundance at the invasive front of infiltrative carcinoma lesions compared to the basement membrane underlying intact transgene induced epithelia (**Figures 6A-B**). However, epithelial cells, particularly TI cluster tumor cells, do not express collagenases (**Figure 6C**). Thus, we hypothesized that $CD45^+$ immune cells within the EY tumor microenvironment (**Figure 6D**) might contribute to remodeling the basement membrane extracellular matrix (ECM). scRNAseq of $CD45^+$ cells identified 11,286 immune cells present in tongue epithelial tissue, which distributed across 13 clusters (**Figures 6E & S6A-B, Table S6**). Myeloid derived suppressor cells (MDSCs) represented less than <1% of immune cells in control (N) and E-expressing epithelia. MDSCs comprised 10% of immune cells in Y-expressing epithelia, and 64.6% of cells in EY-expressing epithelia. The MDSCs infiltrating into EY epithelia distributed nearly equally among granulocytic (31%, G-MDSC) and monocytic (33.6%, M-MDSC) lineages (**Figures 6F-G**). Ligand-receptor analysis revealed TI cluster epithelial cells expressed chemokine ligands (*Cxcl1, Cxcl2, S100a8, S100a9, Csf3)* corresponding to receptors (*Cxcr2, Csf3r*) specifically expressed by G-MDSCs (**Figures 6H-I**). These observations were consistent with RNAseq and cytokine data from bulk epithelia, which showed an enrichment of granulocyte-specific transcripts and increased CXCL1, CXCL2, and G-CSF protein abundance in EY epithelia (see above, **Figure S4C**). Aligned with these transcriptomic and cytokine data, we observed increased abundance of $LY6G^+$ cells (G-MDSCs) at the tumor invasive front in EY epithelia (**Figure 6J & S6C**). In contrast to the absence of collagenase expression in EY epithelial cells, G-MDSCs expressed the collagenases *Mmp8* (neutrophil collagenase) and *Mmp9* (92 kDa type IV collagenase) and, accordingly, we detected an elevated abundance of pro-MMP9 protein in EY epithelial lysates (**Figures 6K & S6D**). Taken together, these findings suggest that EY-transformed CSCs secrete chemokines to recruit collagenase-producing G-MDSCs, and that collagenases made by these G-MDSCs degrade basement membrane collagen fibrils, thereby facilitating EY tumor invasion.

EY-cancer stem cell transcriptional modules are enriched in human HNSC and are associated with poor survival

The oncogenic effects of EY transgene expression in OEPCs support that YAP activation represents a driver of tumor initiation. We thus investigated the status of YAP activation in human malignancies by evaluating immunohistochemical staining for YAP in HNSC tissue microarrays using nuclear localization of YAP as a surrogate for YAP activation.[69] Consistent with its physiologic role in stem cell maintenance and self-renewal,[17,70] nuclear YAP was detected primarily in basal cells in normal oral epithelial tissue (**Figure 7A**). HNSC tumor cells with nuclear YAP staining, on the other hand, were distributed throughout the malignant tissues.

Compared to normal tissue, HNSC had a significantly greater fraction of YAP activated cells (**Figures 7B-C**).

We next tested the clinical significance of features unique to our CSC program in human subjects with HNSC from a molecularly and clinically well-defined cohort, The Cancer Genome Atlas HNSC cohort (TCGA-HNSC). Forty-three TCGA-HNSC subjects had transcriptome profiling data by RNAseq available for HNSC malignant tumor and matched normal solid tissue adjacent to the tumor, enabling comparison of pathway activation across malignant tumor (T) and matched normal tissue (N). Using a set of genes directly regulated by YAP ('YAP Direct Targets') as a reference for YAP-mediated transcriptional activation,[71] we observed enrichment of this YAP activation in tumor compared to normal tissue. A transcriptional signature of mTOR signaling was also enriched in tumors (**Figures 7D & S7A**). Consistent with our findings above, we found a strong positive relationship between YAP and mTOR pathway activation across TCGA-HNSC tumor samples (n=520, R^2=0.448, p<0.001, **Figure 7E**). In addition, subjects with above-median YAP or mTOR pathway activation displayed worse overall (**Figures 7F-G**) and disease-free survival (**Figures S7B-C**), suggesting that beyond tumor initiation YAP and mTOR pathway activation contribute to tumor progression and influence prognostic outcomes.

To define transcriptional modules driving underlying tumor initiating capacity, we performed high dimensional weighted gene co-expression network analysis (hdWGCNA) in TI cluster CSCs.[72,73] hdWGCNA identified 12 co-expressed transcriptional modules (**Figure 7H, Table S7**). To determine if these modules were activated in human HNSC, we subjected each module to ssGSEA analysis in matched TCGA-HNSC tumor and normal tissue samples, and demonstrated enrichment of 8 of 12 modules exclusively in HNSC transcriptomes (**Figures 7I & S7D**). These included modules associated with RNA metabolism and processing, intracellular trafficking, hypoxia response, G1/S and G2/M cell cycle progression, interferon response, cell motility and migration, and cytoskeleton and cell polarity. Of interest, enrichment for G1/S cell cycle, motility and migration, and cytoskeleton and polarity modules was associated with worse overall (**Figures 7J-L**) and disease-free survival (**Figures S7E-G**). These findings suggest that oncogenic gene programs enriched in the TI cluster are associated with tumor initiation and worse survival in human HNSC.

DISCUSSION

Tumor initiation represents the initial step in tumorigenesis during which normal progenitor cells undergo cell fate transition to carcinoma. For solid tumors, the overwhelming majority of work investigating cancer-driving mechanisms has relied on analyses of established malignant tumors, and thus cannot capture the intricacies of the normal-to-malignant cell reprograming process. We designed a genetically engineered murine system to explore tumor initiation through spatiotemporally controlled oncogene activation in OEPCs. Using this system, we demonstrate that unrestrained YAP activation, in the context of E6-E7-mediated tumor suppressor inhibition, is sufficient to rapidly reprogram basal layer OEPCs into invasive CSCs. The fast kinetics of carcinoma formation observed upon concomitant *E6-E7* expression and YAP activation minimizes the likelihood that additional oncogenic events are required for tumor initiation. We thus posited that this system would enable interrogation of the complexity of molecular and cellular processes driving tumor initiation as they occur *in vivo*.

Our findings support that YAP represents a *bona fide* oncogene in OEPCs,[33] but low penetrance, prolonged timing, and reduced tumor sizes suggest that YAP activation alone may not be sufficient to bypass OEPC tumor suppressive mechanisms. Of interest, a recent study demonstrated that the conditional deletion of Hippo pathway kinase activators MOB1A and MOB1B drives HNSC in a YAP-dependent manner.[74] However, MOB1A/B also stimulate Hippo-independent oncogenic mechanisms via the NDR1 and NDR2 kinases,[75,76] suggesting that YAP-independent NDR1/2 mediated activity contributes to HNSC development and augments the limited tumorigenic potential of direct YAP activation. Notably, although isolated expression of *E6-E7* in OEPCs induces a remarkable expansion of the oral epithelial proliferative compartment, it results in only modest dysplastic histological changes and transcriptional effects. Indeed, we and others have shown that *E6-E7* expression alone leads to HNSC formation with very prolonged latency,[26,28,77] supporting that E6-E7-driven expansion of proliferative capacity exposes OEPCs to further genetic insults, albeit with limited efficiency. Importantly, we found that *E6-E7*-mediated inhibition of *TP53* and *CDKN2A* tumor suppressive pathways was sufficient to unleash the full oncogenic potential of YAP.

Given the extensive literature highlighting early and widespread mTOR activation in HNSC,[16,47,78,79] the sufficiency of YAP activation and *E6-E7* expression in carcinogenesis appears at first glance to be counterintuitive. Nonetheless, we observed robust mTOR activation in EY epithelia. Specifically, we found the accumulation of the phosphorylated form of S6, a downstream mTOR target[48] and a hallmark of HNSC premalignancy to carcinoma transition,[80,81] in the basal epithelial layer of pre-invasive lesions and throughout invasive carcinoma. This raises the possibility that YAP may activate the mTOR pathway. A prior study suggested that YAP-mediated transcriptional activation of miR-29 can reduce expression of PTEN, a PI3K-mTOR pathway tumor suppressor protein, thereby activating AKT and mTOR in MCF10A breast epithelial cells.[82] However, we did not observe any changes in PTEN expression in EY lesions. Instead, we demonstrate YAP-driven transcriptional upregulation of AXL and EGFR ligands may underlie EY-mediated mTOR activation. In turn, these YAP-mediated effects may circumvent the requirement for mTOR pathway activating mutations in HNSC tumor initiation and reveal a new mechanism of mTOR activation in the absence of PI3K/AKT/mTOR axis genomic alterations.

The stem cell niche in the basal layer of the oral epithelium was recently defined at single cell resolution.[8] It consists of a single pool of self-renewing oral epithelial progenitor cells and post-mitotic cells fated for differentiation. Single cell transcriptomic profiling of oral epithelia in our genetically engineered murine system recapitulated these physiologic cell states and identified

the rapid emergence of new transgene-specific cell states not observed under physiologic conditions. Notably, Consistent with the possibility that TI cells represent the previously elusive nascent cancer stem cells, the relative abundance of TI cells in Y and EY epithelia correlated with differences in tumor outgrowth upon orthotopic transplantation of Y and EY epithelial cells.

These findings provided an opportunity to interrogate the relevance of our CSC-specific transcriptional modules in human HNSC. CSC-specific expression programs were enriched in human carcinoma but not in paired normal tissue from the same anatomic site, suggesting these modules represented acquired phenotypes required for tumor initiation. These included cell cycle progression, hypoxia and interferon response, and elements of pEMT-mediated epithelial plasticity (e.g., motility and migration, focal adhesion, and cell-ECM interaction). Notably, modules driving G1/S-phase transition, motility and migration, and focal adhesion also strongly correlated with worse prognosis. This suggests that rather than the acquisition of gene expression programs through a series of progressively more aggressive cell states in a stepwise fashion, initial malignant reprograming of OEPCs endows CSCs with most cancer hallmarks previously identified in advanced human HNSC tumors. Importantly, these observations highlight these cellular processes as key in tumor initiation and support the importance of eradicating CSCs to achieve durable cancer remission.

Concordant with observations that modification of the extracellular matrix and invasion are defining features of premalignancy-to-cancer transition, EY-induced carcinoma displayed tumor invasive fronts with extensive modification of basement membrane collagen fibrils. Remarkably, EY-induced CSCs appeared to lack tumor cell-intrinsic matrix degrading potential. Our analyses demonstrated that CSCs do not express collagenases necessary to breach the basement membrane.[83] Instead, we found that CSCs are endowed with the ability to recruit infiltrating MDSCs expressing collagenases including G-MDSCs expressing *Mmp8* and *Mmp9*. Mechanistically, this may involve the expression of chemokines, such as CXCL1, CXCL2, calprotectin (S100A8/9) and G-CSF by CSCs which act on their cognate receptors, CXCR2, TLR4, and CSF3R on G-MDSC cells. These observations provide a model in which CSCs initially exploit host immune cell-cross talk to promote a defining step in cancer initiation: penetration of the basement membrane. In turn, CSCs likely acquire the ability to express and secrete matrix remodeling proteases with tumor progression, as revealed by recent analyses of established HNSC tumors.[84]

These findings provide a model for recent paradoxical observations that the mutational landscape of premalignancies is not fundamentally distinct from that of corresponding established carcinoma.[85–88] In this scenario, tissue specific stem cells, such as OEPCs, may accumulate multiple genomic and epigenomic alterations that converge to promote OEPC conversion to a CSC state, but the final acquisition of the ability to invade the surrounding tissue may represent a distinct and separate step that may be heavily dependent on the microenvironment surrounding the niche in which the already genetically primed CSCs reside. While this exciting possibility requires experimental evaluation, we can expect that this emerging two-step process in cancer initiation may open multiple new avenues for early intervention to intercept cancer progression from premalignant lesions. This may include halting the recruitment of innate inflammatory cells initiating extracellular matrix remodeling and breaching the basement membrane, concomitant with disrupting the ability of cancer cells to avoid tumor-immune surveillance.

In summary, we demonstrate that the use of a genetically defined, spatiotemporally controlled system simultaneously activating oncogenic pathways and disabling tumor suppressive mechanisms in the OEPC compartment enables the identification of a rapidly emerging distinct

CSC population, and provides a first glimpse into the process of cancer initiation at single cell resolution. Importantly, elucidating the cellular changes in OEPCs harboring CSC programs offer new insights to prevent the progression of premalignant lesions to invasive carcinoma, and may lead to new prognostic and therapeutic strategies for human HNSC.

Limitations of the study

We performed paired analysis of matched human HNSC tumors and adjacent normal human tissue and demonstrated that human HNSC is enriched for EY-induced CSC modules. While this suggests that EY-CSC modules can detect potential cancer initiating transcriptional programs maintained in established HNSC, it may offer only a glimpse into the programs underlying the conversion of normal OEPC to CSC in human HNSC. Currently, no single cell or bulk transcriptomic datasets exist of human oral premalignant disease or early stage HNSC, which could serve as ideal platforms to more precisely evaluate transcriptional programs driving tumor initiation. In addition to such datasets, human OEPC-derived organoids may serve as a surrogate for future studies interrogating the oncogenic transcriptional processes that occur upon malignant cell fate transition.

ACKNOWLEDGEMENTS

This work was supported by Stand Up To Cancer No. 308268 (FF & JSG); NIDCD T32DC000028 (FF); American Head and Neck Society Pilot Grant No. 1061310 (FF); NIAID T32AI007036 & AP Giannini Postdoctoral Research Fellowship and Leadership Award (SIR); National Defense Sciences and Engineering Graduate (NDSEG) Fellowship Program (LMC); NCI R25CA221779 (PYAQ); Tobacco-Related Disease Research Program Pre-Doctoral Award T32DT4965 (TSH); Howard Hughes Medical Institute Hanna H. Gray Fellowship (QS)**;** State of California Initiative to Advance Precision Medicine Award OPR18112 & GCBSR shared resources at the UCSD Moores Cancer center NCI P30CA023100 (JAC, KME, & PT); NIDCR R01DE030497 & NIDCR R01DE026644 (JSG); NCI R01CA257505, NIDCR U01DE028227, NIDCR R01DE026870 (PT & JSG).

We are thankful for Katarzyna Dobaczewska, Sarah McArdle, and Brett Laffey of the LJI Microscopy & Histology Core for helpful insights and guidance of experimental designs; Cheryl Kim and Semra Sehic of the LJI Flow Cytometry Core for assistance with FACS experiments; Suzie Alarcon and Hannah Dose of the LJI NGS Core for assistance with scRNAseq experiments. The BD FACSAria II at LJI was funded by NIH equipment grant S10RR027366. The NovaSeq6000 at LJI was acquired through the Shared Instrumentation Grant Program S10OD025052. The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. We appreciate the generous contribution of the TCGA-HNSC donors and participating research groups who made this resource available.

AUTHOR CONTRIBUTIONS

Conceptualization: FF, JSG; Methodology: FF; Software: FF, SIR, LMC, Ksato, AO; Validation: FF, SIR, LMC; Formal Analysis: FF, SIR, LMC, PM, K.Sato; Investigation: FF, SIR, LMC, K.Sato, PAQ, WMGG, TSH, KME, AO, AAM, KK, K.Sakaguchi, ZM; Resources: JSG, PT, PM, ODK, QS, JAC; Data Curation: FF, SIR, LMC, K.Sato; Writing, Original Draft: FF, SIR, JSG; Writing, Review & Editing: FF, SIR, JSG, ODK, QS, KSato, PT, ZM; Visualization: FF; Supervision: JSG, PT, QS, ODK, JAC; Project Administration: FF, JSG; Funding Acquisition: FF, JSG.

DECLARATION OF INTERESTS

J.S.G. has received other commercial research support from Kura Oncology, Mavupharma, Dracen, SpringWorks Therapeutics, is a consultant/advisory board member for Oncoceutics Inc., Vividion Therapeutics, Domain Therapeutics, and Pangea Therapeutics, and io9, and founder of Kadima Pharmaceuticals. The remaining authors declare no competing interests.

INCLUSION AND DIVERSITY STATEMENT

We support inclusive, diverse, and equitable conduct of research.

FIGURE LEGENDS

Figure 1. YAP and E6-E7 activation is sufficient to induce rapid tumor initiation in OEPCs

(A) Frequency of the most commonly altered Cancer Gene Census driver genes in HNSC.

(B) Schematic depicting spatiotemporally controlled activation of YAP and HPV oncogenic pathways in oral epithelial progenitor cells. A tissue-specific inducible Cre recombinase (CreERT) expressed in *Krt14*-expressing cells is activated by intralingual injection of tamoxifen (Tam), resulting in recombination of the Lox_p-STOP-Lox_p (LSL) and expression of the *rtTA* transgene. Feeding tamoxifen-treated mice doxycycline chow (Dox) activates transcription of the tetracycline-inducible transgenes $HPV16^{E6\text{-}E7}$, YAP^{S127A}, and *H2B-GFP. H2B-GFP* is used as a reporter of cells in which transgenes have been activated.

(C) Representative images of gross tongue lesions 20 days after transgene induction.

(D) Kaplan-Meier plot showing the kinetics of tongue lesion formation upon transgene induction.

(E) Hematoxylin and eosin (H&E) stained tongue epithelia sections demonstrating epithelial changes 20 days after transgene activation.

(F) Histopathologic evaluation and scoring of mouse tongue epithelia; n indicates number of tongues examined for each transgenic condition. If no histopathologic changes were identified, tongues were classified as normal. Presence of low, moderate, and high-grade dysplastic changes were classified as “dysplasia.” Infiltrative lesions invading beyond the basement membrane were classified as “carcinoma.”

(G) Number of infiltrative carcinomatous lesions per examined tongue. Panels G and H: n_Y=2, n_{EY}=34 carcinomata.

(H) Cross-sectional area of infiltrative carcinomatous lesions.

(I) Depth of carcinoma invasion as measured by the longest plumb line orthogonal to the tangent of the nearest intact basement membrane.

(J) Schematic showing pulse-chase strategy of transgene induction and time points for histological analysis of tongue epithelia.

(K) Longitudinal measurement of tongue epithelial thickness. n=3-7 mice per condition per time point.

(L) Highest grade lesion identified per time point per tongue for the carcinogenic conditions Y and EY.

Panel D was analyzed by Bonferroni-corrected Mantel-Cox log-rank test. Panels F-I, and K were analyzed by ANOVA with Tukey correction for multiple comparisons. Panel L was analyzed by two-sided Fisher’s exact test. For all panels with asterisks denoting significance: **$p<0.01$, ***$p<0.001$, ****$p<0.0001$.

Figure S1. Related to Figure 1.

(A) The HPV^{E6-E7} and YAP^{S127A} transgenic system models signaling pathway perturbations caused by frequently altered genes in HNSC: *TP53*, *CDKN2A*, and *FAT1*. *TP53* and *CDKN2A* inactivation are modeled by HPV^{E6}-mediated proteolytic decay of p53 and HPV^{E7}-mediated release of E2F1 from Rb inhibition, respectively. *FAT1* inactivation results in persistent YAP activation, modeled by expression of the constitutively active synthetic YAP^{S127A} mutant.

(B) Overview of local transgene induction in lingual epithelia and strategies for validation of transgene activation at genomic, transcript, and histologic levels.

(C) Left: Genomic DNA extracted from tongue epithelia 10 days after tamoxifen treatment. Top three gel images show presence of transgene and internal control (ctrl) PCR products. Bottom gel image confirms Lox_p-Stop-Lox_p (LSL) recombination for *LSL-rtTA*. Lane 1 mouse tongue epithelial gDNA from a tamoxifen and doxycycline untreated mouse bearing all transgenes. NTC, no template control. Right: Schematic of oligonucleotide primer design to assay LSL excision.

(D) qRT-PCR quantification of HPV^{E6}, HPV^{E7}, and YAP^{S127A} expression in tongue epithelia 10 days after transgene induction.

(E) Immunofluorescence image of tongue epithelia from a Krt14-$CreER^{+}$LSL-$rtTA^{+}$tetON_H2B-GFP^{+} mouse 6 days after transgene induction showing basal localization of KRT14 and ITGA6, and expression of H2B-GFP.

(F) Total number of lesions per tongue 20 days after transgene induction.

(G) Hematoxylin and eosin (top) and pan-cytokeratin IHC (bottom) stained tongue epithelia demonstrating infiltrative carcinoma in Y and EY mice.

Panels D and F were analyzed by ANOVA with Tukey correction for multiple comparisons. For all panels with asterisks denoting significance: *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$.

Figure 2. YAP and E6-E7 activation induces the ectopic expansion of a mitotically active stem-like population with invasive potential

(A) Representative immunofluorescence images of H2B-GFP, KI67, and KRT15 expression in tongue epithelia 15 days after transgene induction. White dashed line represents the basement membrane.

(B) Representative immunofluorescence images of H2B-GFP, P63, and ITGA6 expression in tongue epithelia 15 days after transgene induction. White dashed line represents the basement membrane.

(C) Related to Panel A. Percent $KI67^{+}GFP^{+}$ nuclei in tongue epithelia.

(D) Related to Panel B. Percent $P63^{+}GFP^{+}$ nuclei in tongue epithelia.

(E) Related to Panel B. Quantitative spatial analysis showing normalized fluorescent intensity of ITGA6 signal from basement membrane to epidermal surface.

Panels C and D were analyzed by ANOVA with Tukey correction for multiple comparisons. For all panels with asterisks denoting significance: *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$.

Figure S2. Related to Figure 2.
(A) H2B-GFP$^+$ nuclei in basal cells in tongue epithelial whole mounts by fluorescent microscopy 36 hours after transgene induction.
(B) Related to Panel A. Percent basal H2B-GFP$^+$ nuclei.
(C) Representative images of axial tongue sections of H2B-GFP$^+$ nuclei by fluorescent microscopy 6 days after transgene induction.
(D) Related to Panel C. Percent H2B-GFP$^+$ nuclei.
(E) Representative images of KI67$^+$ nuclei by IHC 20 days after transgene induction.
(F) Related to Panel E. Percent KI67$^+$ nuclei in the suprabasal epidermal strata.
(G) Representative images of P63$^+$ nuclei by IHC 20 days after transgene induction.
(H) Related to Panel F. Percent P63$^+$ nuclei in the suprabasal epidermal strata.
(I) Representative images of SOX2$^+$ nuclei by IHC 20 days after transgene induction.
(J) Related to Panel G. Percent SOX2$^+$ nuclei in the suprabasal epidermal strata.
(K) Magnified segment from panel (E) showing KI67$^+$ nuclei among cells at the invasive front.
(L) Magnified segment from panel (G) showing P63$^+$ nuclei among cells at the invasive front.
(M) Magnified segment from panel (I) showing SOX2$^+$ nuclei among cells at the invasive front.
Panels B, D, F, H, and J were analyzed by ANOVA with Tukey correction for multiple comparisons. For all panels with asterisks denoting significance: *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$.

Figure 3. YAP-mediated transcriptional programs induce mTOR activation.
(A) Schematic demonstrating strategy for bulk RNAseq of microdissected tongue epithelia at 15 days post-transgene induction.
(B) Two-dimensional principal component analysis of individual N, E, Y, and EY transcriptome profiles.
(C) Bulk RNAseq data showing YAP target gene expression in transgenic epithelia.
(D) Single sample GSEA for gene sets associated with TEAD- and YAP-target gene expression, and mTORC1 activation. For gene set details, please see Table S1.
(E) Left: GSEA enrichment plots for the HALLMARK_MTORC1_SIGNALING gene set for E-, Y-, and EY-specific differentially expressed genes. Right: GSEA normalized enrichment scores and statistics for E-, Y-, and EY-specific differentially expressed genes.
(F) IHC staining and (G) quantification of phospho-S6-positive cells in basal epithelial cells.
(H) YAP-target gene *Axl* expression by RNAseq.
(I-J) Western blot showing effect of siRNA-mediated knockdown of YAP, TAZ, and AXL on YAP target gene abundance and the abundance and phosphorylation of EGFR and S6 ribosomal protein in (L) Cal27 and (M) Cal33 cells.
(K) Proposed mechanism for YAP-mediated transcriptional activation of mTORC1 via AXL.
Panels C, D, G, and H were analyzed by ANOVA with Tukey correction for multiple comparisons. For all panels with asterisks denoting significance: *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$.

Figure S3. Related to Figure 3.

(A) Two-dimensional principal component analysis of individual N and E transcriptome profiles.

(B) Related to Figure 3C. Bulk RNAseq data showing YAP target gene expression in transgenic tongue epithelia.

(C) Related to Figure 3D. Single sample GSEA for gene sets associated with TEAD- and YAP-target gene expression, and mTORC1 activation

(D) Expression of genes in tongue epithelia transcriptomes that may contribute to AXL-EGFR-mTORC1 activation.

Panels B, C, and D were analyzed by ANOVA with Tukey correction for multiple comparisons. For all panels with asterisks denoting significance: *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$.

Figure 4. Combined YAP and E6-E7 activation induces oncogenic transcriptional programs

(A) Venn analysis of shared and unique differentially expressed genes for each transgenic condition versus control (N). 'EY unique' denotes the 2,421 genes uniquely dysregulated in EY.

(B) Enriched cellular processes among EY unique DEGs by Gene Ontology.

(C) Model of the basal layer oral epithelial stem cell niche depicting physiologic cell states including stem cells, cycling progenitor cells, and differentiating cells. Adapted from Jones et al.[8]

(D-F) GSEA analysis enrichment plots of EY vs N mouse tongue epithelium differentially expressed genes for (D) cycling progenitor cell (G1/S, G2/M), (E) differentiating cell, and (F) stem cell oral epithelial cell programs.

(G) Single sample GSEA for all transgenic conditions for gene signatures related to cell cycle, differentiation, stemness, and EMT. For gene set details, please see Table S1.

Each row in panel G was analyzed by ANOVA with Tukey correction for multiple comparisons: *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$.

Figure S4. Related to Figure 4.

(A) MSigDB Hallmark Pathways enriched in upregulated EY-unique DEGs.

(B) Granulocyte-specific markers and cytokines are among the most highly differentially expressed EY-unique DEGs.

(C) G-CSF, IL-17, TNFα, CXCL1, and CXCL2 protein abundance in epithelial lysates from N, E, Y, and EY mice 15 days after transgene induction. *p<0.05, **p<0.01.

(D) Single sample GSEA showing enrichment or depletion of cell states observed in the oral epithelial stem cell niche including Cell Cycle Progression (G1/S, G2/M); Stem Cell; and Differentiation (Diff 1 and Diff 2). These gene signatures were derived from cluster defining genes from Jones et al. For gene set details, please see Table S1.

(E) Relative expression of EY-unique DEGs regulating of oral epithelial cell identity.

(F) Transcription factor (TF) enrichment analysis in EY vs N upregulated (red) and downregulated (blue) DEGs.

(G) Top: Venn diagram showing overlap of enriched TFs among EY vs N upregulated DEGs and differentially expressed transcription factors. Bottom: Relative expression of differentially expressed TFs across samples from all transgenic conditions.

(H) Top: Venn diagram showing overlap of enriched TFs among EY vs N downregulated DEGs and differentially expressed transcription factors. Bottom: Relative expression of differentially expressed TFs across samples from all transgenic conditions.

(I) Network diagram showing protein-protein interactions across enriched, differentially expressed transcription factors by STRINGdb v11.5 at minimum required interaction score 0.400. Colored concentric circles (nodes) surrounding transcription factors indicate functional annotations. Edges (lines) indicate that the directly linked proteins are part of the same physical complex. Edge thickness indicates the strength of data supporting the physical interactions.

For all DEG analyses: thresholds of Log2FC>1 and p_{adj}<0.01.

Panels C and D were analyzed by ANOVA with Tukey correction for multiple comparisons: *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

Figure 5. YAP and E6-E7 induce the emergence of cancer stem cells

(A) Experimental approach for single cell transcriptome profiling of tongue epithelia 15 days after transgene induction.

(B) UMAP of epithelial cell states (clusters) identified by scRNAseq classified into physiologic and transgene-associated cell states, n=12,771 epithelial cells.

(C) Relative expression of representative cell state associated genes by cluster.

(D) Individual UMAPs by transgenic condition demonstrating contributions to overall epithelial cell UMAP shown in B. n=2 mouse tongue epithelia per transgenic condition.

(E) Distribution of cell states in tongue epithelial cells from each transgenic condition. n = total cells per transgenic condition.

GSEA of (F) physiologic cell states across epithelial cell clusters and (G) molecular signatures and recurring cancer cell states across the transgene-enriched clusters. Dot color indicates enrichment (red in F, purple in G) or depletion (blue in F, green in G). Dot color indicates enrichment or depletion. Dot size encodes the absolute value (abs) of the normalized enrichment score (NES). Dot opacity represents $-\log_{10}$ of the adjusted p-value (p_{adj}); dots hollow if $p_{adj}>0.05$. For gene set details, please see Table S1.

(H) Experimental approach for orthotopic implantation of transgene-induced epithelial cell suspensions.

(I) Implanted tumor outgrowth frequency. n denotes number of orthotopically implanted mouse tongues. Bonferroni-corrected Fisher's exact test *p<0.05.

(J) Tumor volumes 10 days after implantation. *p<0.05.

(K) Representative photographs of mouse tongues 10 days after implantation.

(L) H&E stained sections of a representative tongue with an implanted EY cell-derived tumor 10 days after EY cell implantation. ANOVA with Tukey correction for multiple comparisons: *p<0.05.

Figure S5. Related to Figure 5.

(A) Expression of the top 6 cluster defining genes among epithelial cell clusters. Dot size denotes the percentage of cells within a cluster expressing each transcript and the color indicates average gene expression across all cells in each cluster.

(B) *E6-E7* transgene expression stratified by genotype.

(C) *YAP1*^{S127A} transgene expression stratified by genotype.

(D) Related to Figure 5F. GSEA enrichment plots for the Jones Stem Cell, Jones Differentiation, and Jones G1/S G2/M physiologic OEPC genes sets across epithelial cell clusters.

(E) Average expression of the YAP Direct Target, mTORC1 Signaling, E2F Targets, MYC Targets, and the Barkley *et al* recurring cancer cell state gene sets: partial epithelial to mesenchymal transition (pEMT), hypoxia, and squamous differentiation in single epithelial cells.

(F) Representative images of H2B-GFP and PDPN expression in N and EY epithelia 10 days after transgene induction. Dashed white line indicates basement membrane.

Figure 6. Tumor cell recruitment of MDSCs with the capacity to modify ECM

(A) Representative H&E-stained sections (left) and collagen imaging by second harmonic generation (right) in control epithelia (N) and an infiltrative carcinoma lesion (EY) 20 days after transgene induction.

(B) Fluorescent intensity of second harmonic generation signal at intact basement membrane of transgenic tongue epithelia and at the invasive front of EY carcinoma 20 days after transgene induction. ANOVA with Tukey correction for multiple comparisons: ***$p<0.001$, ****$p<0.0001$.

(C) Collagenase gene expression across epithelial cell states by scRNAseq.

(D) Percent of $CD45^+$ cells present in tongue epithelia at 1, 6, 10, and 15 days after transgene induction by flow cytometry. Each time point was analyzed individually by ANOVA with Tukey correction for multiple comparisons: **$p<0.01$, ****$p<0.0001$.

(E) UMAP of immune cell clusters identified by scRNAseq, n=11,286 immune cells.

(F) Immune cell types stratified by transgenic condition.

(G) Distribution of immune cell types in each transgenic condition.

(H) Expression of chemokine genes (ligands) among epithelial cell clusters.

(I) Expression of receptor genes among immune cell clusters for corresponding ligands in Panel H.

(J) Representative immunofluorescence image of pan-Cytokeratin, LY6G, and IBA1 expression in N, E, Y, and EY tongue epithelia 20 days after transgene induction.

(K) MMP gene expression across immune cell types.

Figure S6. Related to Figure 6.

(A) Expression of select cell type defining genes among immune cells by scRNAseq.

(B) Dot Plot demonstrating GSEA of published cell-type defining gene lists across immune cell types. Dot color indicates enrichment (Orange) or depletion (Blue). Dot size represents the absolute value of the normalized enrichment score (NES). Dot opacity represents $-\log_{10}$ of the adjusted p-value. For gene set details, please see Table S1.

(C) Related to Figure 6K. Single channel and merged immunofluorescence images of pan-Cytokeratin, LY6G, and IBA1 expression in N, E, Y, and EY epithelia 20 days after transgene induction.

(D) MMP8 and pro-MMP9 protein abundance in epithelial lysates from N, E, Y, and EY mice 15 days after transgene induction. ANOVA with Tukey correction for multiple comparisons: $^{*}p<0.05$.

Figure 7. EY-cancer stem cell transcriptional modules are enriched in human HNSC and are associated with poor survival

(A) Representative images YAP stained tissues by IHC in normal human oral epithelial tissue (top) and malignant human HNSC tissue (bottom).

(B) Percent of cells with nuclear YAP staining by IHC in normal (n=12) and malignant HNSC (n=44) human tissues from tissue microarrays. Two-tailed Mann-Whitney test: ****$p<0.0001$.

(C) Percent of samples with binned nuclear YAP IHC staining in normal (n=12) and malignant HNSC (n=44) human tissues from the same tissue microarray samples as Panel B.

(D) mTORC1 and YAP pathway enrichment in malignant tumors (T) compared to matched normal solid tissues (N) by single sample GSEA among subjects in the TCGA-HNSC cohort (n=43 subjects). ****$p<0.0001$. Two-tailed Mann-Whitney test: ****$p<0.0001$.

(E) Correlation of YAP and mTORC1 pathway activity by ssGSEA among TCGA-HNSC samples (n=520).

(F-G) Kaplan-Meier plots demonstrating worse overall survival (n=520) among TCGA-HNSC subjects with greater median (F) mTORC1 and (G) YAP pathway activity. For gene set details, please see Table S1.

(H) Left: Weighted gene co-expression network analysis (WGCNA) heatmap displaying topological overlap matrix (TOM) dissimilarity indices among genes in TP cluster cells. Right: Table of WGCNA modules and selected genes identified by Metascape and Enrichr. See also Table S7.

(I) Module enrichment in malignant tumors (T) compared to matched normal solid tissues (N) by single sample GSEA among subjects in The Cancer Genome Atlas (TCGA) Head and Neck Squamous Carcinoma (HNSC) cohort (n=43 subjects with matched T and N samples). Two-tailed paired T-test: **$p<0.01$, ****$p<0.0001$.

(J-L) Kaplan-Meier plots demonstrating worse overall survival (n=520) among TCGA-HNSC subjects with greater median enrichment for the (J) G1/S, (K) pEMT, and (L) Focal Adhesion modules.

Figure S7. Related to Figure 7.

(A) Related to Figure 7D. mTORC1 and YAP pathway enrichment in malignant tumors (T) compared to matched normal solid tissues (N) by single sample GSEA among subjects in the TCGA-HNSC cohort (n=43 subjects). Two-tailed paired T-test: ****$p<0.0001$.

(B-C) Kaplan-Meier plots demonstrating worse disease-free survival (n=393) among TCGA-HNSC subjects with greater median (A) mTORC1 and (B) YAP pathway activity.

(D) Related to Figure 7I. EY-module enrichment in malignant tumors (T) compared to matched normal solid tissues (N) by single sample GSEA among subjects in The Cancer Genome Atlas (TCGA) Head and (E-G) Neck Squamous Carcinoma (HNSC) cohort (n=43 subjects with matched T and N samples). Two-tailed paired T-test: **$p<0.01$, ****$p<0.0001$.

(E-G) Kaplan-Meier plots demonstrating worse disease-free survival (n=393) among TCGA-HNSC subjects with greater median enrichment for the (E) G1/S, (F) pEMT, and (G) Focal Adhesion EY-modules.

STAR METHODS

KEY RESOURCES TABLE

Genotyping PCR primers

Oligo name	Sequence	Description	Amplicon Size
K14CreERT_F	CGCATCCCTTTCCAATTTAC		169bp
K14CreERT_R	GGGTCCATGGTGATACAAGG		
Col1_wt_F	TCCCTCACTTCTCATCCAGATATT		
Col1_wt_R	AGTCTTGGATACTCCGTGACCATA	wt:	1092 bp
Col_mut_R	GGACAGGATAAGTATGACATCATCAA	mutant:	480 bp (*YAP1*^{S127A})
Trp53_F	CTTGGAGACATAGCCACACTG	Internal Control	
Trp53_R	TTACACATCCAGCCTCTGTGG		166 bp
Il2_F	CTAGGCCACAGAATTGAAAGATCT	Internal Control	324 bp
Il2_R	GTAGGTGGAAATTCTAGCATCATCC		
HPV16-E67_F2	CTGAGAACAGATGGGGCACA		201 bp
HPV16-E67_R2	GACAGCTCAGAGGAGGAGGA		
rtTA_F	TGCCAACAAGGTTTTTCACTAGAGA		90 bp
rtTA_-R	CTCTTGATCTTCCAATACGCAACCTA		
H2B-GFP-F	GCTCGTTTAGTGAACCGTCAG		250 bp
H2B-GFP-R	GACTGTGTCTGATTTCC		
Ctrl_F	AGTGGCCTCTTCCAGAAATGTGC	Internal Control	521 bp
Ctrl_R	TCTTCTGCGCCTTAGTCACC		

Loxp–STOP–Loxp excision assay PCR primers

Oligo name	Sequence	Amplicon Size	Description
LSL_link_pA_F1	GCCTGAAGAACGAGATCAGC	1.5 kb	Intact
LSL_link_recomb_F1	CAAACTCTTCGCGGTCTTTC	400 bp	Recombined
LSL_link_rtTA_R2	AAAATCTTGCCAGCTTTCCCC		

RT-qPCR primers

Oligo name	Sequence	Amplicon Size	Description
hsaYAP1_qPCR_F1	TAGCCCTGCGTAGCCAGTTA	177 bp	Transgene specific Tang et al.[94]
hsaYAP1_qPCR_R1	TCATGCTTAGTCCACTGTCTGT		
HPV16_E6_qPCR_F1	AATGTTTCAGGACCCACAGG	107 bp	Bordigoni et al.[95]
HPV16_E6_qPCR_R1	GTTGCTTGCAGTACACACATTC		
HPV16_E7_qPCR_F1	TCAGAGGAGGAGGATGAAATAGA	111 bp	
HPV16_E7_qPCR_R1	GCACAACCGAAGCGTAGA		
mmuPpia_qPCR_F1	GAGCTGTTTGCAGACAAAGTTC	125 bp	PrimerBank validated[96]
mmuPpia_qPCR_R1	CCCTGGCACATGAATCCTGG		

IHC antibodies

Primary Antibody	Clone	Dilution	Species	Supplier
Pan-cytokeratin	ab9377	1/200	Rabbit	Abcam
phospho-S6	CST2211	1/400	Rabbit	Cell Signaling Technology
KI67	ab15580	1/400	Rabbit	Abcam
P63	CST39692	1/900	Rabbit	Cell Signaling Technology
SOX2	CST14962	1/300	Rabbit	Cell Signaling Technology

Secondary Antibody	Dilution	Catalog #	Supplier

Goat Anti-Rabbit IgG Antibody (H+L), Biotinylated	1/200	BA-1000	Vector Laboratories

Immunofluorescence antibodies

Primary Antibody	Clone	Dilution	Species	Supplier	Secondary Antibody	Secondary dilution	Secondary catalog #
KRT14	poly19053	1/200	Rabbit	BioLegend	AF568 goat anti-rabbit	1/1000	Thermo A11036
PDPN-biotin	8.1.1	1/100	Syrian Hamster	BioLegend	AF647 streptavidin	1/1000	Thermo A78962
KRT15	Poly18339	1/100	Chicken	BioLegend	AF674 goat anti-chicken	1/1000	Thermo A11036
KI67	ab15580 (poly)	1/200	Rabbit	Abcam	AF568 goat anti-rabbit	1/1000	Thermo A32933
ITGA6	GoH3	1/200	Rat	BioLegend	AF647 goat-anti-rat	1/1000	Thermo A21247
P63	D9L7L	1/200	Rabbit	CST	AF568 goat anti-rabbit	1/1000	Thermo A32933
IBA1	E4O4W	1/200	Rab	CST	AF568 goat anti-rabbit	1/1000	Thermo A32933
LY6G	1A8	1/100	Rat	BioLegend	AF647 goat-anti-rat	1/1000	Thermo A21247
Broad Spectrum Cytokeratin	poly ab86734	1/200	Mouse	Abcam	AF488 goat-anti-mouse	1/1000	Thermo A21121

siRNAs

Name	Description	Supplier	Catalog number	Sequence
siCTRL	Non-targeting Pool	Horizon Discovery	D-001810-10-20	UGGUUUACAUGUCGACUAA
				UGGUUUACAUGUUGUGUGA
				UGGUUUACAUGUUUUCUGA
				UGGUUUACAUGUUUUCCUA
siYAP	Human YAP1 Pool	Horizon Discovery	L-012200-00-0005	GCACCUAUCACUCUCGAGA
				UGAGAACAAUGACGACCAA
				GGUCAGAGAUACUUCUUAA
				CCACCAAGCUAGAUAAAGA
siTAZ	Human WWTR1 Pool	Horizon Discovery	L-016083-00-0005	CCGCAGGGCUCAUGAGUAU
				GGACAAACACCCAUGAACA
				AGGAACAAACGUUGACUUA
				CCAAAUCUCGUGAUGAAUC
siAXL	Human AXL Pool	Horizon Discovery	L-003104-00-0005	ACAGCGAGAUUUAUGACUA
				GGUACCGGCUGGCGUAUCA
				GACGAAAUCCUCUAUGUCA
				GAAGGAGACCCUUAUGGA

Immunoblot antibodies

Primary Antibody	Clone	Dilution	Species	Supplier
YAP/TAZ	D24E4	1/1000	Rabbit	Cell Signaling Technology
AXL	C89E7	1/1000	Rabbit	Cell Signaling Technology
CYR61	D4H5D	1/1000	Rabbit	Cell Signaling Technology
pEGFR	1H123	1/1000	Mouse	Cell Signaling Technology
EGFR	D38B1	1/1000	Rabbit	Cell Signaling Technology
pS6	D68F8	1/1000	Rabbit	Cell Signaling Technology

S6	54D2	1/1000	Mouse	Cell Signaling Technology

Secondary Antibody	Dilution	Catalog #	Supplier
HRP-goat anti-rabbit IgG	1/10,000	4030-05	Southern Biotechnology
HRP-goat anti-mouse IgG	1/10,000	1030-05	Southern Biotechnology

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse lines

Mice were house in accordance with University of California San Diego Institutional Animal Care and Use Committee (IACUC) guidelines. The UCSD IACUC approved all mouse experiments (Protocol S15195). The following mouse lines were kindly provided by Dr. Elaine Fuchs (The Rockefeller University): Tg(KRT14-cre/ERT)^{20Efu} and Tg(tetO-HIST1H2BJ/GFP)^{47Efu}.[36,37] The Col1a1tm1(tetO-Yap1*)^{Lrsn} mouse was kindly provided by Dr. Fernando Camargo (Harvard University).[38] The B6.Cg-Gt(ROSA)26Sor$^{tm1(rtTA,EGFP)Nagy/J}$ mouse was obtained from The Jackson Laboratory.[89] The Tg(tetO-HPV16-E6E7)^{SGu} mouse was designed by the Gutkind laboratory and generated in house.[28] All transgenic mouse experiments were performed in age- and sex-balanced groups of 8-16 week old littermates. NSG™ mice (NOD.Cg-*Prkdc*^{scid} *Il2rg*$^{tm1Wjl/SzJ}$) mice were originally obtained from The Jackson Laboratory and propagated at the Moores Cancer Center. Implantation of transgene epithelial cell suspensions were performed in 8-week-old female NSG mice.

Human HNSC cell lines

CAL27 and CAL33 cell lines were obtained from the NIDCR Oral and Pharyngeal Cancer Branch cell collection.[90] Cell identity was confirmed by STR profiling. CAL27 (CVCL_1107) was derived from a 56 year old male with tongue adenosquamous carcinoma. CAL33 (CVCL_1108) was derived from a 69 year old male with tongue squamous cell carcinoma. Both cell lines were cultured in DMEM (D-6429, Sigma-Aldrich, St. Louis, MO), 10% fetal bovine serum, 5% CO_2, at 37 °C, and both tested free of Mycoplasma infection directly prior to experimentation.

TCGA-HNSC

Transcriptome profiling, biospecimens, and clinical data from The Cancer Genome Atlas Program (TCGA) was downloaded from the National Cancer Institute GDC Data Portal for patients with the cancer type head and neck squamous cell carcinoma (HNSC) (https://portal.gdc.cancer.gov/projects/TCGA-HNSC)][13] Additional clinical data for this project was downloaded from cBioPortal (https://www.cbioportal.org/study/summary?id=hnsc_tcga_pan_can_atlas_2018).[91,92]

Tissue microarrays

YAP staining was evaluated in samples from the Head and Neck Tissue Microarray Initiative.[78]

METHOD DETAILS

Husbandry and genotyping

Krt14-CreERT$^{+/+}$*/LSL-rtTA*$^{+/+}$*/H2B-GFP*$^{+/+}$*/E6-E7*$^{+/-}$ mice were crossed to *Krt14-CreERT*$^{+/+}$*/LSL-rtTA*$^{+/+}$*/H2B-GFP*$^{+/+}$*/YAP1*$^{S127A+/-}$ resulting in Mendelian proportions of N, E, Y, and EY littermates. At 3-4 weeks of age, a tail fragment was obtained for initial screening genotype confirmation. Directly prior to transgene induction for experiments, mice were assigned to age and sex-balanced groups, and an ear fragment was obtained for confirmatory genotyping. Genomic DNA was isolated by incubating tissue in 25mM NaOH and 0.2 mM EDTA at 100°C for

1 hour, followed by neutralization with an equal volume of 40 mM Tris-HCl (pH 5.5).[93] Multiplex polymerase chain reaction (PCR)-based genotyping was performed using REDTaq® polymerase per manufacturer recommendations (Millipore Sigma). Oligonucleotides were multiplexed as follows: (1) *LSL-rtTA* and *E6-E7* and *Il2* (positive control), (2) *Yap1*^{S127A} and *Trp53* (positive control), (3) *Krt14-CreERT* and *Il2* (positive control). All PCR products were subjected to electrophoresis on 2% agarose gel in Tris acetate EDTA buffer.

Transgene induction

Mice were anesthetized with isoflurane and 100µL of tamoxifen solution (20 mg/mL in miglyol) was administered into the tongue under stereomicroscopic visualization. One dose of tamoxifen was administered every other day for a total of 3 doses.

Epithelia isolation

After *in situ* infiltration with 500uL collagenase+dispase solution (1mg/mL, 2.5 mg/mL) (Millipore Sigma), the tongues of euthanized mice were dissected free and incubated for 30 minutes at 37°C. The tongue epithelium was then dissected free from the underlying muscle under stereomicroscopic visualization.

LoxP–STOP–LoxP excision assay

For floxed stop cassette excision assay, high-quality genomic DNA was isolated from whole epithelia using the DNeasy Blood and Tissue Kit per manufacturer protocol (Qiagen). PCR products were generated using REDTaq® polymerase and LSL excision primers, and were subjected to electrophoresis on 2% agarose gel in Tris acetate EDTA buffer.

RT-qPCR

RNA was prepared by homogenization of whole tongue epithelia in TRIzol® (Invitrogen) followed by phenol:chloroform extraction and RNeasy Mini Kt based column purification with on-column DNase treatment (Qiagen). For quantitative PCR (qPCR), cDNA library preparation was performed using Bio-Rad iScript™ reverse transcriptase and qPCR was performed using Applied Biosystems Fast SYBR® Green Master Mix per manufacturer's instructions.

Evaluation of gross tongue lesions

Following transgene induction, mice were examined under anesthesia using a stereomicroscope every 3-7 days for the appearance of tongue epithelial lesions. Lesion free survival in days was defined as the time from first tamoxifen treatment to the appearance of the first gross lesion.

Histopathology and immunohistochemical staining

Mouse tongues were transected from the pharynx and floor of mouth, and placed in 10% aqueous buffered zinc formalin for 24-36 hours at room temperature and transferred to 70% ethanol. Tissues were paraffin embedded, sectioned (5µm), and stained with hematoxylin and eosin (H&E) by standard protocols (Histoserv or LJI histology). H&E slides were prepared according to a standard protocol. https://www.protocols.io/view/hematoxylin-amp-eosin-protocol-for-leica-st5020-au-x54v9mozqg3e/v1

Immunohistochemistry (IHC) was performed as previously described.[97] After deparaffinization, antigen retrieval was performed using IHC Antigen Retrieval Solution (ThermoFisher, 00-4955-58) in a steamer for 40 minutes. Endogenous peroxidases were inactivated using Bloxall Blocking Solution (Vector Labs, SP-6000, 30-min incubation, room temperature). Tissues were incubated with primary antibody overnight at 4°C then exposed to biotinylated anti-rabbit secondary antibody (Vector Labs, BA-1000, 1:400 dilution, 30 min at room temperature

followed by avidin-biotin complex formation (Vector Laboratories, # PK-6100), staining with DAB substrate (Vector Laboratories, # SK-4105), and hematoxylin counterstain (Mayer's Hematoxylin solution (Sigma MHS1-100ML)). All H&E and IHC stained slides were scanned using the Leica Aperio AT2 slide scanner at 40x magnification.

Fluorescence microscopy

For whole mount fluorescence imaging, epithelial sheets were isolated from mice euthanized 36 hours after a single dose of intralingual tamoxifen, washed in HBSS, stained with NucBlue for 3 hours at room temperature, washed again, mounted immersed in HBSS between two cover glasses, and Z-stacks were acquired with a confocal microscope.

For cross-sectional fluorescence imaging, immediately after euthanasia, mice underwent intracardiac perfusion first with 2mM EDTA in PBS followed by 1.6% paraformaldehyde in PBS. Perfusion fixed tongues were dissected and incubated in 1.6% paraformaldehyde at room temperature overnight, then transferred to 30% sucrose in PBS for 2-3 days at 4°C, then washed in PBS, then embedded in OCT media and snap frozen in cryomolds for frozen section slide preparation. For fluorescent analyses, slides were thawed in the dark, blocked, incubated overnight at 4°C with primary antibodies, and then incubated with fluorophore-conjugated secondary antibodies for 2 hours at room temperature. Nuclei were then stained with Hoechst 33342 in PBS for 15 minutes and slides were mounted with ProLong Diamond mounting medium.

Second harmonic generation for collagen imaging

The second-harmonic generation imaging was done on an upright Leica SP8 microscope with a resonant scanner and hybrid non-descanned detectors. Ti-Sapphire femtosecond pulsed Chameleon Ultra II (Coherent Inc.) laser was tuned to 855 nm and the beam was focused on the sample with an HC PL APO CS 10x/0.40 dry objective. The light was routed to the detectors with 560 nm, 495 nm, and 640 nm long-pass dichroic mirrors. The SHG signal was recorded with a 425/26 nm bandpass filter, the autofluorescence was recorded with a 650/60 nm bandpass filter. The pixel size was set to 0.746 µm, and 16x line averaging was used to improve the signal-to-noise ratio. Data were digitized in an 8-bit mode. The sample navigator software module was used to create autofocus support points and individual fields of view were tiled and stitched.

RNAseq

Tongue epithelia were isolated and RNA was prepared as described above. RNA samples passing purity, concentration, and integrity quality metrics by NanoDrop and TapeStation were submitted to Novogene for oligo-dT-based mRNA selection, cDNA library preparation, and sequencing on Illumina NovaSeq6000.

siRNA transfection in human cell lines

All human cells were transfected at 60% confluency using Lipofectamine RNAiMAX reagent according to the manufacturer's instructions, using 20nM of each siRNA. Culture media was refreshed at 24 hours after transfection. Cells were placed under serum free conditions at 48 hours, and collected for experimentation at 72 hours post-transfection.

Immunoblot assay

Cells rinsed with ice cold PBS and lysates were harvested in RIPA buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1% NP-40) supplemented with Halt™ Protease and Phosphatase Inhibitor Cocktail (#78440, ThermoFisher Scientific) and cleared by centrifugation for 15 minutes. The concentration of supernatants was measured using Bradford colorimetric

assay. Equal amounts of protein were loaded onto 10% polyacrylamide gels, subjected to electrophoresis in Tris/Glycine/SDS buffer, and transferred to PVDF membranes. The membranes were blocked with 5% milk in TBS with 0.1% Tween-20 (TBS-T) buffer for 1 hour, incubated with primary antibodies diluted in 5% BSA overnight at 4°C. After washing 3 times with TBS-T, the membranes were incubated with HRP-conjugated secondary antibodies diluted in 5% milk in TBS-T for 1 h at room temperature. Immobilon Western Chemiluminescent HRP substrate (Millipore, MA) was used for detection.

Cytokine array

Whole epithelia were homogenized in RIPA lysis buffer with protease and phosphatase inhibitors, snap frozen, and sent to Eve Technologies for Mouse Cytokine/Chemokine 44-Plex and Mouse MMP 5-plex Discovery Assay® Arrays.

Generation of epithelial cell suspensions

Isolated epithelia were minced in 0.25% trypsin-EDTA (Thermo) and subjected to mechanical dissociation in the gentleMACS dissociator C tubes (Miltenyi #130-095-937) for 12 minutes at 37°C, followed by inactivation of trypsin and filtration.

Flow cytometry

Epithelial cell suspensions were stained for viability using LIVE/DEAD™ Fixable Blue Dead Cell Stain Kit (Thermo #L23105) and BUV737 Rat Anti-Mouse CD45 Clone 30-F11 (BD Biosciences # 568344) and analyzed using a 5-laser Cytek Aurora.

scRNAseq

Two mice per genotype were used to generate tongue epithelial cell suspensions, as described above. Cells were sorted on a BD FACSAria-II. Viable single cells were selected by size (FSC x SSC) and viability (double negative for propridium iodide and Fixable Viability Dye eFluor780 (eBioscience) staining) parameters. Sorted cells were then loaded on a Chromium Controller (10x Genomics) using the Chromium Next GEM Single Cell 3' Kit v3.1 (10x #1000269) according to the manufacturer's protocol with a target of 10,000 cells per GEM reaction. The resulting cDNA library was sequenced on an Illumina NovaSeq 6000 using the S1 100 cycle Reagent Kit v1.5 (Illumina 200228319), with a targeted read depth of 20,000 reads/cell.

Orthotopic implantation

Epithelial cell suspensions were generated as described above. Cell count and viability was performed using trypan blue on the Countess III. Cells were only implanted for epithelia that showed minimum viability of 75%. After one wash in HBSS, $2x10^5$ viable cells were implanted orthotopically into the tongues of NSG mice. After implantation, mice tongues were first evaluated at 5 days after implantation then every other day until endpoint was reached.

Imaging equipment and software

Gross evaluation of tongue lesions (Figures 1 & 5) was performed using the Motic K-400P stereo microscope. Fluorescent cross-sectional images of were acquired using the Zeiss LSM780 confocal microscope system with Zeiss Black software (Figures S1, 2, S2A, 6), Zeiss AxioZ1 slide scanner (Figure S2C), or Zeiss LSM990 confocal microscope system with Zeiss Blue software (Figure S5). Histological images (H&E, IHC, immunofluorescence) were analyzed using QuPath 0.2.3, ImageJ/FIJI, or MATLAB.

QUANTIFICATION AND STATISTICAL ANALYSIS

Histopathological analyses

Histopathological changes for each experimental condition were independently evaluated by at least two board-certified veterinary pathologists (KK, KSakaguchi, AAM). Carcinoma was defined as atypical epithelial cells deep to the basement membrane. Average epithelial thickness was determined on mid-tongue axial sections by measuring 8-10 orthogonal lines from the basement membrane to the epithelial surface. Carcinoma burden was defined as the number of independent carcinoma foci identified in individual tongues. Carcinoma size was defined as the cross-sectional area of atypical epithelial cells invading deep to the basement membrane. Carcinoma depth of invasion was measured using a line orthogonal to the basement membrane of the closest adjacent normal mucosa to the deepest point of tumor invasion.

Quantification of IHC

Basal phospho-S6 staining was quantified in QuPath using a trained pixel classifier applied to manually-segmented epithelial basal layer regions of interest (ROI). At least three ROIs each with a minimum area of 20,000 um^2 were analyzed per sample. The fraction of phosphoS6 positive pixels for each sample was calculated using the mean of the ROIs after adjusting for relative area per ROI. Suprabasal Ki67, p63, Sox2+ nuclei were quantified in a similar fashion using a trained object classifier applied to manually-segmented epithelial subrabasal layer regions of interest. At least three suprabasal layer ROIs with minimum area of 50,000 um^2 were selected per sample.The fraction of Ki67, p63, or Sox2+ positive nuclei for each sample was calculated using the mean of the ROIs after adjusting for relative area per ROI.

IF nuclear segmentation

Instance segmentation was performed using Stardist, a deep-learning based segmentation FIJI plugin. Distinct grayscales values were assigned to each nucleus called by Stardist. MATLAB scripts were then developed to parse the label images, placing the linear indices corresponding to each pixel within a nucleus into a cell array. The relative size of each cell array corresponded to the number of segmented nuclei per image. The Hoechst and GFP channels from each confocal image were separated for independent segmentation and cell array formation. The ratio of GFP+ to Hoechst+ nuclei were calculated by comparing the resulting nuclear pixel cell arrays for the GFP and Hoechst channels for each image. The Hoechst channel was used for normalization and calculation of the mean fluorescence intensity (MFI) in the GFP channel. A threshold value was calculated using this normalized MFI to call Hoechst+/GFP+ and Hoechst+/GFP- nuclei.

IF spatial context analysis

To determine the spatial fluorescent intensity distribution of ITGA6 as a function of distance from the basement membrane, MFI was calculated along the manually traced basement membrane by calculating vertical shifts from the basement membrane on a per-pixel basis. An array of each pixel location was created, and the MFI along the length of the traced basement membrane was calculated. As epithelial intensity distributions varied based on the length of the basement membrane tracing, they were interpolated using a spline with 50,000 query points, allowing the arrays to be combined and a normalized average intensity distribution to be calculated for each set of image arrays for a given mouse.

Statistical analyses

Statistical analyses were performed in GraphPad Prism 9.5.1 with an alpha threshold of 0.05. Groupwise comparisons were tested using Kruskal-Wallis (one-way ANOVA on ranks) test with Dunn's post-hoc correction for multiple comparison. Differences in survival were compared by Mantel-Cox Log-Rank test with Bonferroni correction for multiple comparisons. Pairwise comparisons between normal and malignant tumors was conducted using two-tailed paired t

test. The correlation between mTOR and YAP signatures among TCGA tumors was determined initially by performing a simple linear regression and tested for significance using a two-tailed Spearman's test. Additional statistical analyses on bulk RNAseq and scRNAseq data were performed in R (v4.1.2, 2021-11-01, "Bird Hippie"); see respective sections for details.

Bulk RNAseq analyses

Paired-end reads were aligned using STAR v2.7.9 using default settings. STAR index was created using the GRCm39 primary genome FASTA and annotation files. The resulting BAM files were sorted by name using samtools v1.7 then gene counts were quantified using HTSeq-count v0.13.5. Pairwise differential expression was calculated and principal component analysis plots were created using DESeq2 v1.34.0.

GO & GSEA analyses

Gene ontology (GO) analyses were performed using GeneOntology.org (Panther 17.0) using significantly differentially expressed genes at ($|\log2FC| > 1$, p-value < 0.01). Gene set enrichment analysis (GSEA[53]) was conducted using the Julia packages Match.jl and BioLab.jl, which contain bioinformatics and computational biology functions under active development. Prior to GSEA analyses, raw bulk RNAseq reads were aligned to the human reference transcriptome with the pseudo-aligner Kallisto[98] using the "quant" command. Transcript expression values were normalized to transcript per million. Transcript expression was converted to gene expression using the maximum individual transcript expression. Single sample GSEA[46] was performed with rank normalization against MSigDB[52,54] gene set collections c2, c3, c5, and c6, with 10,000 permutations. Enriched and depleted gene sets were prioritized based on respective information coefficients[99–101] [6, 7, 8] and Bonferroni-corrected chi-square p-values.

Global clustering

Single cell gene expression data was processed from the Illumina sequencer files using Cell Ranger (v5.0.0) and its prebuilt mouse reference genome. Individual sample data was then processed and merged using the Seurat (v4.3.0) SCTransform pipeline. Low quality cells (mitochondrial percentage >7, features <1000 and >5500, transcripts per cell >30,000) were filtered prior to data scaling and normalization. After filtering, data was transformed using SCTransform with default parameters, regressing on percent mitochondrial content. Principal component analysis was performed with RunPCA, using the top 50 PCs. Dimensionality reduction was performed with RunUMAP, using the top 30 dimensions. Nearest-neighbor analysis was performed using FindNeighbors using the top 30 dimensions and with k.param set to 50. Clustering was performed with FindClusters with resolution 0.3. Marker genes were calculated using FindAllMarkers with default parameters. Cluster identities were assigned by analysis of differential gene expression.

Epithelial and immune cell sub-clustering

Epithelial and immune cell subset Seurat objects were generated and analyzed individually using Seurat version 4.3.0. Following assignment of cells to epithelial or immune cell subsets, the analysis pipeline used for the combined analysis was run again for each individual subset with identical parameters. Following initial subset clustering, contaminating residual immune or epithelial cells were removed from the subset Seurat objects, and the subset data reanalyzed in Seurat.

Transgene alignment

For quantification of the transgene expression in single cells, STARsolo algorithm implemented to STAR aligner version 2.7.9a was applied. Briefly, the custom FASTA file was generated by

merging the mm10 mouse genome and the transgene sequences. The index file for this custom genome was generated by STAR using the custom GTF file including the annotations for transgenes with the following parameters; *--sjdbOverhang 100*, *--genomeSAsparseD 3*. Subsequently, the FASTQ files including cDNA reads and cell barcodes of each sample were aligned to the custom genome by STARsolo with the following parameters; *--soloType CB_UMI_Simple*, *--clipAdapterType CellRanger4*, *--outFilterScoreMin 30*, *--soloCBmatchWLtype 1MM_multi_Nbase_pseudocounts*, *--soloUMIfiltering MultiGeneUMI_CR*, *-soloUMIdedup 1MM_CR*, *--soloCellFilter EmptyDrops_CR*. Finally, the cell-gene count arrays of the transgenes for each sample were obtained as the output of STARsolo. The data was imported to R and normalized with Seurat R package with "NormalizeData" function using the following options; *normalization.method = "LogNormalize"* and *scale.factor = 10000*. The normalized transgene expression arrays were merged with the Seurat object of the epithelial cell cluster by cell barcodes for downstream analysis.

Sequence: HPV16$^{E6\text{-}E7}$

ATGCACCAAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTACCA
CAGTTATGCACAGAGCTGCAAACAACTATACATGATATAATATTAGAATGTGTGTACTGCAA
GCAACAGTTACTGCGACGTGAGGTATATGACTTTGCTTTTCGGGATTTATGCATAGTATATA
GAGATGGGAATCCATATGCTGTATGTGATAAATGTTTAAAGTTTTATTCTAAAATTAGTGAGT
ATAGACATTATTGTTATAGTTTGTATGGAACAACATTAGAACAGCAATACAACAAACCGTTG
TGTGATTTGTTAATTAGGTGTATTAACTGTCAAAAGCCACTGTGTCCTGAAGAAAAGCAAAG
ACATCTGGACAAAAAGCAAAGATTCCATAATATAAGGGGTCGGTGGACCGGTCGATGTATG
TCTTGTTGCAGATCATCAAGAACACGTAGAGAAACCCAGCTGTAATCATGCATGGAGATAC
ACCTACATTGCATGAATATATGTTAGATTTGCAACCAGAGACAACTGATCTCTACTGTTATG
AGCAATTAAATGACAGCTCAGAGGAGGAGGATGAAATAGATGGTCCAGCTGGACAAGCAG
AACCGGACAGAGCCCATTACAATATTGTAACCTTTTGTTGCAAGTGTGACTCTACGCTTCG
GTTGTGCGTACAAAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAATGGGCACA
CTAGGAATTGTGTGCCCCATCTGTTCTCAGAAACCATAA

Sequence: YAP1^{S127A} (S127A codon underlined and in bold)

ATGGATCCCGGGCAGCAGCCGCCGCCTCAACCGGCCCCCCAGGGCCAAGGGCAGCCGC
CTTCGCAGCCCCCGCAGGGGCAGGGCCCGCCGTCCGGACCCGGGCAACCGGCACCCGC
GGCGACCCAGGCGGCGCCGCAGGCACCCCCCGCCGGGCATCAGATCGTGCACGTCCGC
GGGGACTCGGAGACCGACCTGGAGGCGCTCTTCAACGCCGTCATGAACCCCAAGACGGC
CAACGTGCCCCAGACCGTGCCCATGAGGCTCCGGAAGCTGCCCGACTCCTTCTTCAAGCC
GCCGGAGCCCAAATCCCACTCCCGACAGGCCAGTACTGATGCAGGCACTGCAGGAGCCC
TGACTCCACAGCATGTTCGAGCTCAT**<u>gCC</u>**TCTCCAGCTTCTCTGCAGTTGGGAGCTGTTTC
TCCTGGGACACTGACCCCCACTGGAGTAGTCTCTGGCCCAGCAGCTACACCCACAGCTCA
GCATCTTCGACAGTCTTCTTTTGAGATACCTGATGATGTACCTCTGCCAGCAGGTTGGGAG
ATGGCAAAGACATCTTCTGGTCAGAGATACTTCTTAAATCACATCGATCAGACAACAACATG
GCAGGACCCCAGGAAGGCCATGCTGTCCCAGATGAACGTCACAGCCCCCACCAGTCCAC
CAGTGCAGCAGAATATGATGAACTCGGCTTCAGGTCCTCTTCCTGATGGATGGGAACAAG
CCATGACTCAGGATGGAGAAATTTACTATATAAACCATAAGAACAAGACCACCTCTTGGCTA
GACCCAAGGCTTGACCCTCGTTTTGCCATGAACCAGAGAATCAGTCAGAGTGCTCCAGTG
AAACAGCCACCACCCCTGGCTCCCCAGAGCCCACAGGGAGGCGTCATGGGTGGCAGCAA
CTCCAACCAGCAGCAACAGATGCGACTGCAGCAACTGCAGATGGAGAAGGAGAGGCTGC
GGCTGAAACAGCAAGAACTGCTTCGGCAGGCAATGCGGAATATCAATCCCAGCACAGCAA
ATTCTCCAAAATGTCAGGAGTTAGCCCTGCGTAGCCAGTTACCAACACTGGAGCAGGATG
GTGGGACTCAAAATCCAGTGTCTTCTCCCGGGATGTCTCAGGAATTGAGAACAATGACGAC
CAATAGCTCAGATCCTTTCCTTAACAGTGGCACCTATCACTCTCGAGATGAGAGTACAGAC
AGTGGACTAAGCATGAGCAGCTACAGTGTCCCTCGAACCCCAGATGACTTCCTGAACAGT

GTGGATGAGATGGATACAGGTGATACTATCAACCAAAGCACCCTGCCCTCACAGCAGAAC
CGTTTCCCAGACTACCTTGAAGCCATTCCTGGGACAAATGTGGACCTTGGAACACTGGAAG
GAGATGGAATGAACATAGAAGGAGAGGAGCTGATGCCAAGTCTGCAGGAAGCTTTGAGTT
CTGACATCCTTAATGACATGGAGTCTGTTTTGGCTGCCACCAAGCTAGATAAAGAAAGCTTT
CTTACATGGTTA

Weighted Gene Co-expression Network Analysis (WGCNA) of scRNAseq data.
R package hdWGCNA version 0.2.03 (https://smorabit.github.io/hdWGCNA/) was used for WGCNA analysis in the scRNAseq dataset. Normalization of the integrated Seurat object containing cell-gene expression arrays of EY-genotype epithelial cells was performed using NormalizeMetacells using parmeters k=10, max_shared=10, min_cells=20. A soft thresholding power was determined as 8 using the function TestSoftPowers and applied for estimation of co-expressing network in the EY-genotype scRNAseq dataset. Significantly co-expressed module genes and highly connected genes within each module (hub genes) were identified by computing eigengene-based connectivity (kME). The heatmap representing topology overlap matrix (TOM) of module genes was generated using R package ComplexHeatmap version 2.14.0. Genes with signed module eigengene-based connectivity measure (kME) greater 0.3 were considered as moderate to high confidence module genes. Modules were assigned functional annotations based on enrichment of member genes for biological processes using Enrichr[102,103] and MetaScape.[104]

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the Lead Contact, J. Silvio Gutkind (sgutkind@health.ucsd.edu).

Materials availability

Transgenic mice and cell lines generated in this study are available from the Lead Contact upon reasonable request and completion of Material Transfer Agreements (MTA). There may restrictions to the availability of these reagents due to cost or limited quantities.

Data and Code Availability

Original data and code will be deposited to the respective standardized accessible databases including NCBI SRA (https://www.ncbi.nlm.nih.gov/sra, RNAseq data), Image Data Resource (IDR, https://idr.openmicroscopy.org/, microscopic images), and Mendeley Data (https://data.mendeley.com/, analytic code markdowns and additional datasets).

SUPPLEMENTARY INFORMATION

Table S1 – Gene sets used for GSEA analyses
Table S2 – Differentially expressed genes on bulk RNAseq (EY vs N, Y vs N, E vs N)
Table S3 – EY-unique DEGs and gene ontology analysis
Table S4 – Transcription factor analysis of EY vs N DEGs
Table S5 – Epithelial cluster marker genes
Table S6 – Immune cluster marker genes
Table S7 – EY-cluster WGCNA module genes and gene ontology analysis

REFERENCES

1. Clayton, E., Doupé, D.P., Klein, A.M., Winton, D.J., Simons, B.D., and Jones, P.H. (2007). A single type of progenitor cell maintains normal epidermis. Nature *446*, 185–189. 10.1038/nature05574.
2. Sánchez-Danés, A., and Blanpain, C. (2018). Deciphering the cells of origin of squamous cell carcinomas. Nat Rev Cancer *18*, 549–561. 10.1038/s41568-018-0024-5.
3. Barker, N., Ridgway, R.A., van Es, J.H., van de Wetering, M., Begthel, H., van den Born, M., Danenberg, E., Clarke, A.R., Sansom, O.J., and Clevers, H. (2009). Crypt stem cells as the cells-of-origin of intestinal cancer. Nature *457*, 608–611. 10.1038/nature07602.
4. Ying, Z., Sandoval, M., and Beronja, S. (2018). Oncogenic activation of PI3K induces progenitor cell differentiation to suppress epidermal growth. Nat Cell Biol *20*, 1256–1266. 10.1038/s41556-018-0218-9.
5. Bartkova, J., Rezaei, N., Liontos, M., Karakaidos, P., Kletsas, D., Issaeva, N., Vassiliou, L.-V.F., Kolettas, E., Niforou, K., Zoumpourlis, V.C., et al. (2006). Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. Nature *444*, 633–637. 10.1038/nature05268.
6. Evan, G.I., Wyllie, A.H., Gilbert, C.S., Littlewood, T.D., Land, H., Brooks, M., Waters, C.M., Penn, L.Z., and Hancock, D.C. (1992). Induction of apoptosis in fibroblasts by c-myc protein. Cell *69*, 119–128. 10.1016/0092-8674(92)90123-t.
7. Berger, A.H., Knudson, A.G., and Pandolfi, P.P. (2011). A continuum model for tumour suppression. Nature *476*, 163–169. 10.1038/nature10275.
8. Jones, K.B., Furukawa, S., Marangoni, P., Ma, H., Pinkard, H., D'Urso, R., Zilionis, R., Klein, A.M., and Klein, O.D. (2019). Quantitative Clonal Analysis and Single-Cell Transcriptomics Reveal Division Kinetics, Hierarchy, and Fate of Oral Epithelial Progenitor Cells. Cell Stem Cell *24*, 183-192.e8. 10.1016/j.stem.2018.10.015.
9. Chen, D., Wu, M., Li, Y., Chang, I., Yuan, Q., Ekimyan-Salvo, M., Deng, P., Yu, B., Yu, Y., Dong, J., et al. (2017). Targeting BMI1 + Cancer Stem Cells Overcomes Chemoresistance and Inhibits Metastases in Squamous Cell Carcinoma. Cell Stem Cell *20*, 621-634.e6. 10.1016/j.stem.2017.02.003.
10. Blanpain, C., and Fuchs, E. (2009). Epidermal homeostasis: a balancing act of stem cells in the skin. Nat Rev Mol Cell Biol *10*, 207–217. 10.1038/nrm2636.
11. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin *68*, 394–424. 10.3322/caac.21492.
12. Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D.M., Piñeros, M., Znaor, A., and Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer *144*, 1941–1953. 10.1002/ijc.31937.
13. Cancer Genome Atlas Network (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature *517*, 576–582. 10.1038/nature14129.
14. Stransky, N., Egloff, A.M., Tward, A.D., Kostic, A.D., Cibulskis, K., Sivachenko, A., Kryukov, G.V., Lawrence, M.S., Sougnez, C., McKenna, A., et al. (2011). The mutational landscape of head and neck squamous cell carcinoma. Science *333*, 1157–1160. 10.1126/science.1208130.
15. Agrawal, N., Frederick, M.J., Pickering, C.R., Bettegowda, C., Chang, K., Li, R.J., Fakhry, C., Xie, T.-X., Zhang, J., Wang, J., et al. (2011). Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science *333*, 1154–1157. 10.1126/science.1206923.
16. Iglesias-Bartolome, R., Martin, D., and Gutkind, J.S. (2013). Exploiting the Head and Neck Cancer Oncogenome: Widespread PI3K-mTOR Pathway Alterations and Novel Molecular Targets. Cancer Discovery *3*, 722–725. 10.1158/2159-8290.CD-13-0239.

17. Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., Chinnaiyan, A., Israel, M.A., Goldstein, L.S.B., Abujarour, R., et al. (2010). The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. Genes Dev *24*, 1106–1118. 10.1101/gad.1903310.
18. Puram, S.V., Tirosh, I., Parikh, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C.L., Mroz, E.A., Emerick, K.S., et al. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell *171*, 1611-1624.e24. 10.1016/j.cell.2017.10.044.
19. Puram, S.V., Mints, M., Pal, A., Qi, Z., Reeb, A., Gelev, K., Barrett, T.F., Gerndt, S., Liu, P., Parikh, A.S., et al. (2023). Cellular states are coupled to genomic and viral heterogeneity in HPV-related oropharyngeal carcinoma. Nat Genet *55*, 640–650. 10.1038/s41588-023-01357-3.
20. Kürten, C.H.L., Kulkarni, A., Cillo, A.R., Santos, P.M., Roble, A.K., Onkar, S., Reeder, C., Lang, S., Chen, X., Duvvuri, U., et al. (2021). Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. Nat Commun *12*, 7338. 10.1038/s41467-021-27619-4.
21. Bai, Y., Gotz, C., Chincarini, G., Zhao, Z., Slaney, C., Boath, J., Furic, L., Angel, C., Jane, S.M., Phillips, W.A., et al. (2023). YBX1 integration of oncogenic PI3K/mTOR signalling regulates the fitness of malignant epithelial cells. Nat Commun *14*, 1591. 10.1038/s41467-023-37161-0.
22. Barkley, D., Moncada, R., Pour, M., Liberman, D.A., Dryg, I., Werba, G., Wang, W., Baron, M., Rao, A., Xia, B., et al. (2022). Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. Nat Genet *54*, 1192–1201. 10.1038/s41588-022-01141-9.
23. Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J., and Howley, P.M. (1990). The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell *63*, 1129–1136. 10.1016/0092-8674(90)90409-8.
24. Martinez-Zapien, D., Ruiz, F.X., Poirson, J., Mitschler, A., Ramirez, J., Forster, A., Cousido-Siah, A., Masson, M., Vande Pol, S., Podjarny, A., et al. (2016). Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. Nature *529*, 541–545. 10.1038/nature16481.
25. Münger, K., Werness, B.A., Dyson, N., Phelps, W.C., Harlow, E., and Howley, P.M. (1989). Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. EMBO J *8*, 4099–4105. 10.1002/j.1460-2075.1989.tb08594.x.
26. Carper, M.B., Troutman, S., Wagner, B.L., Byrd, K.M., Selitsky, S.R., Parag-Sharma, K., Henry, E.C., Li, W., Parker, J.S., Montgomery, S.A., et al. (2019). An Immunocompetent Mouse Model of HPV16(+) Head and Neck Squamous Cell Carcinoma. Cell Rep *29*, 1660-1674.e7. 10.1016/j.celrep.2019.10.005.
27. Schiffman, M., Doorbar, J., Wentzensen, N., de Sanjosé, S., Fakhry, C., Monk, B.J., Stanley, M.A., and Franceschi, S. (2016). Carcinogenic human papillomavirus infection. Nat Rev Dis Primers *2*, 16086. 10.1038/nrdp.2016.86.
28. Callejas-Valera, J.L., Iglesias-Bartolome, R., Amornphimoltham, P., Palacios-Garcia, J., Martin, D., Califano, J.A., Molinolo, A.A., and Gutkind, J.S. (2016). mTOR inhibition prevents rapid-onset of carcinogen-induced malignancies in a novel inducible HPV-16 E6/E7 mouse model. Carcinogenesis *37*, 1014–1025. 10.1093/carcin/bgw086.
29. Petitjean, A., Achatz, M.I.W., Borresen-Dale, A.L., Hainaut, P., and Olivier, M. (2007). TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. Oncogene *26*, 2157–2165. 10.1038/sj.onc.1210302.
30. Donehower, L.A., and Lozano, G. (2009). 20 years studying p53 functions in genetically engineered mice. Nat Rev Cancer *9*, 831–841. 10.1038/nrc2731.

31. Kim, N.-G., and Gumbiner, B.M. (2015). Adhesion to fibronectin regulates Hippo signaling via the FAK-Src-PI3K pathway. J Cell Biol *210*, 503–515. 10.1083/jcb.201501025.
32. Polager, S., and Ginsberg, D. (2009). p53 and E2f: partners in life and death. Nat Rev Cancer *9*, 738–748. 10.1038/nrc2718.
33. Martin, D., Degese, M.S., Vitale-Cross, L., Iglesias-Bartolome, R., Valera, J.L.C., Wang, Z., Feng, X., Yeerna, H., Vadmal, V., Moroishi, T., et al. (2018). Assembly and activation of the Hippo signalome by FAT1 tumor suppressor. Nat Commun *9*, 2372. 10.1038/s41467-018-04590-1.
34. Su, L., Morgan, P.R., and Lane, E.B. (1996). Keratin 14 and 19 expression in normal, dysplastic and malignant oral epithelia. A study using in situ hybridization and immunohistochemistry. J Oral Pathol Med *25*, 293–301. 10.1111/j.1600-0714.1996.tb00265.x.
35. Tang, X.-H., Scognamiglio, T., and Gudas, L.J. (2013). Basal stem cells contribute to squamous cell carcinomas in the oral cavity. Carcinogenesis *34*, 1158–1164. 10.1093/carcin/bgt021.
36. Vasioukhin, V., Degenstein, L., Wise, B., and Fuchs, E. (1999). The magical touch: genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin. Proc Natl Acad Sci U S A *96*, 8551–8556. 10.1073/pnas.96.15.8551.
37. Tumbar, T., Guasch, G., Greco, V., Blanpain, C., Lowry, W.E., Rendl, M., and Fuchs, E. (2004). Defining the epithelial stem cell niche in skin. Science *303*, 359–363. 10.1126/science.1092436.
38. Camargo, F.D., Gokhale, S., Johnnidis, J.B., Fu, D., Bell, G.W., Jaenisch, R., and Brummelkamp, T.R. (2007). YAP1 increases organ size and expands undifferentiated progenitor cells. Curr Biol *17*, 2054–2060. 10.1016/j.cub.2007.10.039.
39. Cuylen, S., Blaukopf, C., Politi, A.Z., Müller-Reichert, T., Neumann, B., Poser, I., Ellenberg, J., Hyman, A.A., and Gerlich, D.W. (2016). Ki-67 acts as a biological surfactant to disperse mitotic chromosomes. Nature *535*, 308–312. 10.1038/nature18610.
40. Melino, G., Memmi, E.M., Pelicci, P.G., and Bernassola, F. (2015). Maintaining epithelial stemness with p63. Sci Signal *8*, re9. 10.1126/scisignal.aaa1033.
41. Arnold, K., Sarkar, A., Yram, M.A., Polo, J.M., Bronson, R., Sengupta, S., Seandel, M., Geijsen, N., and Hochedlinger, K. (2011). Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. Cell Stem Cell *9*, 317–329. 10.1016/j.stem.2011.09.001.
42. Senoo, M., Pinto, F., Crum, C.P., and McKeon, F. (2007). p63 Is essential for the proliferative potential of stem cells in stratified epithelia. Cell *129*, 523–536. 10.1016/j.cell.2007.02.045.
43. Igarashi, T., Shimmura, S., Yoshida, S., Tonogi, M., Shinozaki, N., and Yamane, G. -y (2008). Isolation of oral epithelial progenitors using collagen IV. Oral Dis *14*, 413–418. 10.1111/j.1601-0825.2007.01390.x.
44. Krebsbach, P.H., and Villa-Diaz, L.G. (2017). The Role of Integrin α6 (CD49f) in Stem Cells: More than a Conserved Biomarker. Stem Cells Dev *26*, 1090–1099. 10.1089/scd.2016.0319.
45. Wang, Y., Xu, X., Maglic, D., Dill, M.T., Mojumdar, K., Ng, P.K.-S., Jeong, K.J., Tsang, Y.H., Moreno, D., Bhavana, V.H., et al. (2018). Comprehensive Molecular Characterization of the Hippo Signaling Pathway in Cancer. Cell Rep *25*, 1304-1317.e5. 10.1016/j.celrep.2018.10.001.
46. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112. 10.1038/nature08460.

47. Amornphimoltham, P., Patel, V., Sodhi, A., Nikitakis, N.G., Sauk, J.J., Sausville, E.A., Molinolo, A.A., and Gutkind, J.S. (2005). Mammalian Target of Rapamycin, a Molecular Target in Squamous Cell Carcinomas of the Head and Neck. Cancer Research *65*, 9953–9961. 10.1158/0008-5472.CAN-05-0921.
48. Hay, N., and Sonenberg, N. (2004). Upstream and downstream of mTOR. Genes Dev *18*, 1926–1945. 10.1101/gad.1212704.
49. Chai, A.W.Y., Yee, P.S., Price, S., Yee, S.M., Lee, H.M., Tiong, V.K., Gonçalves, E., Behan, F.M., Bateson, J., Gilbert, J., et al. (2020). Genome-wide CRISPR screens of oral squamous cell carcinoma reveal fitness genes in the Hippo pathway. Elife *9*, e57761. 10.7554/eLife.57761.
50. Mi, H., Muruganujan, A., and Thomas, P.D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res *41*, D377-386. 10.1093/nar/gks1118.
51. Thomas, P.D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L.-P., and Mi, H. (2022). PANTHER: Making genome-scale phylogenetics accessible to all. Protein Sci *31*, 8–22. 10.1002/pro.4218.
52. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst *1*, 417–425. 10.1016/j.cels.2015.12.004.
53. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545–15550. 10.1073/pnas.0506580102.
54. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740. 10.1093/bioinformatics/btr260.
55. Iwakura, Y., and Ishigame, H. (2006). The IL-23/IL-17 axis in inflammation. J Clin Invest *116*, 1218–1222. 10.1172/JCI28508.
56. Keenan, A.B., Torre, D., Lachmann, A., Leong, A.K., Wojciechowicz, M.L., Utti, V., Jagodnik, K.M., Kropiwnicki, E., Wang, Z., and Ma'ayan, A. (2019). ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Res *47*, W212–W224. 10.1093/nar/gkz446.
57. Min, M., Chen, X.-B., Wang, P., Landeck, L., Chen, J.-Q., Li, W., Cai, S.-Q., Zheng, M., and Man, X.-Y. (2017). Role of keratin 24 in human epidermal keratinocytes. PLoS One *12*, e0174626. 10.1371/journal.pone.0174626.
58. Tu, S., Zheng, J., Gao, X., Guan, C., Cai, B., and Xiang, L. (2018). The role of Foxq1 in proliferation of human dental pulp stem cell. Biochem Biophys Res Commun *497*, 543–549. 10.1016/j.bbrc.2018.02.077.
59. Rezza, A., Wang, Z., Sennett, R., Qiao, W., Wang, D., Heitman, N., Mok, K.W., Clavel, C., Yi, R., Zandstra, P., et al. (2016). Signaling Networks among Stem Cell Precursors, Transit-Amplifying Progenitors, and their Niche in Developing Hair Follicles. Cell Rep *14*, 3001–3018. 10.1016/j.celrep.2016.02.078.
60. Godwin, A.R., and Capecchi, M.R. (1998). Hoxc13 mutant mice lack external hair. Genes Dev *12*, 11–20. 10.1101/gad.12.1.11.
61. Jonker, L., Kist, R., Aw, A., Wappler, I., and Peters, H. (2004). Pax9 is required for filiform papilla development and suppresses skin-specific differentiation of the mammalian tongue epithelium. Mech Dev *121*, 1313–1322. 10.1016/j.mod.2004.07.002.
62. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2016). Fast gene set enrichment analysis (Bioinformatics) 10.1101/060012.

63. Wong, D.J., Liu, H., Ridky, T.W., Cassarino, D., Segal, E., and Chang, H.Y. (2008). Module map of stem cell genes guides creation of epithelial cancer stem cells. Cell Stem Cell *2*, 333–344. 10.1016/j.stem.2008.02.009.
64. Kuo, P., Tuong, Z.K., Teoh, S.M., Frazer, I.H., Mattarollo, S.R., and Leggatt, G.R. (2018). HPV16E7-Induced Hyperplasia Promotes CXCL9/10 Expression and Induces CXCR3+ T-Cell Migration to Skin. J Invest Dermatol *138*, 1348–1359. 10.1016/j.jid.2017.12.021.
65. Ambühl, L.M.M., Villadsen, A.B., Baandrup, U., Dybkær, K., and Sørensen, S. (2017). HPV16 E6 and E7 Upregulate Interferon-Induced Antiviral Response Genes ISG15 and IFIT1 in Human Trophoblast Cells. Pathogens *6*, 40. 10.3390/pathogens6030040.
66. Martín-Villar, E., Fernández-Muñoz, B., Parsons, M., Yurrita, M.M., Megías, D., Pérez-Gómez, E., Jones, G.E., and Quintanilla, M. (2010). Podoplanin associates with CD44 to promote directional cell migration. Mol Biol Cell *21*, 4387–4399. 10.1091/mbc.E10-06-0489.
67. Martín-Villar, E., Borda-d'Agua, B., Carrasco-Ramirez, P., Renart, J., Parsons, M., Quintanilla, M., and Jones, G.E. (2015). Podoplanin mediates ECM degradation by squamous carcinoma cells through control of invadopodia stability. Oncogene *34*, 4531–4544. 10.1038/onc.2014.388.
68. Chen, X., Nadiarynkh, O., Plotnikov, S., and Campagnola, P.J. (2012). Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure. Nat Protoc *7*, 654–669. 10.1038/nprot.2012.009.
69. Chen, Q., Zhang, N., Xie, R., Wang, W., Cai, J., Choi, K.-S., David, K.K., Huang, B., Yabuta, N., Nojima, H., et al. (2015). Homeostatic control of Hippo signaling activity revealed by an endogenous activating mutation in YAP. Genes Dev *29*, 1285–1297. 10.1101/gad.264234.115.
70. Schlegelmilch, K., Mohseni, M., Kirak, O., Pruszak, J., Rodriguez, J.R., Zhou, D., Kreger, B.T., Vasioukhin, V., Avruch, J., Brummelkamp, T.R., et al. (2011). Yap1 acts downstream of α-catenin to control epidermal proliferation. Cell *144*, 782–795. 10.1016/j.cell.2011.02.031.
71. Zanconato, F., Forcato, M., Battilana, G., Azzolin, L., Quaranta, E., Bodega, B., Rosato, A., Bicciato, S., Cordenonsi, M., and Piccolo, S. (2015). Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. Nat Cell Biol *17*, 1218–1227. 10.1038/ncb3216.
72. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559. 10.1186/1471-2105-9-559.
73. Morabito, S., Miyoshi, E., Michael, N., Shahin, S., Martini, A.C., Head, E., Silva, J., Leavy, K., Perez-Rosendahl, M., and Swarup, V. (2021). Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. Nat Genet *53*, 1143–1155. 10.1038/s41588-021-00894-z.
74. Omori, H., Nishio, M., Masuda, M., Miyachi, Y., Ueda, F., Nakano, T., Sato, K., Mimori, K., Taguchi, K., Hikasa, H., et al. (2020). YAP1 is a potent driver of the onset and progression of oral squamous cell carcinoma. Sci Adv *6*, eaay3324. 10.1126/sciadv.aay3324.
75. Enomoto, A., Kido, N., Ito, M., Morita, A., Matsumoto, Y., Takamatsu, N., Hosoi, Y., and Miyagawa, K. (2008). Negative regulation of MEKK1/2 signaling by serine-threonine kinase 38 (STK38). Oncogene *27*, 1930–1938. 10.1038/sj.onc.1210828.
76. Bichsel, S.J., Tamaskovic, R., Stegert, M.R., and Hemmings, B.A. (2004). Mechanism of activation of NDR (nuclear Dbf2-related) protein kinase by the hMOB1 protein. J Biol Chem *279*, 35228–35235. 10.1074/jbc.M404542200.
77. Gillison, M.L., Chaturvedi, A.K., Anderson, W.F., and Fakhry, C. (2015). Epidemiology of Human Papillomavirus–Positive Head and Neck Squamous Cell Carcinoma. JCO *33*, 3235–3242. 10.1200/JCO.2015.61.6995.

78. Molinolo, A.A., Hewitt, S.M., Amornphimoltham, P., Keelawat, S., Rangdaeng, S., Meneses García, A., Raimondi, A.R., Jufe, R., Itoiz, M., Gao, Y., et al. (2007). Dissecting the Akt/mammalian target of rapamycin signaling network: emerging results from the head and neck cancer tissue array initiative. Clin Cancer Res *13*, 4964–4973. 10.1158/1078-0432.CCR-07-1041.
79. Molinolo, A.A., Marsh, C., El Dinali, M., Gangane, N., Jennison, K., Hewitt, S., Patel, V., Seiwert, T.Y., and Gutkind, J.S. (2012). mTOR as a Molecular Target in HPV-Associated Oral and Cervical Squamous Carcinomas. Clinical Cancer Research *18*, 2558–2568. 10.1158/1078-0432.CCR-11-2824.
80. Gutkind, J.S., Molinolo, A.A., Wu, X., Wang, Z., Nachmanson, D., Harismendy, O., Alexandrov, L.B., Wuertz, B.R., Ondrey, F.G., Laronde, D., et al. (2021). Inhibition of mTOR signaling and clinical activity of metformin in oral premalignant lesions. JCI Insight *6*, e147096. 10.1172/jci.insight.147096.
81. Vitale-Cross, L., Molinolo, A.A., Martin, D., Younis, R.H., Maruyama, T., Patel, V., Chen, W., Schneider, A., and Gutkind, J.S. (2012). Metformin Prevents the Development of Oral Squamous Cell Carcinomas from Carcinogen-Induced Premalignant Lesions. Cancer Prevention Research *5*, 562–573. 10.1158/1940-6207.CAPR-11-0502.
82. Tumaneng, K., Schlegelmilch, K., Russell, R.C., Yimlamai, D., Basnet, H., Mahadevan, N., Fitamant, J., Bardeesy, N., Camargo, F.D., and Guan, K.-L. (2012). YAP mediates crosstalk between the Hippo and PI(3)K–TOR pathways by suppressing PTEN via miR-29. Nat Cell Biol *14*, 1322–1329. 10.1038/ncb2615.
83. Liotta, L.A. (2016). Adhere, Degrade, and Move: The Three-Step Model of Invasion. Cancer Research *76*, 3115–3117. 10.1158/0008-5472.CAN-16-1297.
84. Rosenthal, E.L., and Matrisian, L.M. (2006). Matrix metalloproteases in head and neck cancer. Head Neck *28*, 639–648. 10.1002/hed.20365.
85. Spira, A., Yurgelun, M.B., Alexandrov, L., Rao, A., Bejar, R., Polyak, K., Giannakis, M., Shilatifard, A., Finn, O.J., Dhodapkar, M., et al. (2017). Precancer Atlas to Drive Precision Prevention Trials. Cancer Research *77*, 1510–1541. 10.1158/0008-5472.CAN-16-2346.
86. Wood, H.M., Conway, C., Daly, C., Chalkley, R., Berri, S., Senguven, B., Stead, L., Ross, L., Egan, P., Chengot, P., et al. (2015). The clonal relationships between pre-cancer and cancer revealed by ultra-deep sequencing. J Pathol *237*, 296–306. 10.1002/path.4576.
87. Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B., Tubio, J.M., et al. (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science *348*, 880–886. 10.1126/science.aaa6806.
88. Cooper, C.S., Eeles, R., Wedge, D.C., Van Loo, P., Gundem, G., Alexandrov, L.B., Kremeyer, B., Butler, A., Lynch, A.G., Camacho, N., et al. (2015). Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. Nat Genet *47*, 367–372. 10.1038/ng.3221.
89. Belteki, G., Haigh, J., Kabacs, N., Haigh, K., Sison, K., Costantini, F., Whitsett, J., Quaggin, S.E., and Nagy, A. (2005). Conditional and inducible transgene expression in mice through the combinatorial use of Cre-mediated recombination and tetracycline induction. Nucleic Acids Res *33*, e51. 10.1093/nar/gni051.
90. Martin, D., Abba, M.C., Molinolo, A.A., Vitale-Cross, L., Wang, Z., Zaida, M., Delic, N.C., Samuels, Y., Lyons, J.G., and Gutkind, J.S. (2014). The head and neck cancer cell oncogenome: a platform for the development of precision molecular therapies. Oncotarget *5*, 8906–8923. 10.18632/oncotarget.2417.
91. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an

open platform for exploring multidimensional cancer genomics data. Cancer Discov *2*, 401–404. 10.1158/2159-8290.CD-12-0095.

92. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal *6*, pl1. 10.1126/scisignal.2004088.
93. Truett, G.E., Heeger, P., Mynatt, R.L., Truett, A.A., Walker, J.A., and Warman, M.L. (2000). Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). Biotechniques *29*, 52, 54. 10.2144/00291bm09.
94. Tang, J.-Y., Yu, C.-Y., Bao, Y.-J., Chen, L., Chen, J., Yang, S.-L., Chen, H.-Y., Hong, J., and Fang, J.-Y. (2018). TEAD4 promotes colorectal tumorigenesis via transcriptionally targeting YAP1. Cell Cycle *17*, 102–109. 10.1080/15384101.2017.1403687.
95. Bordigoni, A., Motte, A., Tissot-Dupont, H., Colson, P., and Desnues, C. (2021). Development and validation of a multiplex qPCR assay for detection and relative quantification of HPV16 and HPV18 E6 and E7 oncogenes. Sci Rep *11*, 4039. 10.1038/s41598-021-83489-2.
96. Wang, X., Spandidos, A., Wang, H., and Seed, B. (2012). PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. Nucleic Acids Research *40*, D1144–D1149. 10.1093/nar/gkr1013.
97. Czerninski, R., Amornphimoltham, P., Patel, V., Molinolo, A.A., and Gutkind, J.S. (2009). Targeting mammalian target of rapamycin by rapamycin prevents tumor progression in an oral-specific chemical carcinogenesis model. Cancer Prev Res (Phila) *2*, 27–36. 10.1158/1940-6207.CAPR-08-0147.
98. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol *34*, 525–527. 10.1038/nbt.3519.
99. Joe, H. (1989). Relative Entropy Measures of Multivariate Dependence. Journal of the American Statistical Association *84*, 157–164. 10.1080/01621459.1989.10478751.
100. Kim, J.W., Botvinnik, O.B., Abudayyeh, O., Birger, C., Rosenbluh, J., Shrestha, Y., Abazeed, M.E., Hammerman, P.S., DiCara, D., Konieczkowski, D.J., et al. (2016). Characterizing genomic alterations in cancer by complementary functional associations. Nat Biotechnol *34*, 539–546. 10.1038/nbt.3527.
101. Linfoot, E.H. (1957). An informational measure of correlation. Information and Control *1*, 85–89. 10.1016/S0019-9958(57)90116-X.
102. Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene Set Knowledge Discovery with Enrichr. Curr Protoc *1*, e90. 10.1002/cpz1.90.
103. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90-97. 10.1093/nar/gkw377.
104. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun *10*, 1523. 10.1038/s41467-019-09234-6.

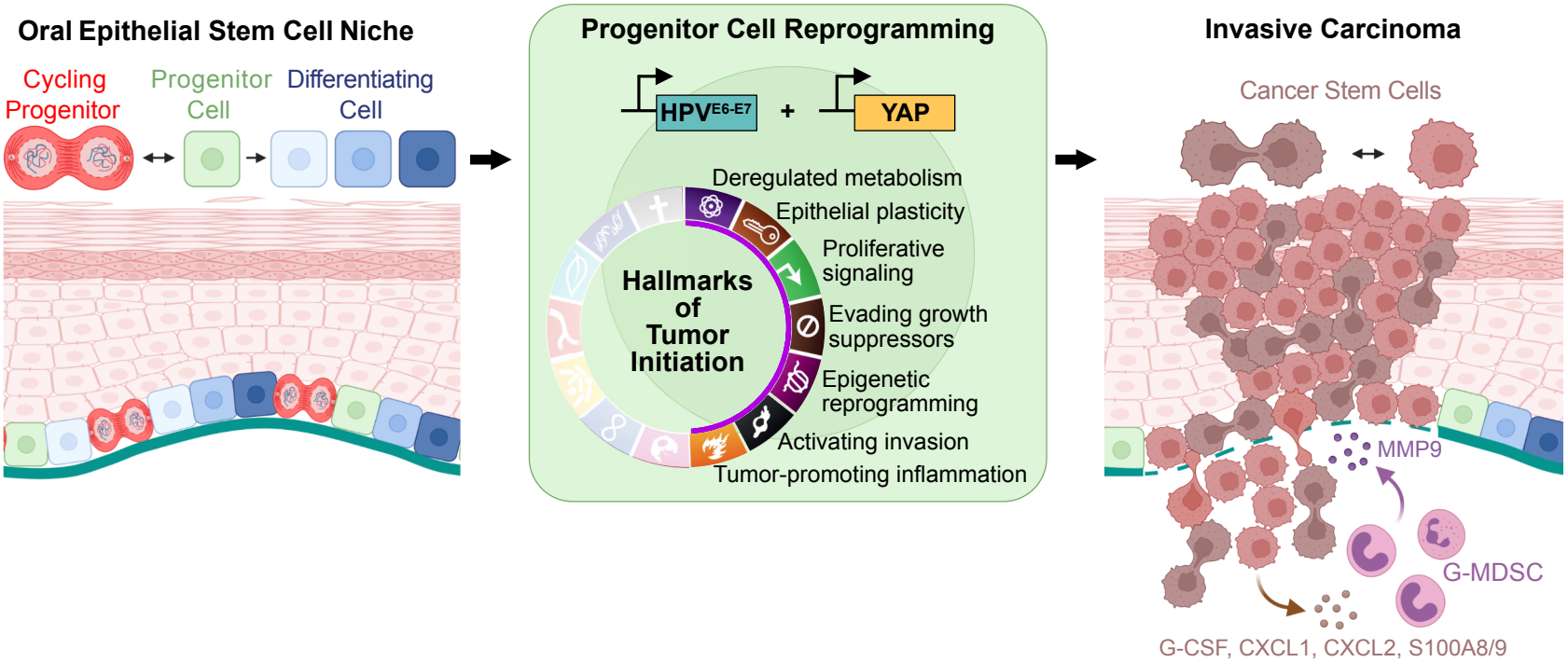

Oral Epithelial Stem Cell Niche
Cycling Progenitor
Progenitor Cell
Differentiating Cell
Progenitor Cell Reprogramming
HPV E6-E7
+
YAP
Hallmarks of Tumor Initiation
Deregulated metabolism
Epithelial plasticity
Proliferative signaling
Evading growth suppressors
Epigenetic reprogramming
Activating invasion
Tumor-promoting inflammation
Invasive Carcinoma
Cancer Stem Cells
MMP9
G-MDSC
G-CSF, CXCL1, CXCL2, S100A8/9

Figure 1

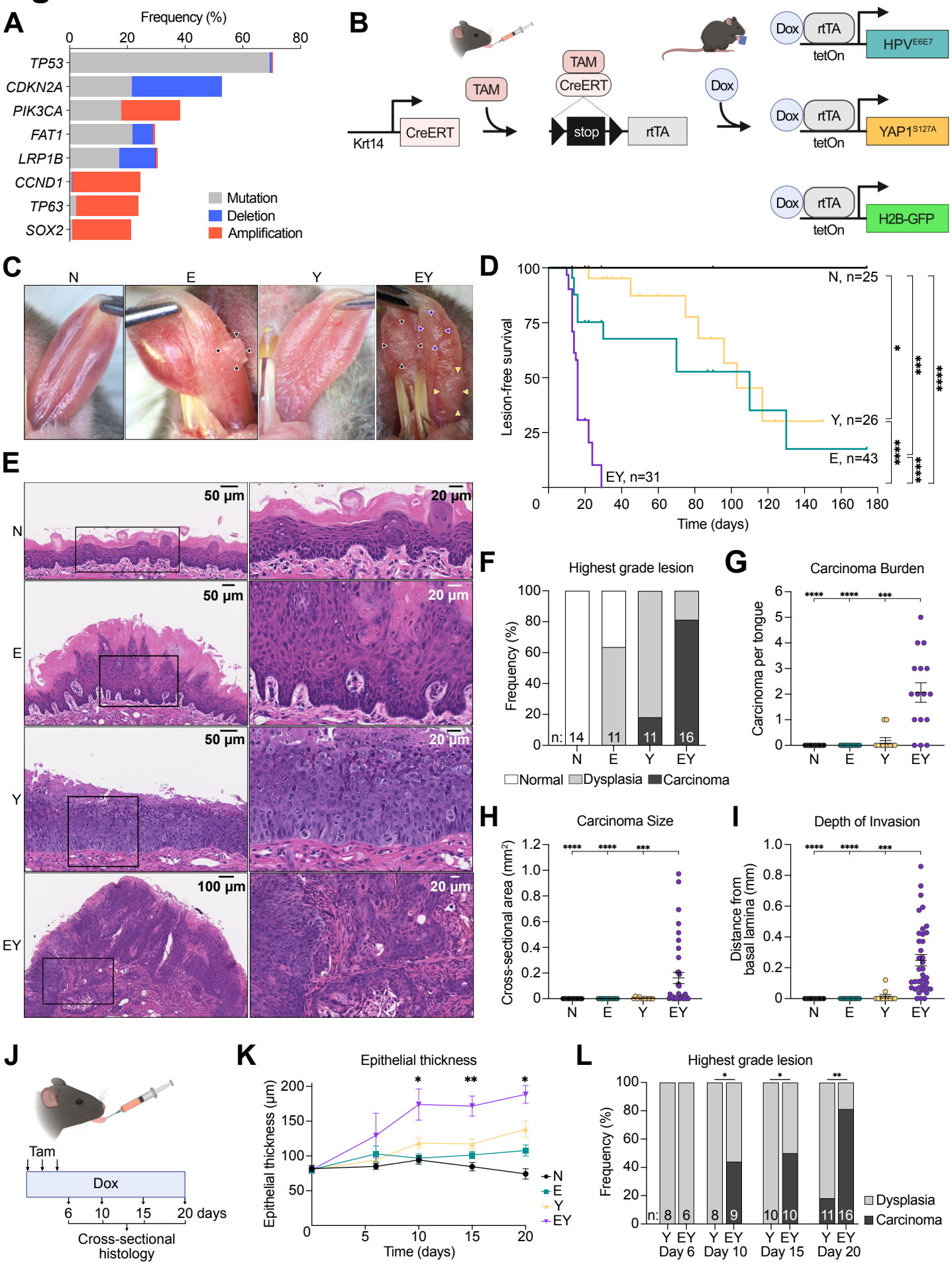

A
Frequency (%)
0
20
40
60
80
TP53
CDKN2A
PIK3CA
FAT1
LRP1B
CCND1
TP63
SOX2
Mutation
Deletion
Amplification
B
Krt14
CreERT
TAM
TAM
CreERT
stop
rtTA
Dox
Dox
rtTA
tetOn
HPV^E6E7
Dox
rtTA
tetOn
YAP1^S127A
Dox
rtTA
tetOn
H2B-GFP
C
N
E
Y
EY
D
Lesion-free survival
Time (days)
N, n=25
Y, n=26
E, n=43
EY, n=31
*
***
****
****
****
E
N
E
Y
EY
50 µm
20 µm
100 µm
F
Highest grade lesion
Frequency (%)
n: 14
11
11
16
N
E
Y
EY
Normal
Dysplasia
Carcinoma
G
Carcinoma Burden
Carcinoma per tongue
****
****
***
H
Carcinoma Size
Cross-sectional area (mm²)
I
Depth of Invasion
Distance from basal lamina (mm)
J
Tam
Dox
6
10
15
20 days
Cross-sectional histology
K
Epithelial thickness
Epithelial thickness (µm)
Time (days)
N
E
Y
EY
L
Highest grade lesion
Frequency (%)
n: 8
6
8
9
10
10
11
16
Y
EY
Day 6
Day 10
Day 15
Day 20
Dysplasia
Carcinoma

Figure 2

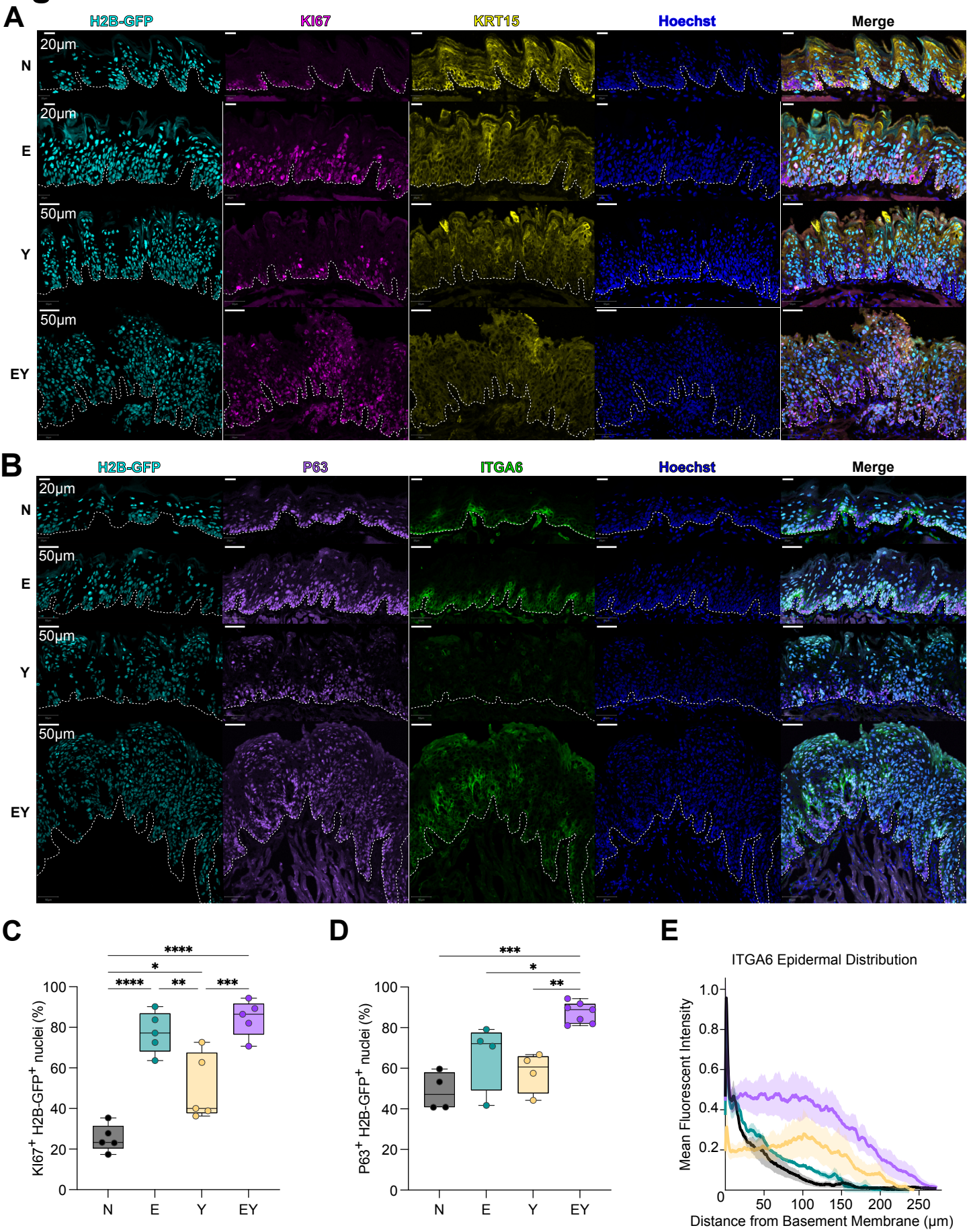

A
H2B-GFP
KI67
KRT15
Hoechst
Merge
N
E
Y
EY
20µm
50µm
B
H2B-GFP
P63
ITGA6
Hoechst
Merge
C
KI67+ H2B-GFP+ nuclei (%)
N
E
Y
EY
D
P63+ H2B-GFP+ nuclei (%)
E
ITGA6 Epidermal Distribution
Mean Fluorescent Intensity
Distance from Basement Membrane (µm)

Figure 3

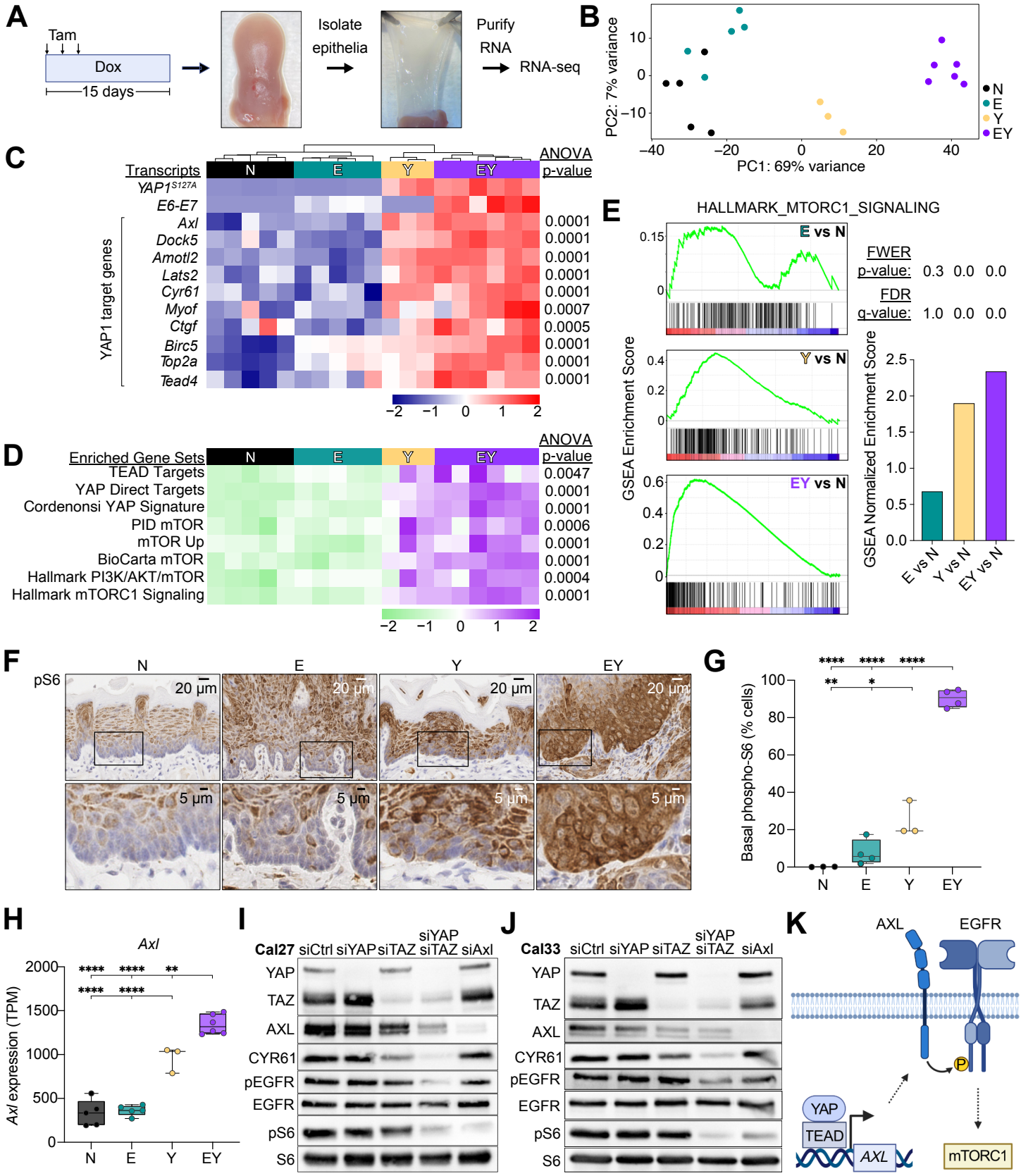

A
Tam
Dox
15 days
Isolate epithelia
Purify RNA
RNA-seq
B
PC2: 7% variance
PC1: 69% variance
N
E
Y
EY
C
Transcripts
N
E
Y
EY
ANOVA p-value
YAP1 target genes
YAP1S127A
E6-E7
Axl 0.0001
Dock5 0.0001
Amotl2 0.0001
Lats2 0.0001
Cyr61 0.0001
Myof 0.0007
Ctgf 0.0005
Birc5 0.0001
Top2a 0.0001
Tead4 0.0001
D
Enriched Gene Sets
ANOVA p-value
TEAD Targets 0.0047
YAP Direct Targets 0.0001
Cordenonsi YAP Signature 0.0001
PID mTOR 0.0006
mTOR Up 0.0001
BioCarta mTOR 0.0001
Hallmark PI3K/AKT/mTOR 0.0004
Hallmark mTORC1 Signaling 0.0001
E
HALLMARK_MTORC1_SIGNALING
E vs N
Y vs N
EY vs N
FWER p-value: 0.3 0.0 0.0
FDR q-value: 1.0 0.0 0.0
GSEA Enrichment Score
GSEA Normalized Enrichment Score
E vs N
Y vs N
EY vs N
F
pS6
N
E
Y
EY
20 µm
5 µm
G
Basal phospho-S6 (% cells)
H
Axl
Axl expression (TPM)
I
Cal27
siCtrl
siYAP
siTAZ
siYAP siTAZ
siAxl
YAP
TAZ
AXL
CYR61
pEGFR
EGFR
pS6
S6
J
Cal33
K
AXL
EGFR
P
YAP
TEAD
AXL
mTORC1

Figure 4

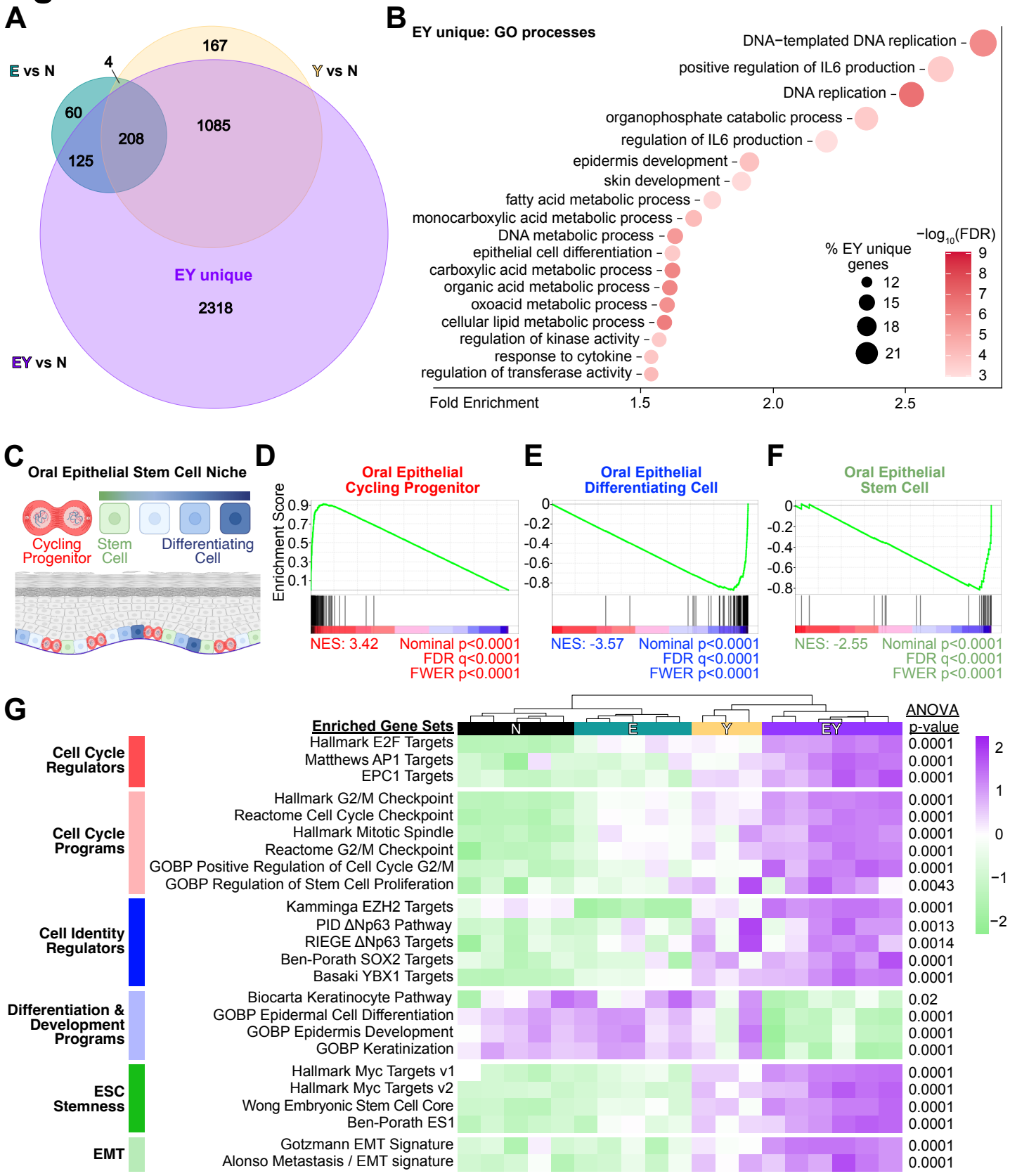

A
E vs N
Y vs N
EY vs N
167
4
60
208
1085
125
EY unique
2318
B
EY unique: GO processes
DNA−templated DNA replication
positive regulation of IL6 production
DNA replication
organophosphate catabolic process
regulation of IL6 production
epidermis development
skin development
fatty acid metabolic process
monocarboxylic acid metabolic process
DNA metabolic process
epithelial cell differentiation
carboxylic acid metabolic process
organic acid metabolic process
oxoacid metabolic process
cellular lipid metabolic process
regulation of kinase activity
response to cytokine
regulation of transferase activity
Fold Enrichment
1.5
2.0
2.5
% EY unique genes
12
15
18
21
−log10(FDR)
9
8
7
6
5
4
3
C
Oral Epithelial Stem Cell Niche
Cycling Progenitor
Stem Cell
Differentiating Cell
D
Oral Epithelial Cycling Progenitor
Enrichment Score
NES: 3.42
Nominal p<0.0001
FDR q<0.0001
FWER p<0.0001
E
Oral Epithelial Differentiating Cell
NES: -3.57
Nominal p<0.0001
FDR q<0.0001
FWER p<0.0001
F
Oral Epithelial Stem Cell
NES: -2.55
Nominal p<0.0001
FDR q<0.0001
FWER p<0.0001
G
Enriched Gene Sets
N
E
Y
EY
ANOVA p-value
Cell Cycle Regulators
Hallmark E2F Targets 0.0001
Matthews AP1 Targets 0.0001
EPC1 Targets 0.0001
Cell Cycle Programs
Hallmark G2/M Checkpoint 0.0001
Reactome Cell Cycle Checkpoint 0.0001
Hallmark Mitotic Spindle 0.0001
Reactome G2/M Checkpoint 0.0001
GOBP Positive Regulation of Cell Cycle G2/M 0.0001
GOBP Regulation of Stem Cell Proliferation 0.0043
Cell Identity Regulators
Kamminga EZH2 Targets 0.0001
PID ΔNp63 Pathway 0.0013
RIEGE ΔNp63 Targets 0.0014
Ben-Porath SOX2 Targets 0.0001
Basaki YBX1 Targets 0.0001
Differentiation & Development Programs
Biocarta Keratinocyte Pathway 0.02
GOBP Epidermal Cell Differentiation 0.0001
GOBP Epidermis Development 0.0001
GOBP Keratinization 0.0001
ESC Stemness
Hallmark Myc Targets v1 0.0001
Hallmark Myc Targets v2 0.0001
Wong Embryonic Stem Cell Core 0.0001
Ben-Porath ES1 0.0001
EMT
Gotzmann EMT Signature 0.0001
Alonso Metastasis / EMT signature 0.0001
2
1
0
−1
−2

Figure 5

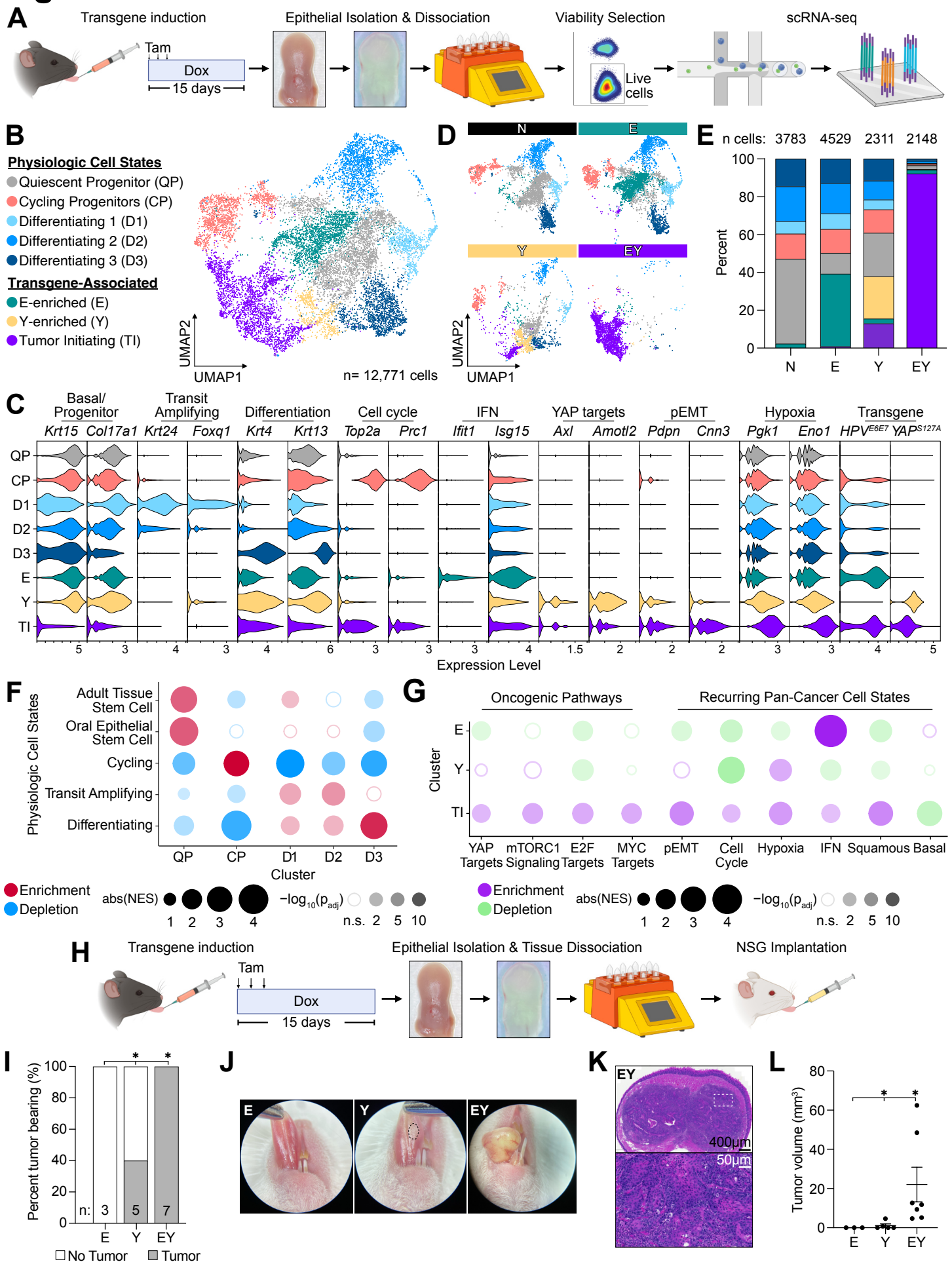

A
Transgene induction
Tam
Dox
15 days
Epithelial Isolation & Dissociation
Viability Selection
Live cells
scRNA-seq
B
Physiologic Cell States
Quiescent Progenitor (QP)
Cycling Progenitors (CP)
Differentiating 1 (D1)
Differentiating 2 (D2)
Differentiating 3 (D3)
Transgene-Associated
E-enriched (E)
Y-enriched (Y)
Tumor Initiating (TI)
UMAP2
UMAP1
n= 12,771 cells
D
N
E
Y
EY
E
n cells: 3783 4529 2311 2148
Percent
C
Basal/ Progenitor
Transit Amplifying
Differentiation
Cell cycle
IFN
YAP targets
pEMT
Hypoxia
Transgene
Krt15 Col17a1 Krt24 Foxq1 Krt4 Krt13 Top2a Prc1 Ifit1 Isg15 Axl Amotl2 Pdpn Cnn3 Pgk1 Eno1 HPV E6E7 YAP S127A
QP
CP
D1
D2
D3
E
Y
TI
Expression Level
F
Physiologic Cell States
Adult Tissue Stem Cell
Oral Epithelial Stem Cell
Cycling
Transit Amplifying
Differentiating
QP CP D1 D2 D3
Cluster
Enrichment
Depletion
abs(NES)
1 2 3 4
−log10(padj)
n.s. 2 5 10
G
Oncogenic Pathways
Recurring Pan-Cancer Cell States
Cluster
YAP Targets
mTORC1 Signaling
E2F Targets
MYC Targets
pEMT
Cell Cycle
Hypoxia
IFN
Squamous
Basal
H
Transgene induction
Tam
Dox
15 days
Epithelial Isolation & Tissue Dissociation
NSG Implantation
I
Percent tumor bearing (%)
n: 3 5 7
No Tumor
Tumor
J
K
400μm
50μm
L
Tumor volume (mm³)

Figure 6

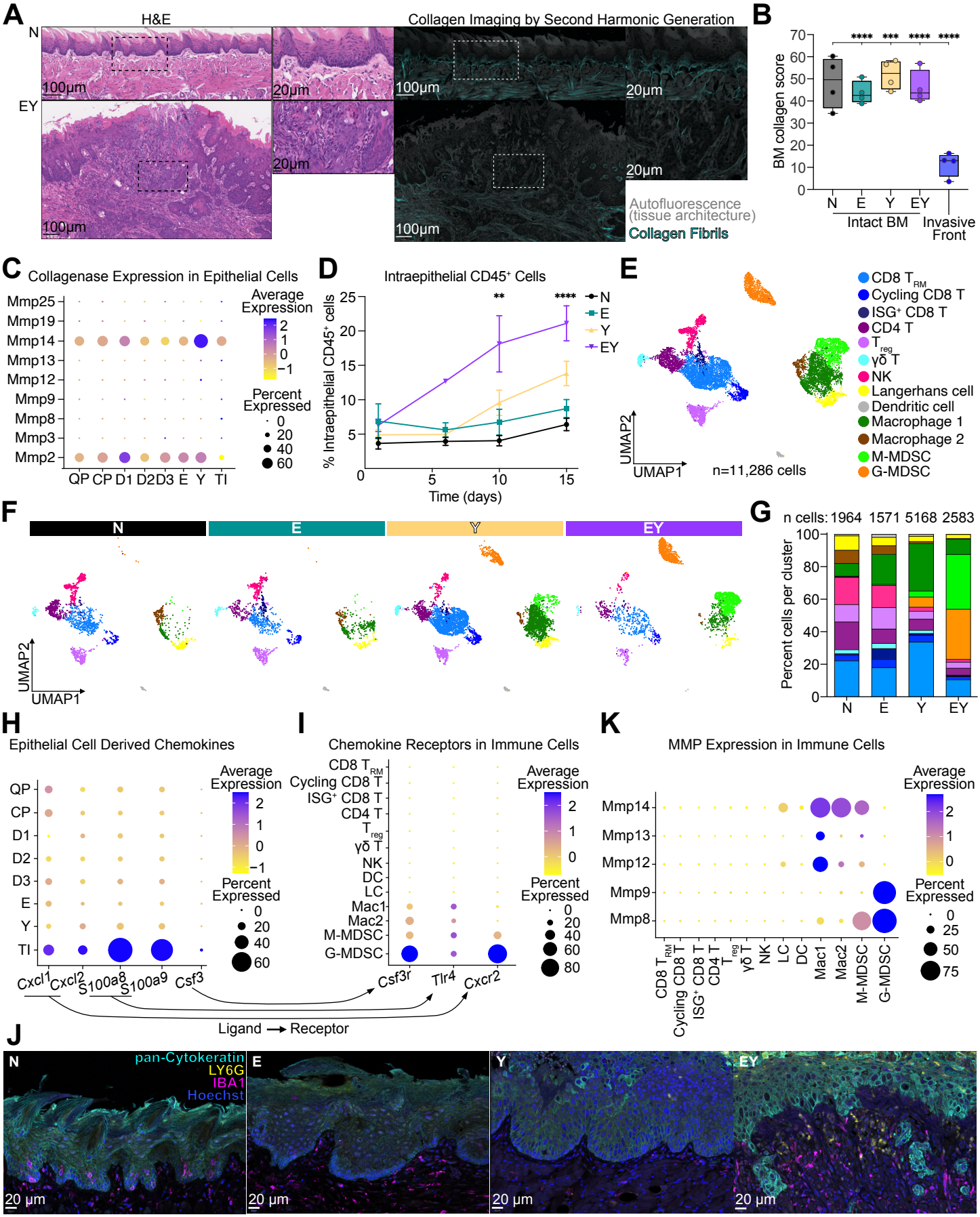
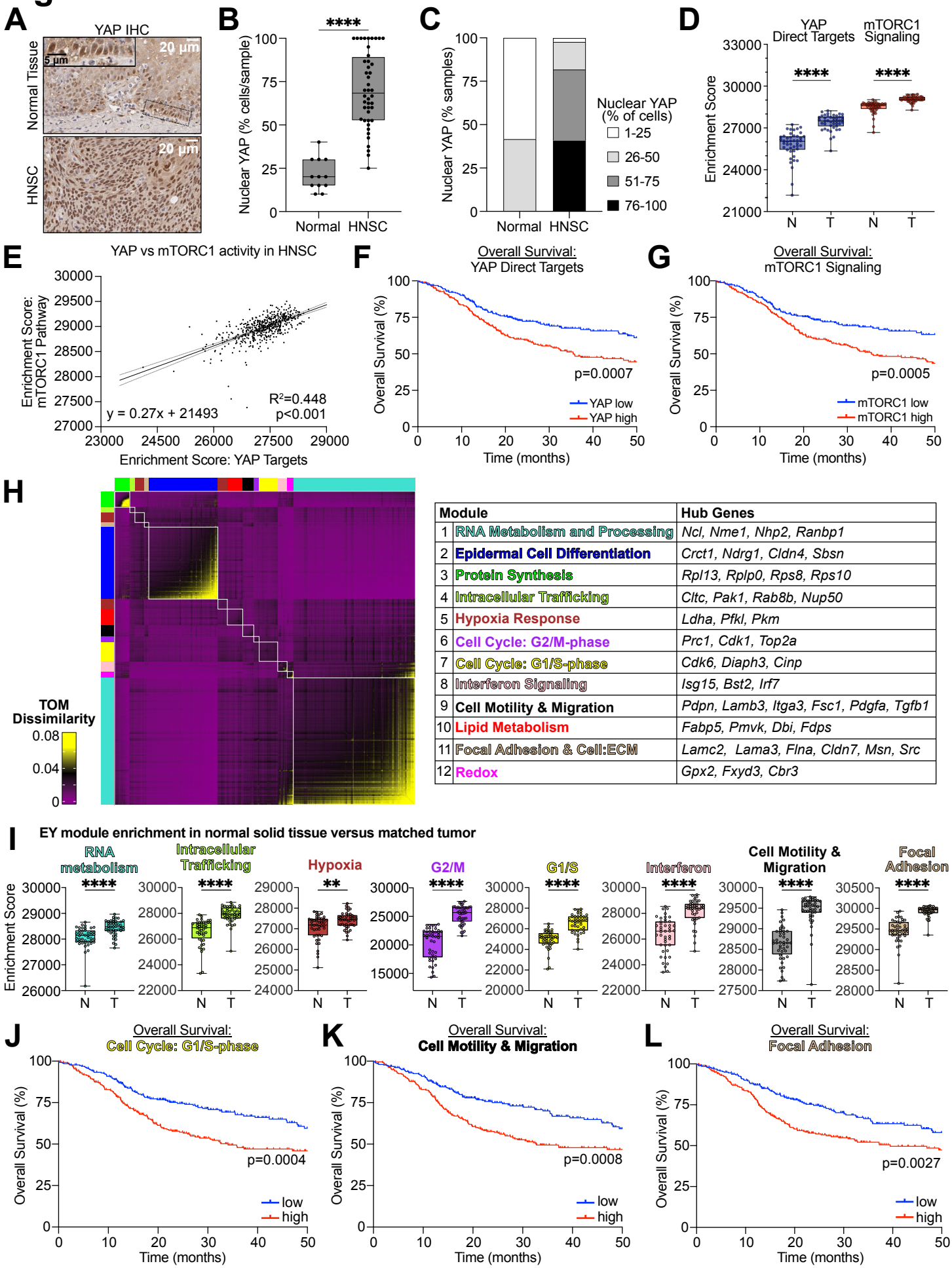

A
H&E
Collagen Imaging by Second Harmonic Generation
N
EY
100µm
20µm
Autofluorescence (tissue architecture)
Collagen Fibrils
B
BM collagen score
N E Y EY
Intact BM
Invasive Front
C
Collagenase Expression in Epithelial Cells
Mmp25 Mmp19 Mmp14 Mmp13 Mmp12 Mmp9 Mmp8 Mmp3 Mmp2
QP CP D1 D2 D3 E Y TI
Average Expression
Percent Expressed
D
Intraepithelial CD45⁺ Cells
% Intraepithelial CD45⁺ cells
Time (days)
N E Y EY
E
CD8 T_RM
Cycling CD8 T
ISG⁺ CD8 T
CD4 T
T_reg
γδ T
NK
Langerhans cell
Dendritic cell
Macrophage 1
Macrophage 2
M-MDSC
G-MDSC
UMAP1
UMAP2
n=11,286 cells
F
N E Y EY
G
n cells: 1964 1571 5168 2583
Percent cells per cluster
H
Epithelial Cell Derived Chemokines
QP CP D1 D2 D3 E Y TI
Cxcl1 Cxcl2 S100a8 S100a9 Csf3
I
Chemokine Receptors in Immune Cells
CD8 T_RM
Cycling CD8 T
ISG⁺ CD8 T
CD4 T
T_reg
γδ T
NK
DC
LC
Mac1
Mac2
M-MDSC
G-MDSC
Csf3r Tlr4 Cxcr2
Ligand → Receptor
K
MMP Expression in Immune Cells
Mmp14 Mmp13 Mmp12 Mmp9 Mmp8
CD8 T_RM Cycling CD8 T ISG⁺ CD8 T CD4 T T_reg γδ T NK LC DC Mac1 Mac2 M-MDSC G-MDSC
J
N E Y EY
pan-Cytokeratin
LY6G
IBA1
Hoechst
20 µm

Figure 7



	Module	Hub Genes
1	RNA Metabolism and Processing	*Ncl, Nme1, Nhp2, Ranbp1*
2	Epidermal Cell Differentiation	*Crct1, Ndrg1, Cldn4, Sbsn*
3	Protein Synthesis	*Rpl13, Rplp0, Rps8, Rps10*
4	Intracellular Trafficking	*Cltc, Pak1, Rab8b, Nup50*
5	Hypoxia Response	*Ldha, Pfkl, Pkm*
6	Cell Cycle: G2/M-phase	*Prc1, Cdk1, Top2a*
7	Cell Cycle: G1/S-phase	*Cdk6, Diaph3, Cinp*
8	Interferon Signaling	*Isg15, Bst2, Irf7*
9	Cell Motility & Migration	*Pdpn, Lamb3, Itga3, Fsc1, Pdgfa, Tgfb1*
10	Lipid Metabolism	*Fabp5, Pmvk, Dbi, Fdps*
11	Focal Adhesion & Cell:ECM	*Lamc2, Lama3, Flna, Cldn7, Msn, Src*
12	Redox	*Gpx2, Fxyd3, Cbr3*

Supplementary Figure 1, related to Figure 1

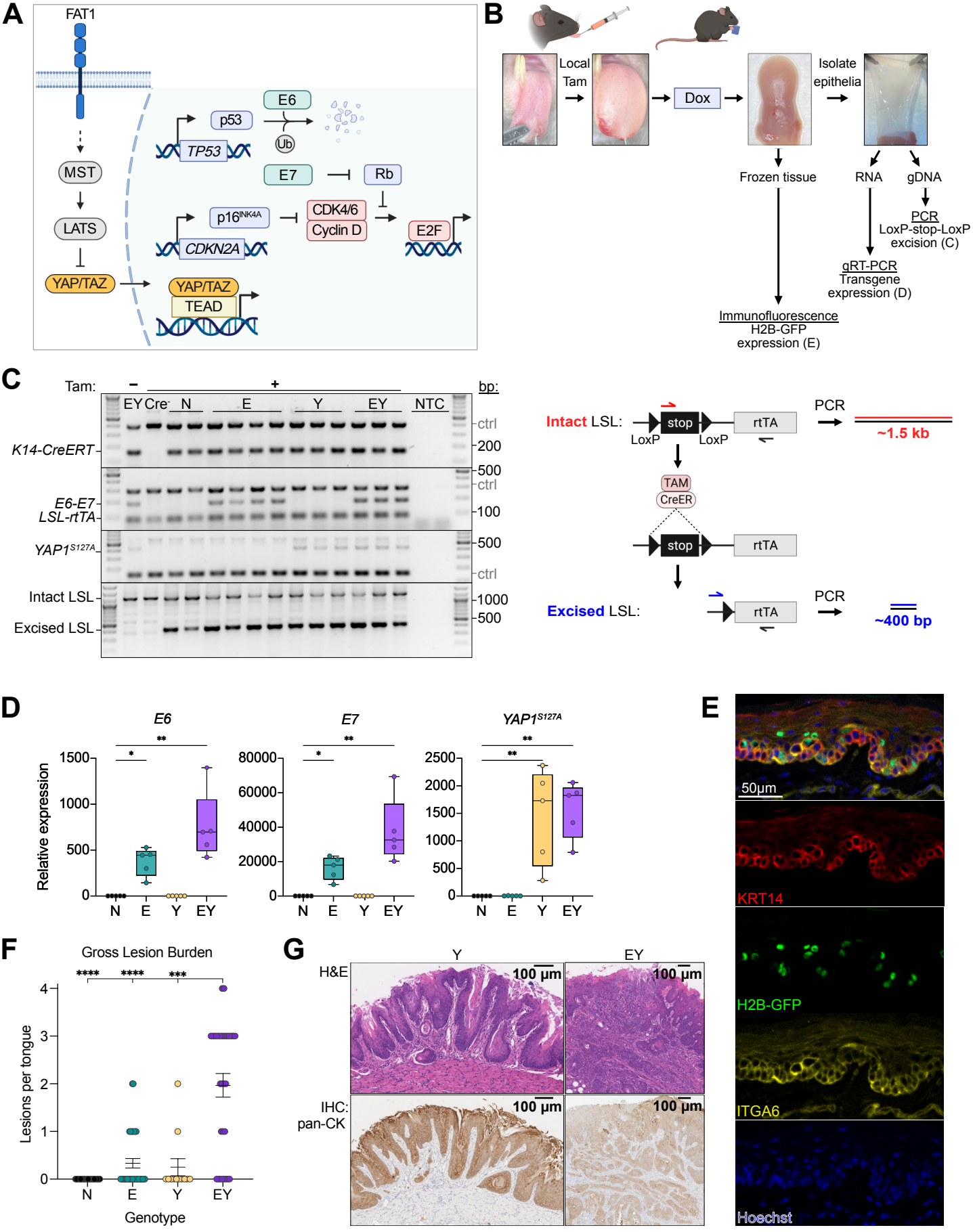


Figure S2, related to Figure 2

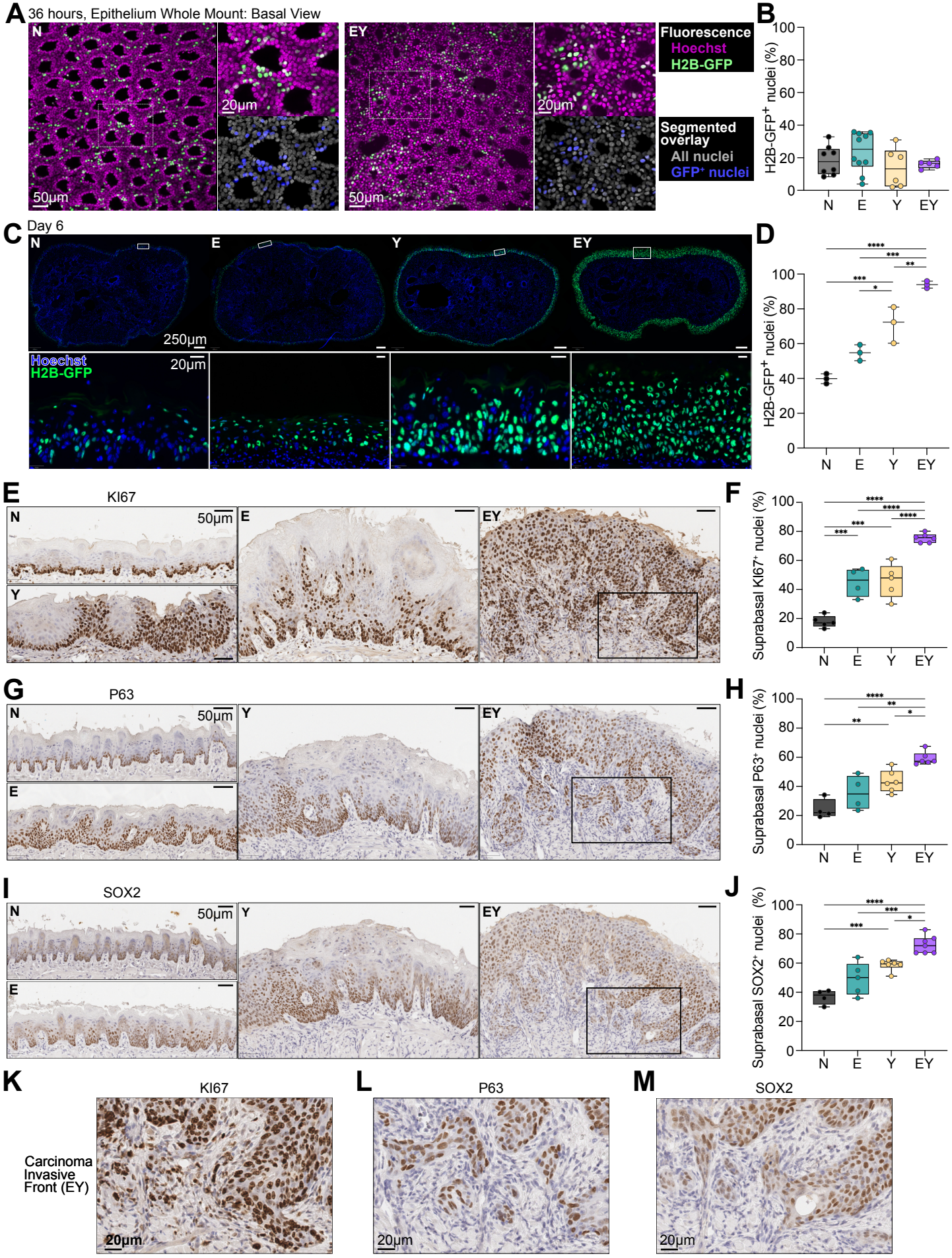


Figure S3, related to Figure 3

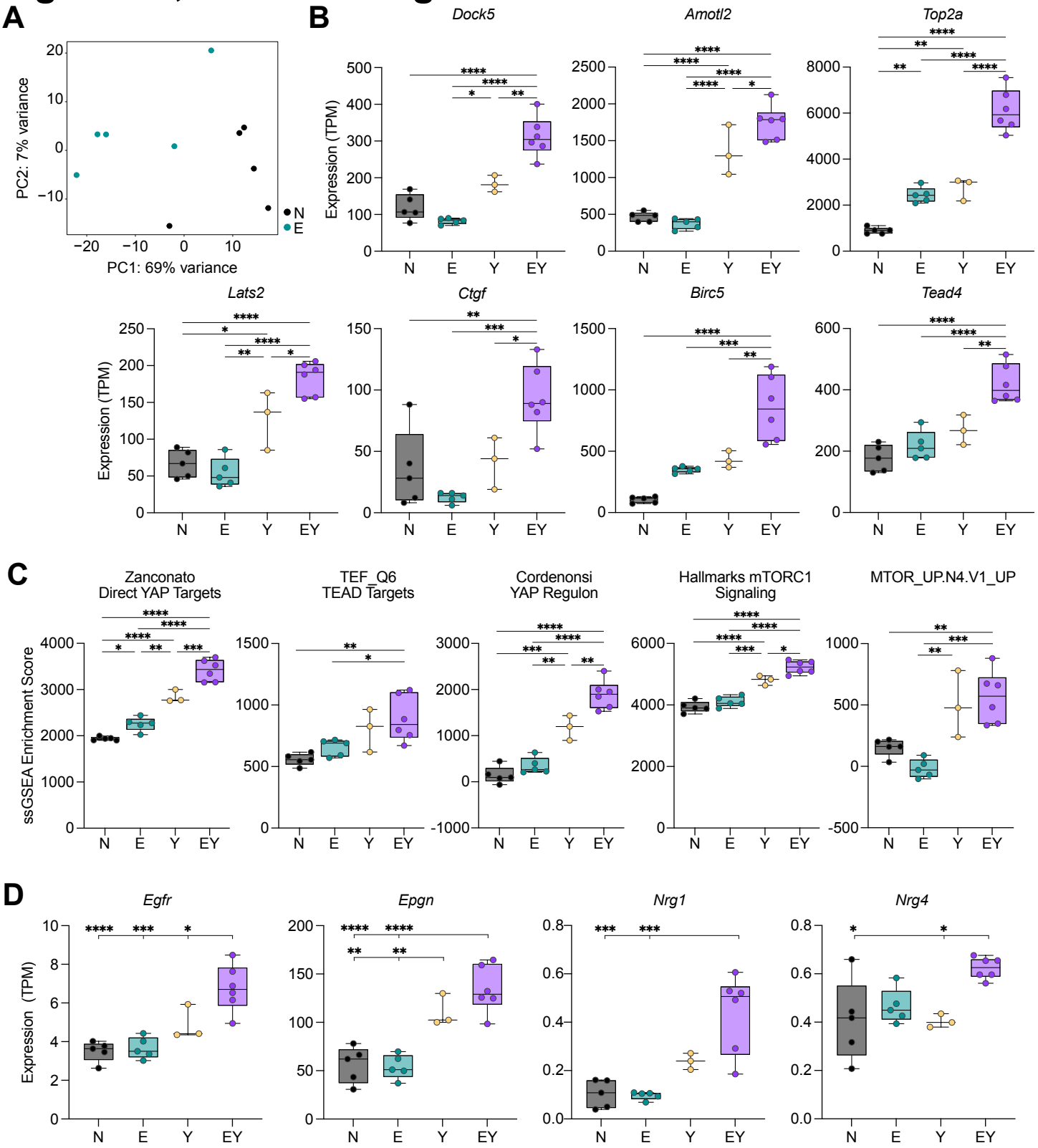


Figure S4, related to Figure 4

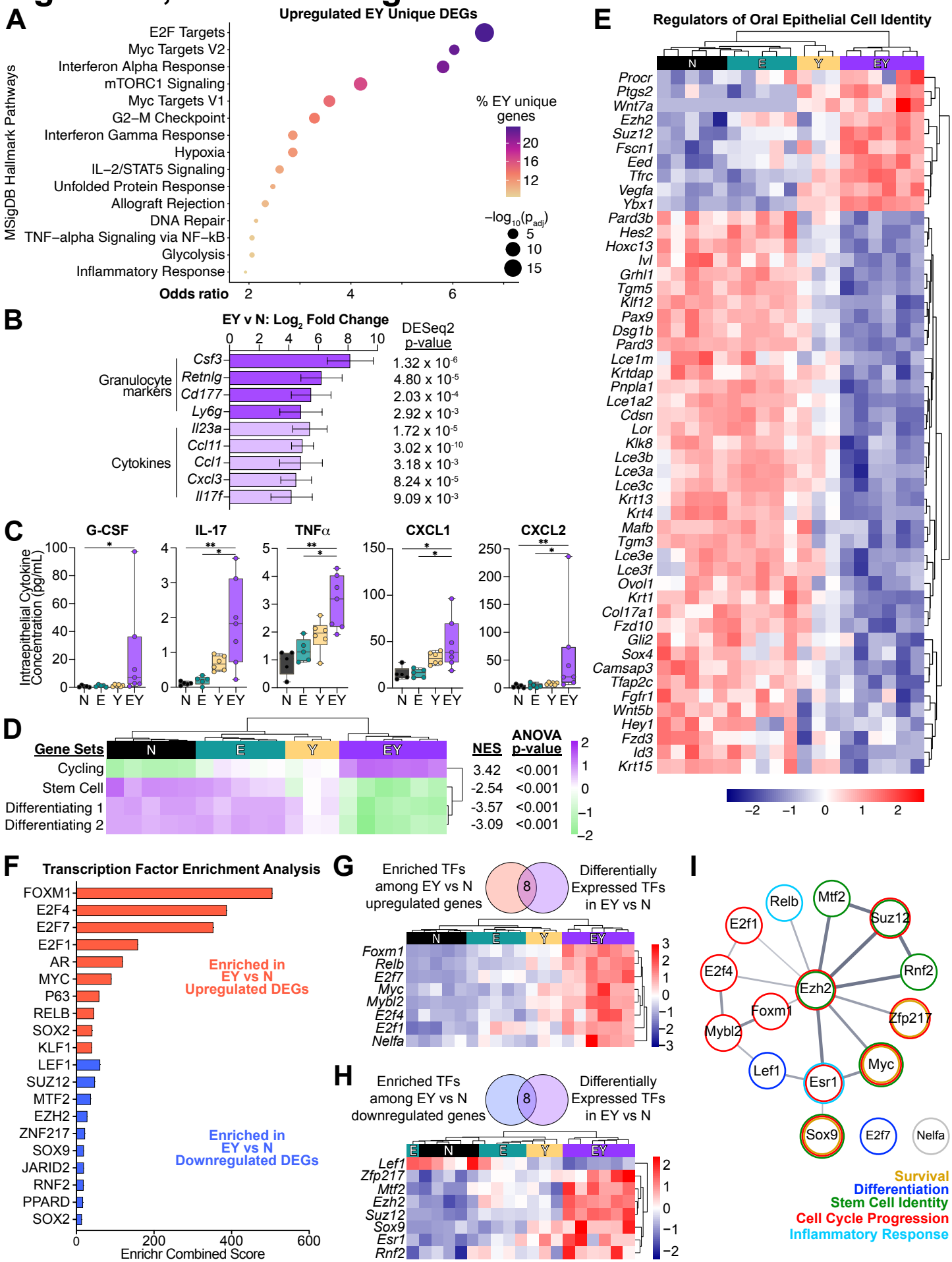


Figure S5, related to Figure 5

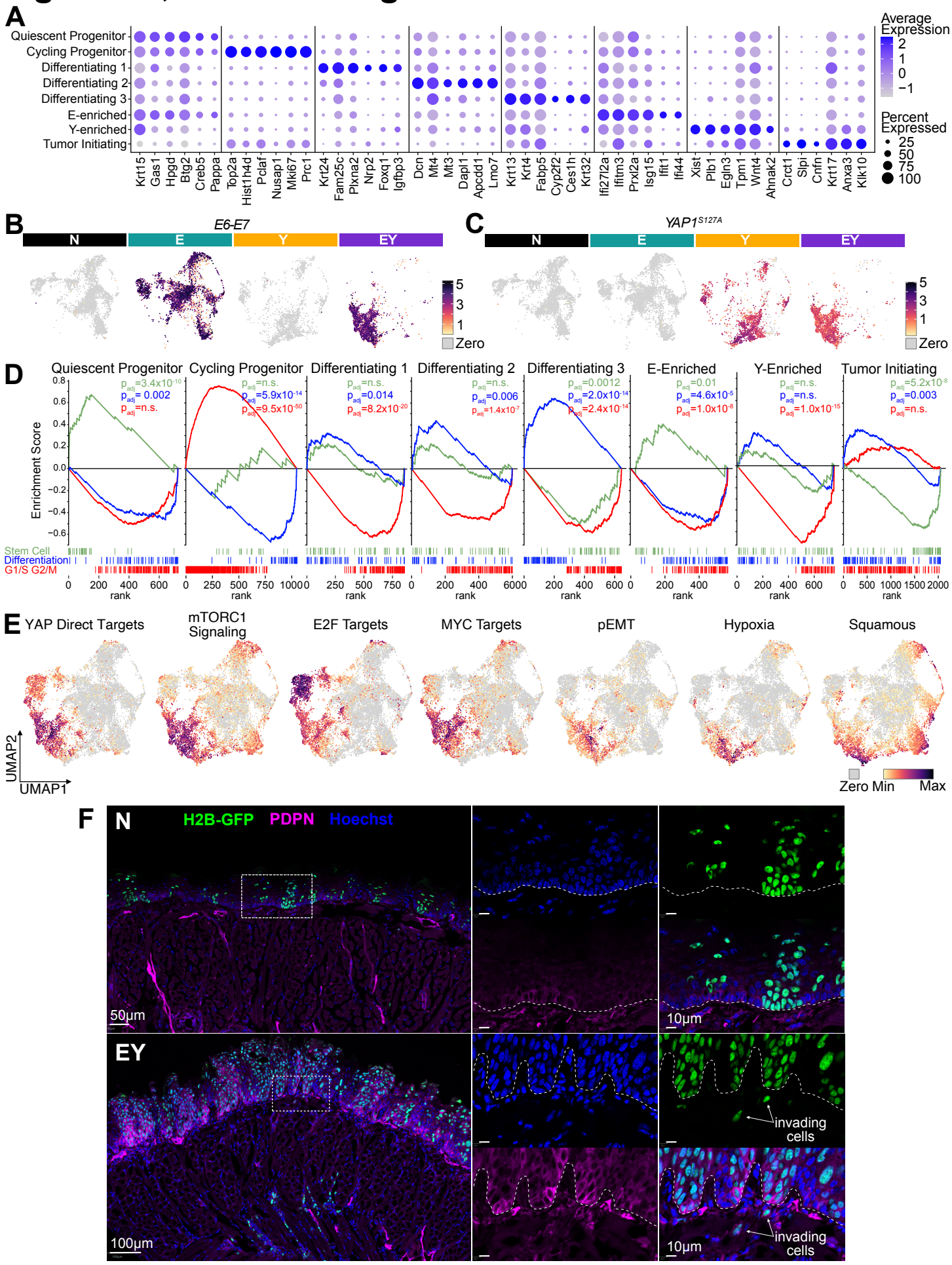


Figure S6, related to Figure 6

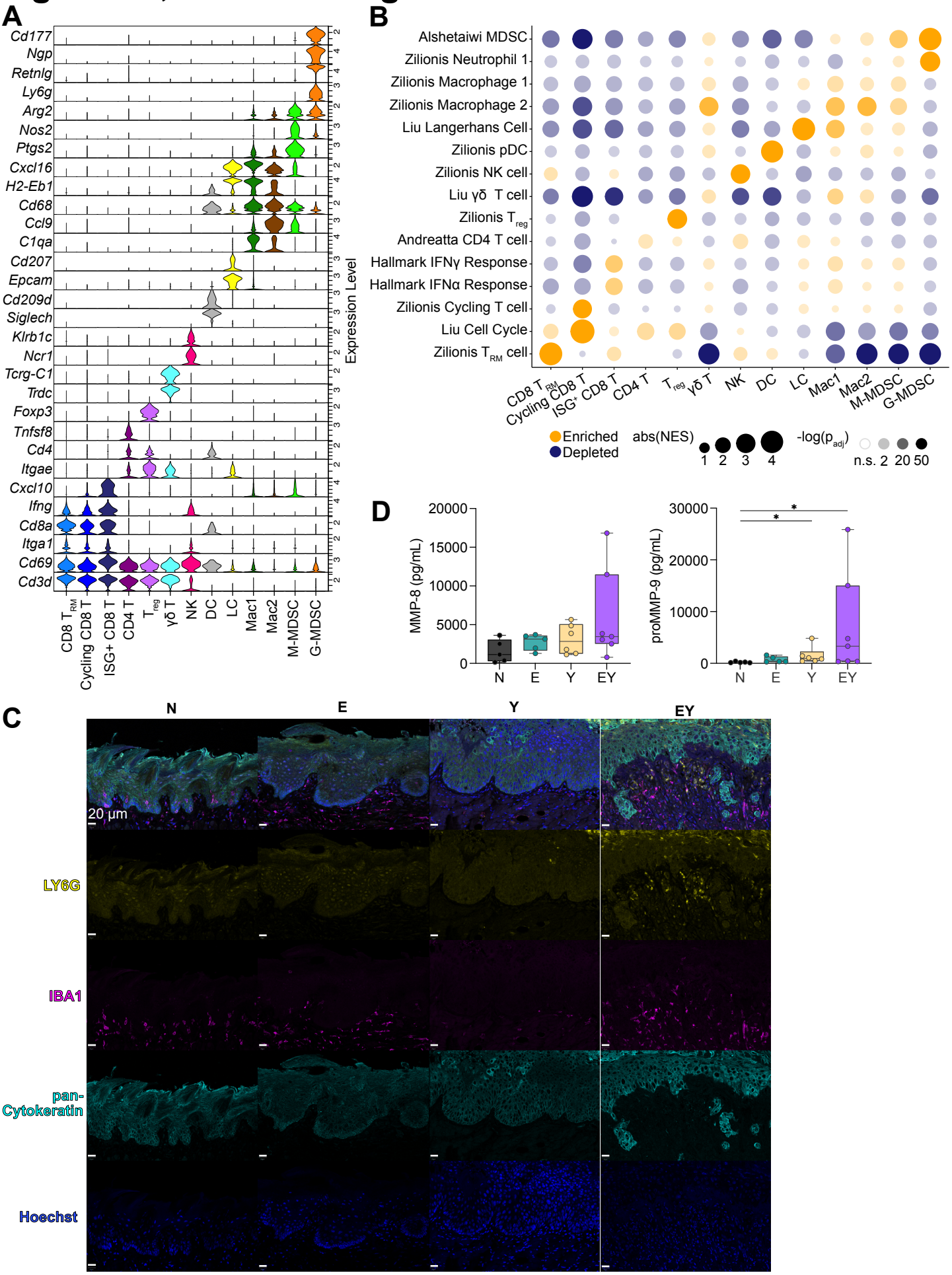
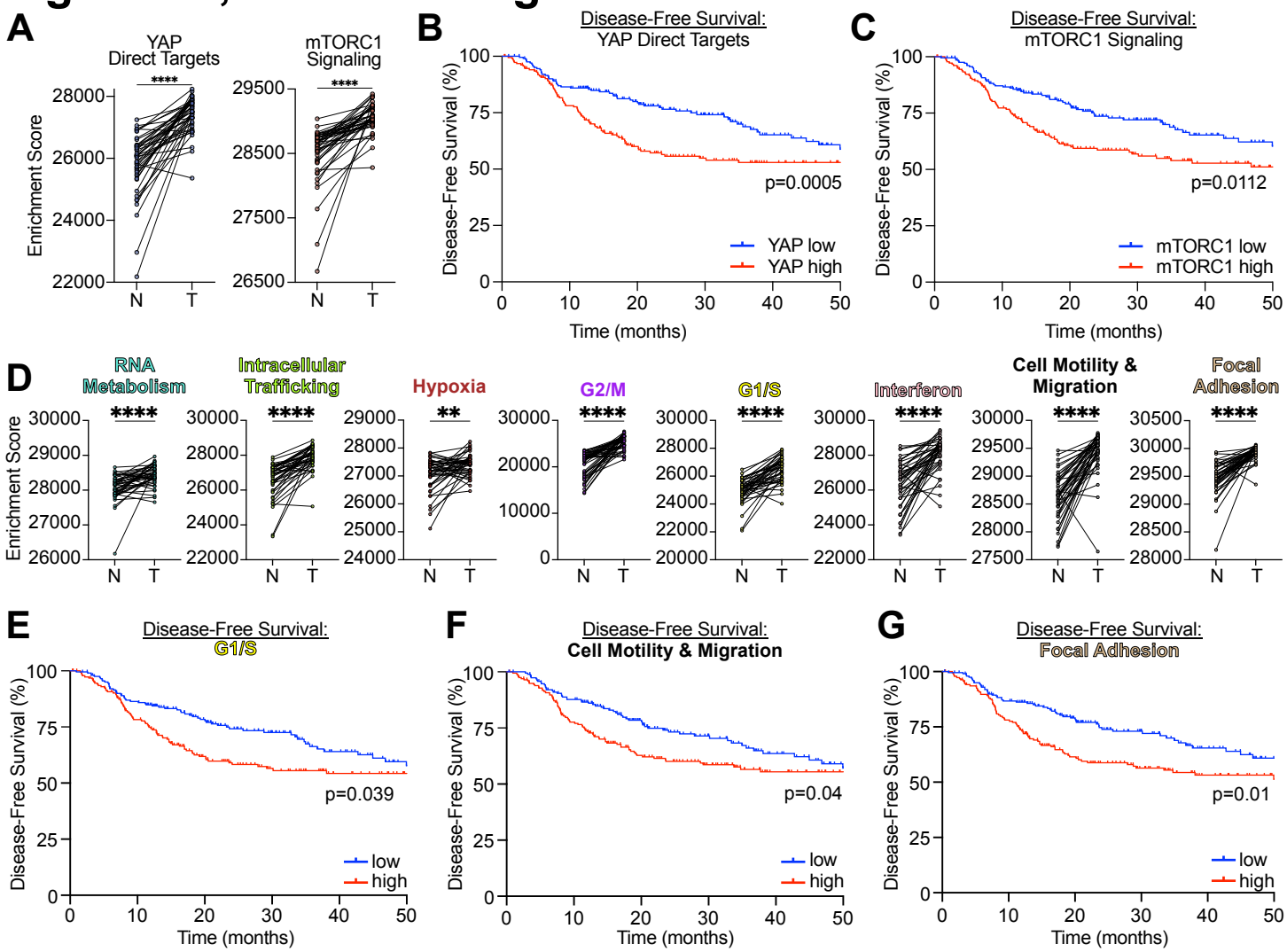



Figure S7, related to Figure 7