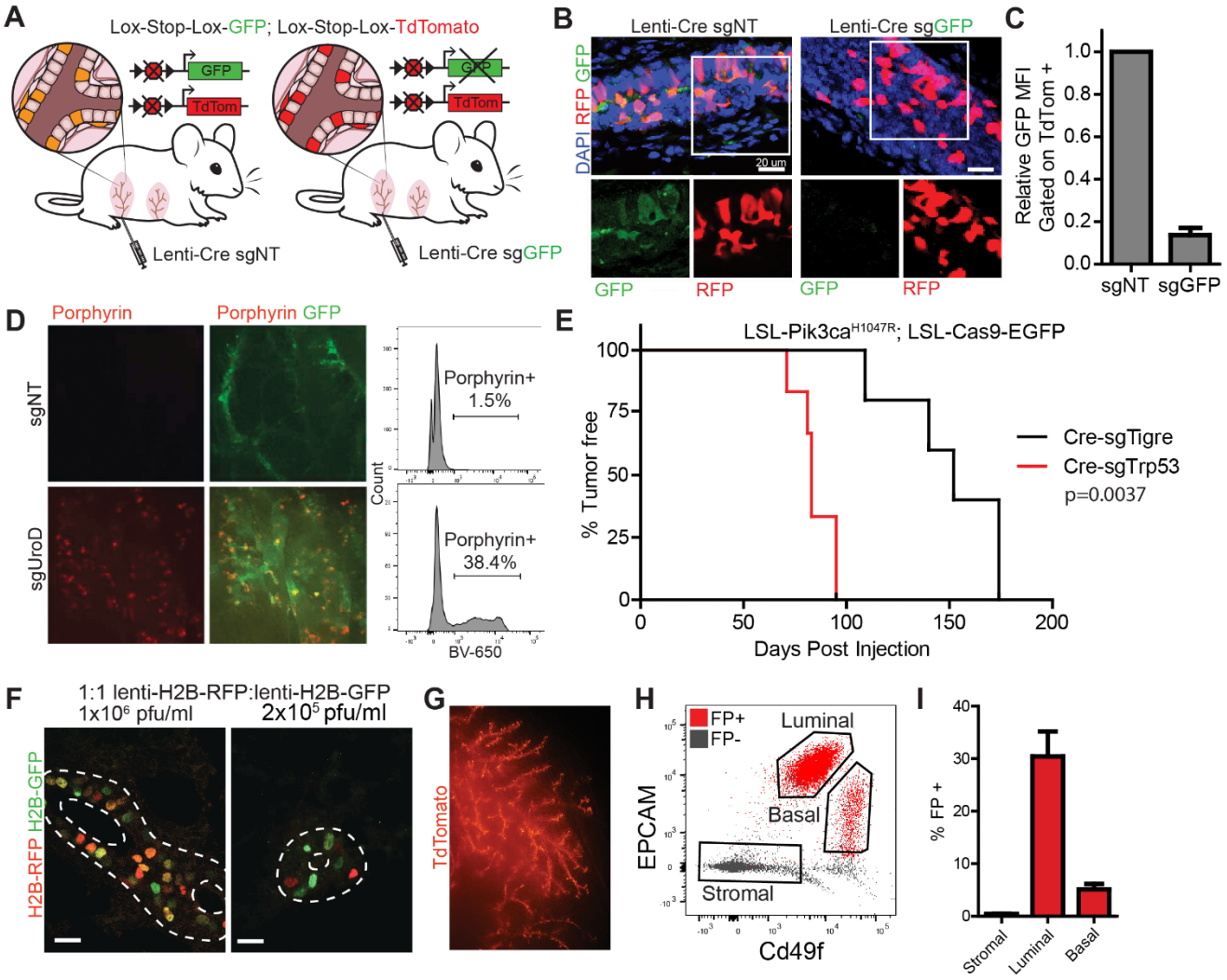
Supplementary Figures

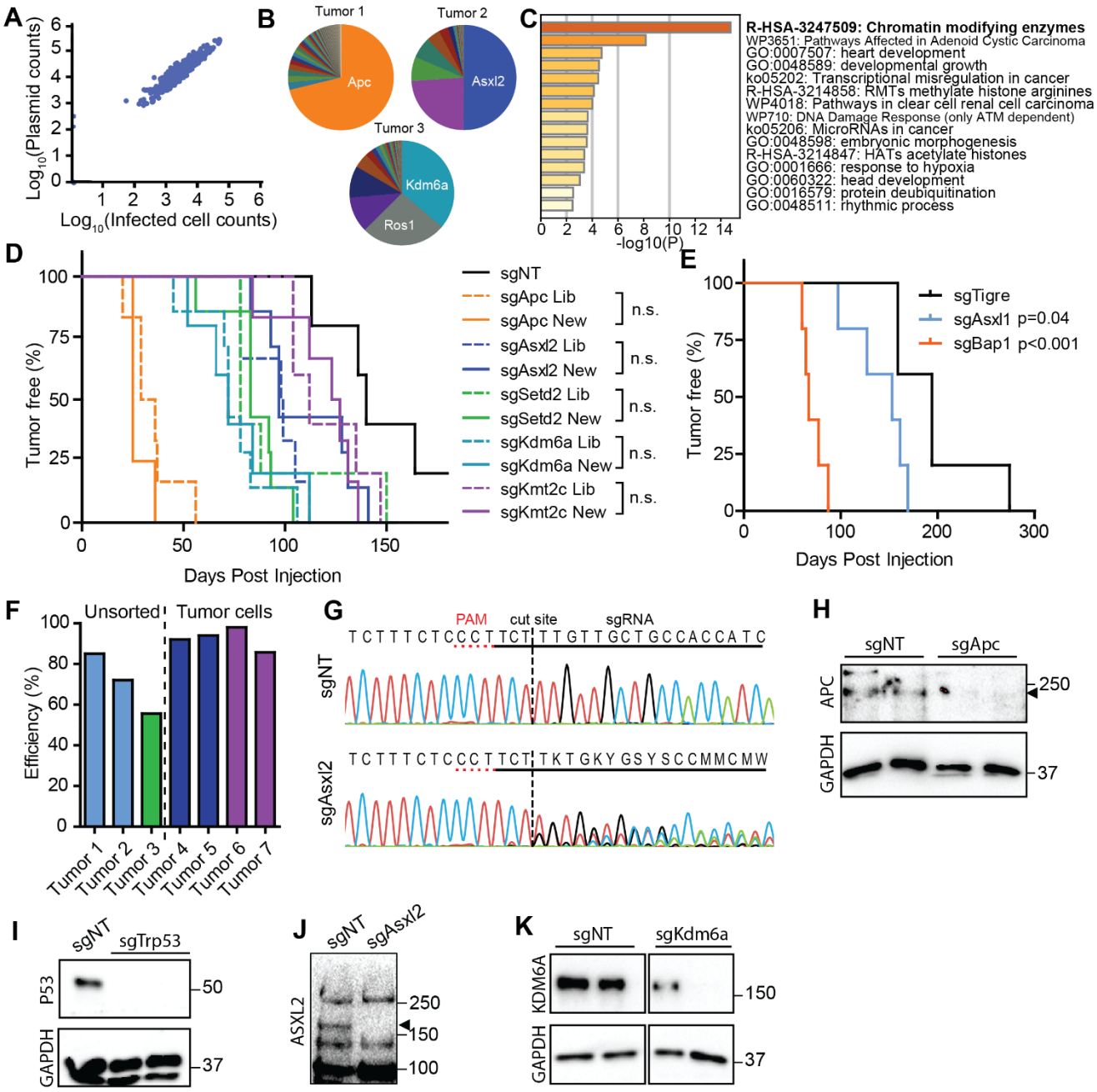
Supplementary Fig. S1



Supplementary Fig. S1. Direct *in vivo* CRISPR knock-out efficiency in mouse mammary epithelium.

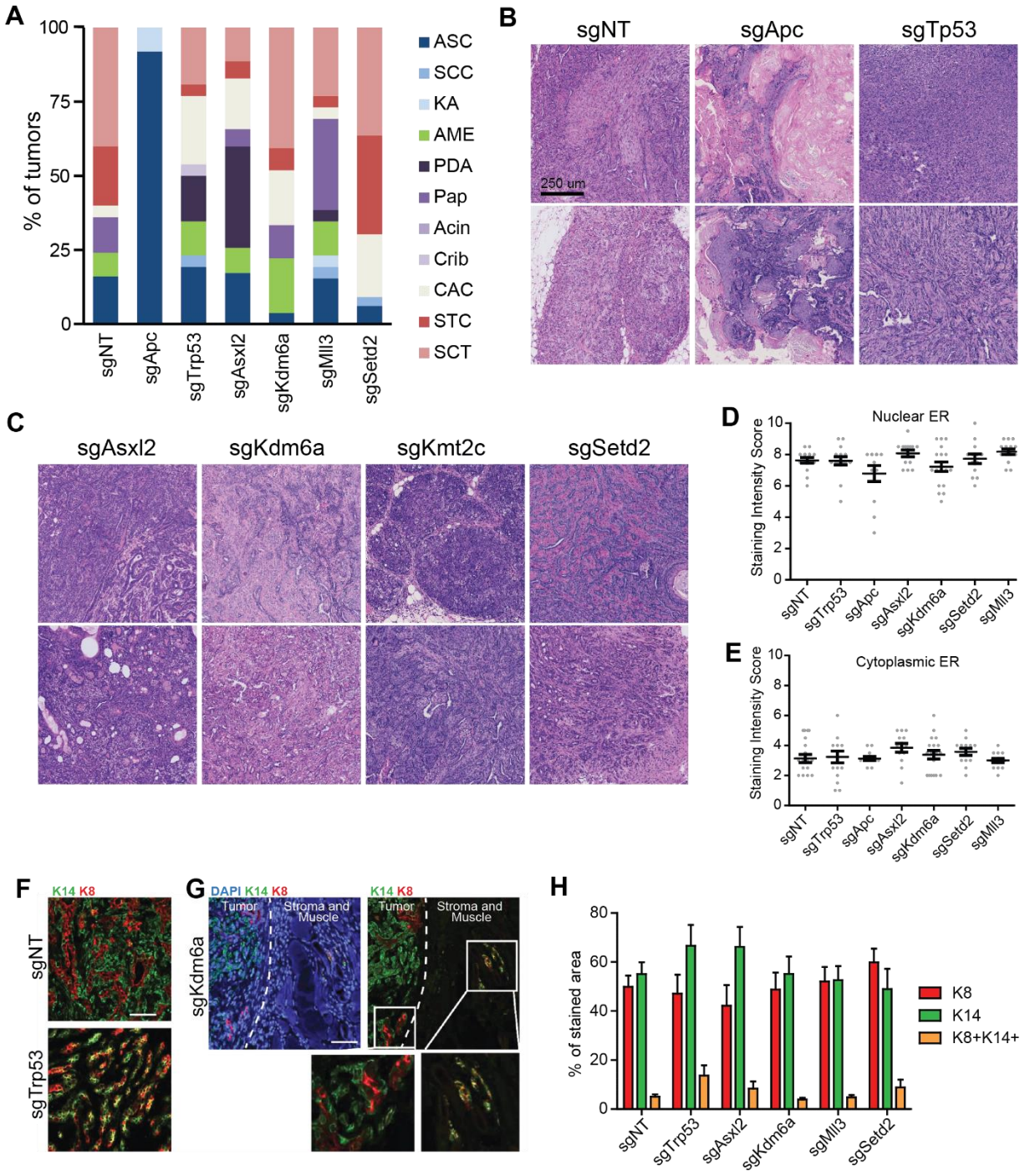
A, Experimental schematic showing genetic ablation of green fluorescent protein (GFP) from double reporter R26-LSL-Cas9-GFP^{het}; R26-LSL-tdTomato^{het} mice transduced with an sgGFP-Cre lentivirus that targets GFP or non-targeting control lentivirus (sgNT). **B**, Representative images showing GFP and tdTomato expression in R26-LSL-Cas9-GFP^{het}; R26-LSL-tdTomato^{het} mammary epithelium transduced with sgGFP-Cre lentivirus or control non-targeting sgNT-Cre lentivirus. **C,** Flow cytometry analysis showing relative Mean Fluorescent Intensity (MFI) of GFP in TdTomato+ cells for sgGFP infected glands relative to sgNT. **D**, Representative whole mount images and FACS plots depicting red fluorescence in the mammary epithelium of *Urod* mosaic knock-out mice indicative of efficient, bi-allelic mutagenesis of *UroD* gene. **E**, Tumor-free survival for *Pik3ca*^{H1047R};Cas9 mice transduced with lentiviral sgRNAs targeting the *Trp53* gene or the Tigre 'safe harbor' locus as control. (n≥5 per group). **F**, Representative images of mammary epithelium transduced with a Lenti-GFP/Lenti-RFP mixture showing cells transduced with GFP or RFP or double-infected expressing GFP+/RFP+ cells. More double positive cells are observed at higher viral titre. **G**, Whole mount image of a mammary gland transduced with Lenti-Cre in a R26-LSL-tdTomato^{het} mouse. **H and I,** Representative FACS plot (**H**) and quantification (**I**) of fluorescent protein (FP)+ LV-infected cells.

Supplementary Fig. S2



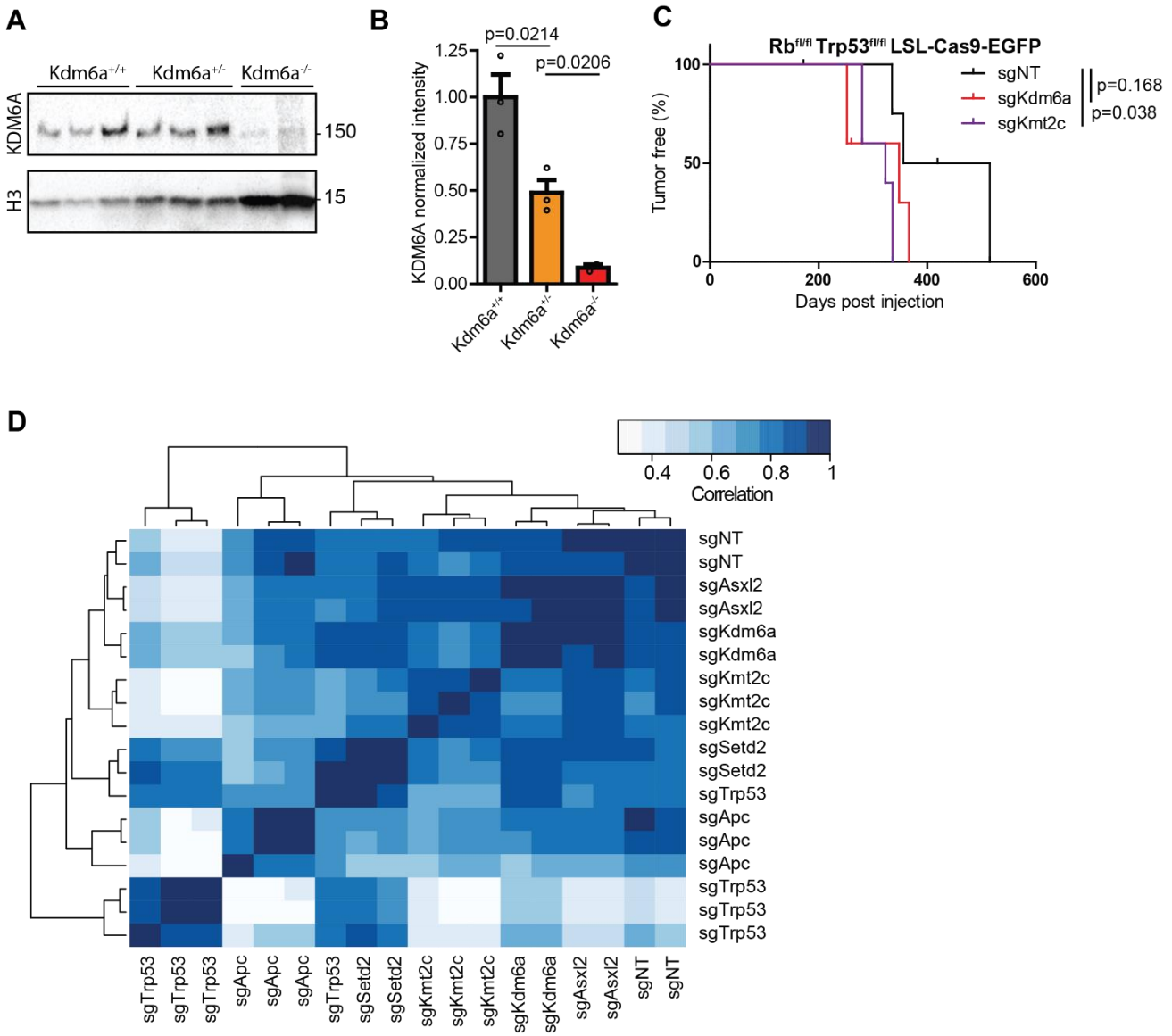
Supplementary Fig. S2. Library screening and hit validation. A, Graph showing sgRNA representation from breast cancer libraries in plasmid DNA versus DNA from infected MEFs. Each dot represents a guide. Representation of all genes is maintained with high correlation in abundance. **B,** Representative pie charts showing tumor suppressor genes with enriched sgRNAs in tumor DNA obtained from three different mammary gland tumors. **C,** Metascape pathway enrichment analysis within screen hits. **D,** Tumor-free survival of Pik3ca^{H1047R} transduced individually with sgRNAs targeting the indicated genes. sgRNAs from the original library screen ('Lib') or independently designed sgRNAs ('New') were used. No significant difference in tumor-free survival was found between pairs of sgRNAs targeting the same gene. **E,** Tumor-free survival of Pik3ca^{H1047R} transduced individually with sgRNAs targeting *Bap1*, *Asxl1* or the inert Tigre locus as control (n=5, p<0.05). **F,** Gene editing efficiency determined using sanger-sequencing data of PCR-amplified sgRNA target sites followed by Tracking of Indels by Decomposition (TIDE) on bulk or GFP+ sorted tumor cells. Colour denotes sgRNA as shown in (**C**). **G,** Representative sanger sequencing plots showing discordance of DNA from a sgAsxl2-targeted sample compared to a wildtype sample. **H-K,** Western blot analysis showing loss of protein expression after CRISPR-mediated knockout of Apc (**H**), p53 (**I**), Asxl2 (**J**) and Kdm6a (**K**).

Supplementary Fig. S3



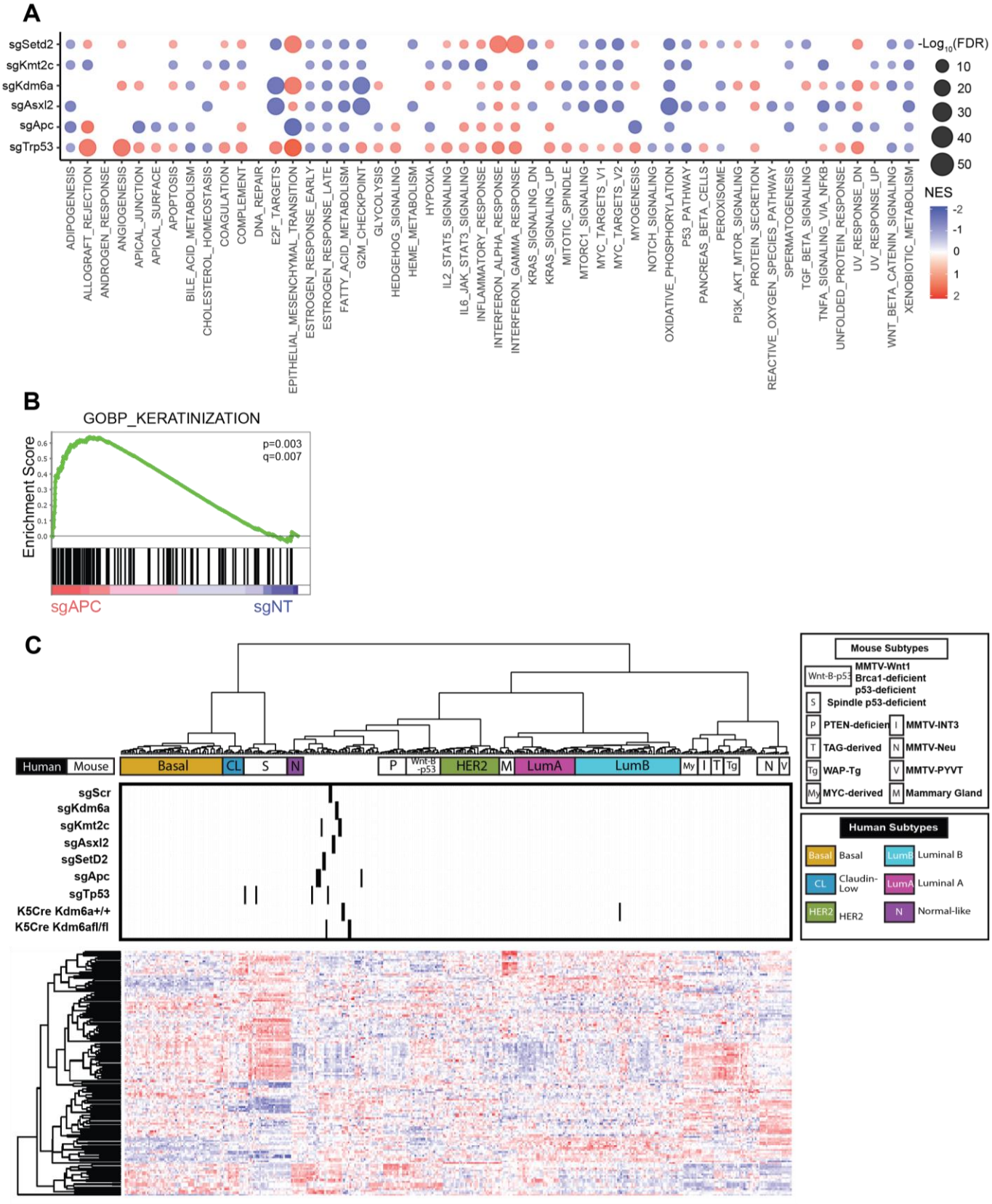
Supplementary Fig. S3. Histological characterization of CRISPR knockout tumors. A, Pathological characterization of mammary tumors from each knockout. **B and C,** Representative tumor H&E from each sgRNA. **D and E,** Quantification of estrogen receptor staining intensity in the nucleus (D) or cytoplasm (E) determined by anti-ERα immunohistochemical staining. **F and G,** Representative images of K8 and K14 staining in sgTp53 and sgNT tumors (G) and in a sgKdm6a tumor (G). **H,** Quantification of K8 and K14 staining in tumors transduced with the indicated sgRNAs.

Supplementary Fig. S4



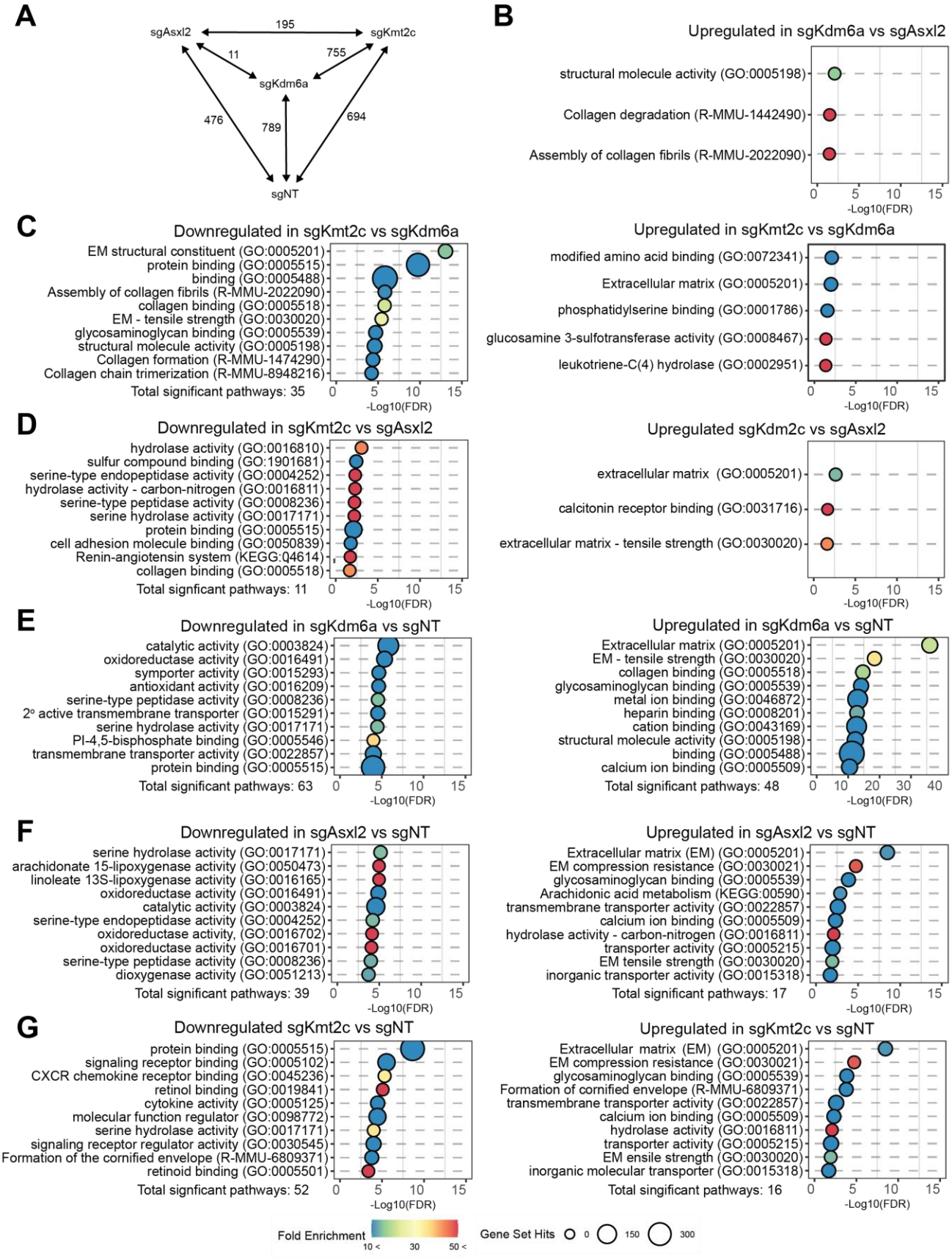
Supplementary Fig. S4. Heterozygous loss of Kdm6a and transcriptional clustering. **A,** Western blot of KDM6A in wild-type, *Kdm6a*$^{fl/+}$ heterozygous and *Kdm6a*$^{fl/fl}$ homozygous mammary tumors. **B,** densitometry of KDM6A normalized to loading control H3 from the blot shown in **A**. **C,** Tumor-free survival of Rb$^{fl/fl}$; Trp53$^{fl/fl}$; LSL-Cas9-EGFP mice transduced individually with sgRNAs targeting *Kmt2c*, *Kdm6a* or non-targeting control (n>4, p vs. control indicated). **D,** Heatmap depicting Pearson's correlation between transcriptional profiles of sgRNA knockout tumors grouped by unsupervised clustering.

Supplementary Fig. S5



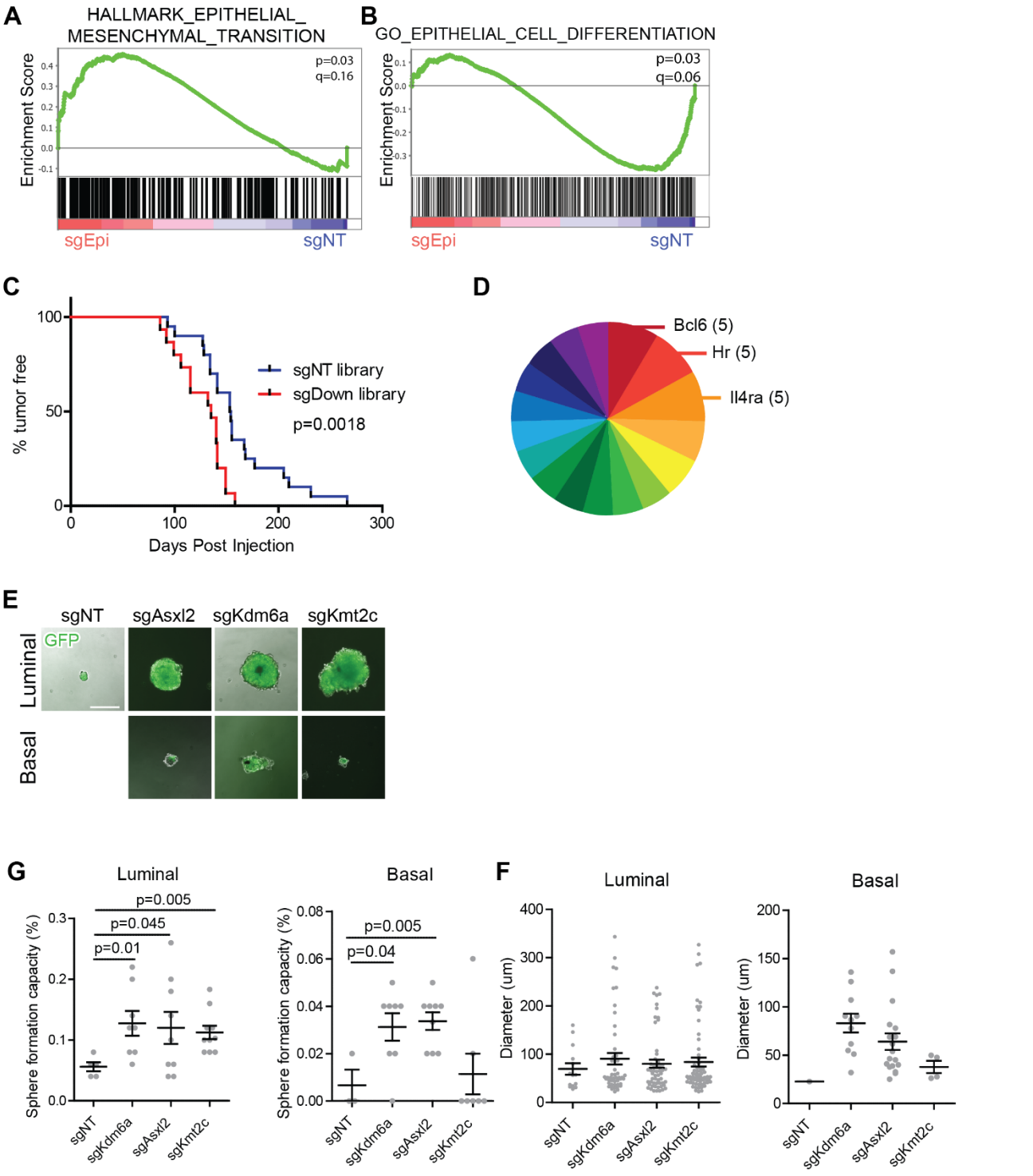
Supplementary Fig. S5. Transcriptional profiling of epigenetic knockout tumors. **A,** Dot plot showing GSEA for hallmark pathways in knockout tumors relative to control tumors. Circle size denotes -log(FDR) with lower cutoff of -log(FDR)=0.6 (FDR=0.25). Color depicts Normalized Enrichment Score (NES) with red denoting enriched pathways and blue denoting depleted pathways. **B,** GSEA plot showing enrichment of 'GO_keratinization' in sgApc tumors compared to control sgNT tumors. **C,** Cross- and interspecies unsupervised clustering of CRISPR/Cas9 EpiDriver knockout Pik3ca^{H1047R} tumors as well as Ad-K5-Cre *Pik3ca^{H1047R}* and Ad-K5-Cre *Pik3ca^{H1047R};Kdm6a$^{fl/fl}$* tumors compared to previously described mouse breast cancer models and human breast cancer subtypes.

Supplementary Fig. S6



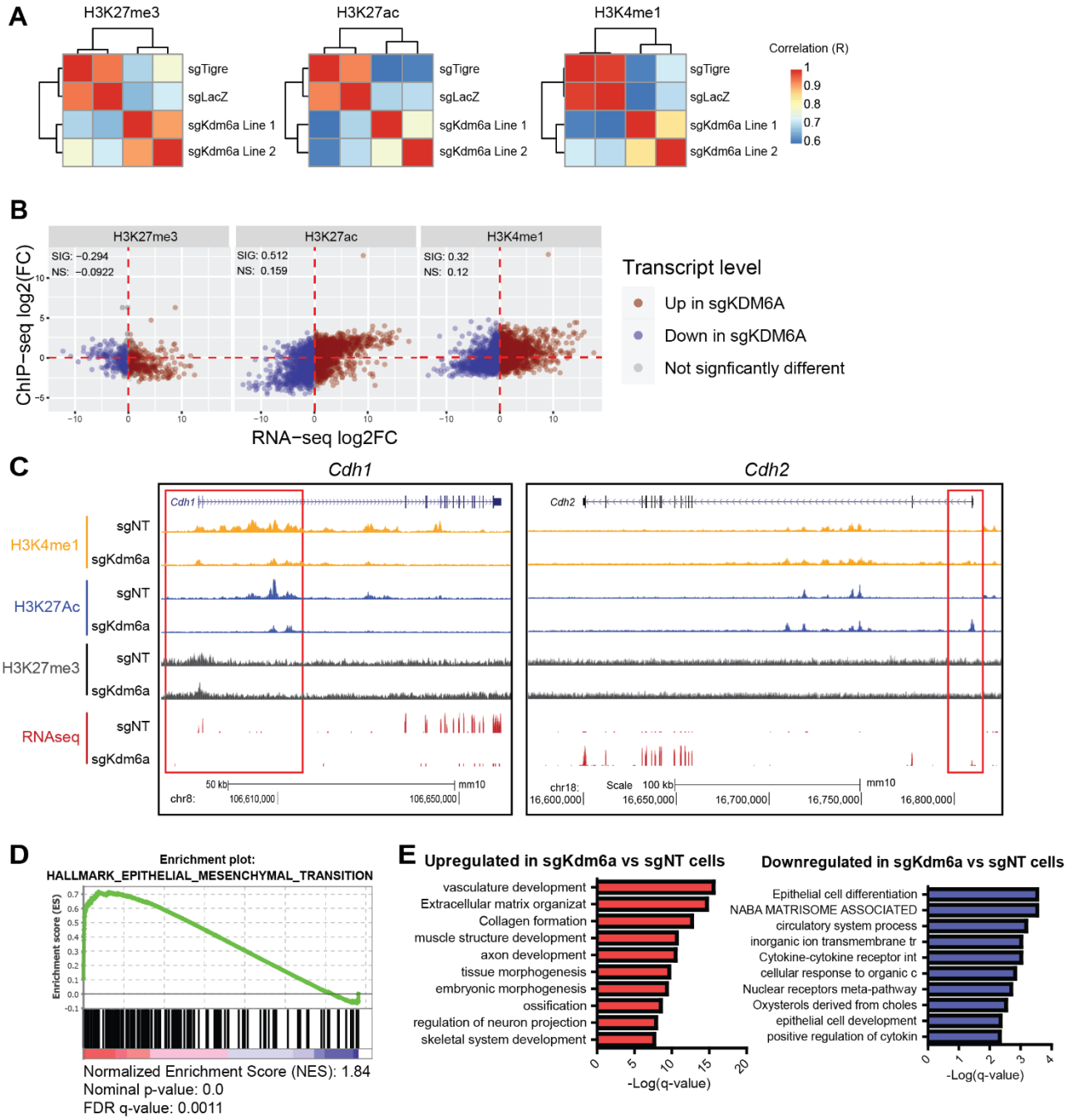
Supplementary Fig. S6. Pairwise comparison of transcriptional profiles of sgNT and sgEpiDriver tumors. A, Number of differentially expressed genes between each pair (fold-change > 2, $p_{adj} < 0.05$). **B-D,** g-profiler pathway enrichment displaying top 10 enriched pathways in significantly downregulated (left) or upregulated (right) genes found by pairwise comparisons between sgKdm6a, sgAsxl2, and sgKmt2c. If more than 10 pathways were enriched, the total number of differential pathways is listed. No significantly downregulated pathways were observed when comparing sgKdm6a to sgAsxl2. **E-G**, g-profiler pathway enrichment for significantly down-regulated (left) or up-regulated (right) genes found by comparison of sgKdm6a, sgAsxl2, and sgKmt2c to sgNT. Legend applies to panels **B-G**.

Supplementary Fig. S7



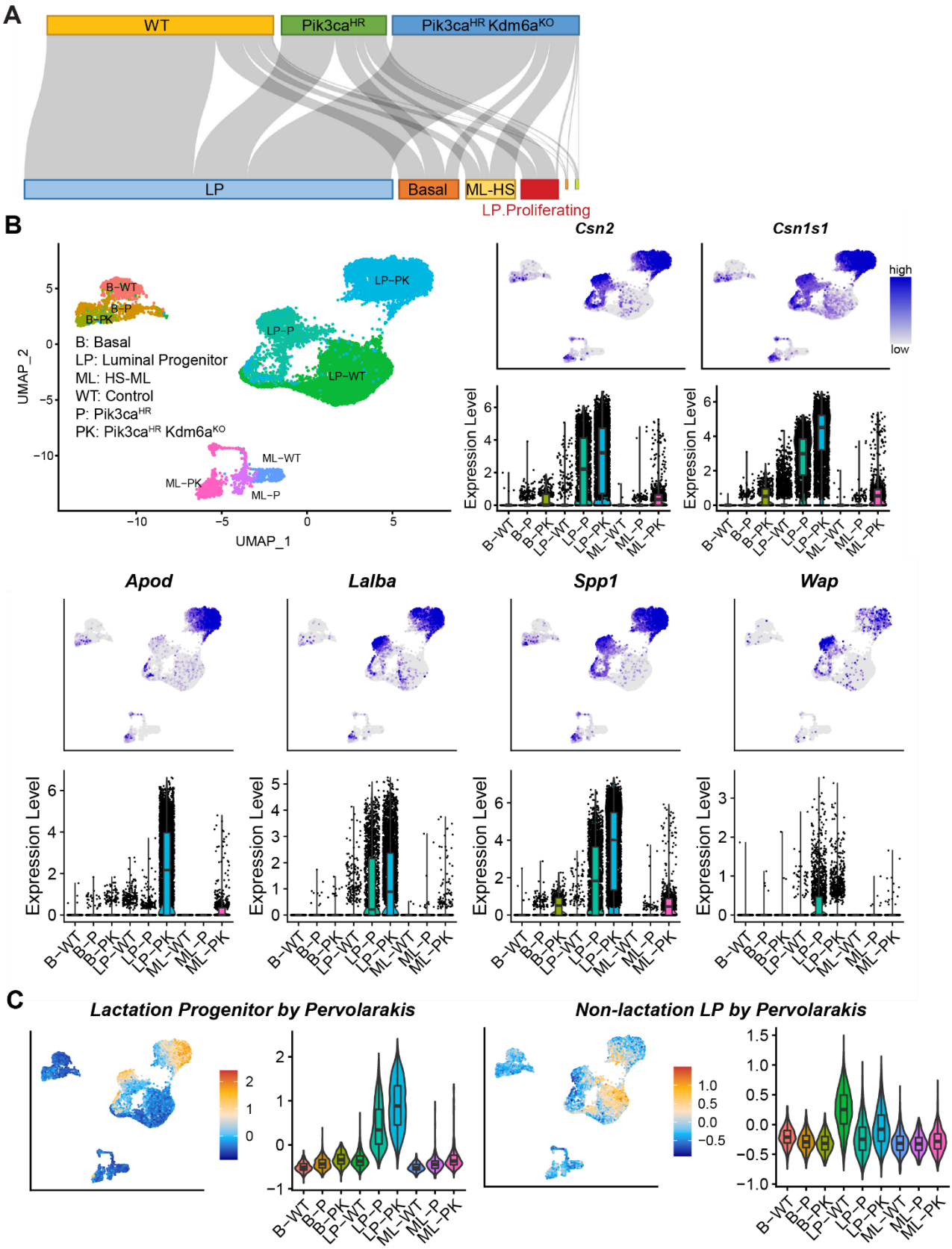
Supplementary Fig. S7. Screen of downregulated genes and sphere forming capacity of EpiDrivers. A and B, GSEA plots showing enrichment of 'HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION'(**A**) and depletion of 'GO_EPITHELIAL_ CELL_DIFFERENTIATION' (**B)** in EpiDriver knockout vs control sgNT tumors. GSEA plots display Nominal p-value (p=) and FDR q-value (q=). **C,** Tumor-free survival of LSL-Pik3ca^{H1047R}; LSL-Cas9-EGFP transduced with lentiviral sgRNA libraries. The sgDown library is composed of 498 genes downregulated in tumors of at least 3 of 4 epigenetic regulators (Asxl2, Kdm6a, Kmt2c and Setd2) compared to sgNT control tumors. **D,** Pie chart showing putative tumor suppressor genes with enriched sgRNAs in tumor DNA (number of tumors are denoted in brackets). Top 3 genes are labelled. **E,** Representative images of spheres grown from sgNT, sgAsxl2, sgKdm6a, or sgKmt2c luminal or basal mammary epithelial cells 10 days after plating. Images show overlap of GFP and brightfield images. Both GFP and brightfield images were taken with auto-exposure. Scale bar = 200 μm. **F,** Sphere forming capacity in sgNT, sgAsxl2, sgKdm6a, or sgKmt2c luminal or basal mammary epithelial cells. **G,** Sphere size of sgNT, sgAsxl2, sgKdm6a, or sgKmt2c luminal or basal mammary epithelial cells.

Supplementary Fig. S8



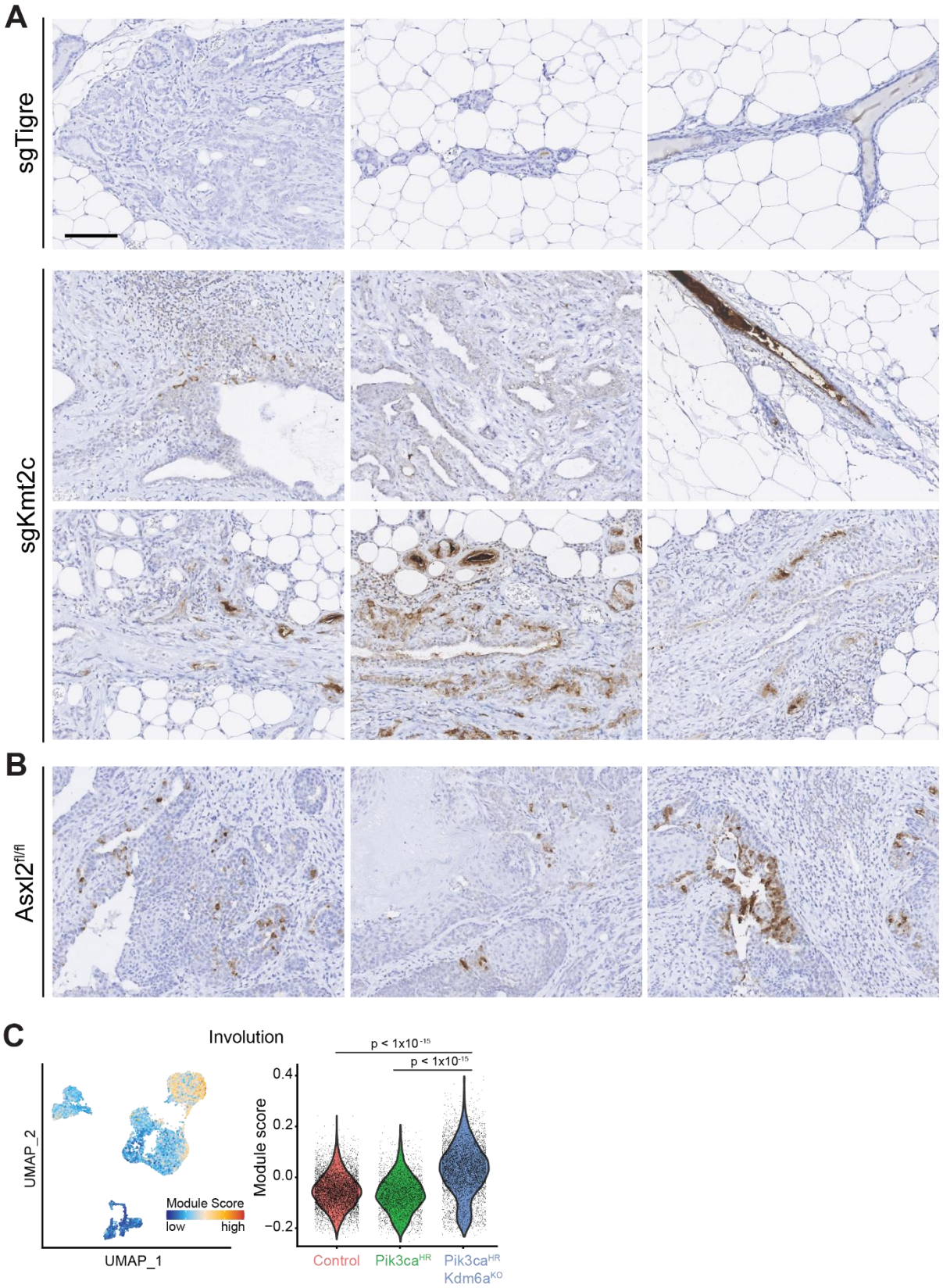
Supplementary Fig. S8. Transcriptional changes show underlying histone modification in cultured primary mouse tumor cells. A, Clustering of ChIP-seq libraries based on pairwise Pearson correlation of raw count values across peak regions. **B,** Scatterplot showing gene expression log2FE (x-axis) and ChIP-seq log2FE at the ChIP-seq peak nearest to the TSS for each respective histone mark. Genes are color-coded based on the direction of gene expression change. Pearson correlation coefficients are reported for significantly (SIG) and non-significant (NS) differentially expressed genes. **C,** UCSC genome browser tracks depicting histone ChIP-seq and RNA-seq signal at prominent DEGs in mouse mammary tumor cell samples. All ChIP-tracks are shown as fold-change over input; all RNA-seq tracks are normalized to total library size. All tracks are shown on the same scale within each experiment for sgKdm6a and control. All tracks show merged signal from 2 replicates with the exception of control H3K27Ac, for which only the higher quality replicate is shown. **D,** GSEA plot showing enrichment of 'HALLMARK_ EPITHELIAL_ MESENCHYMAL_TRANSITION' in *Kdm6a* knockout vs control sgNT cultured tumor cells. **E,** Metascape analysis showing enriched (left) and depleted (right) pathways in common de-regulated genes in cultured sgKdm6a tumor cells compared to control tumor cells.

Supplementary Fig. S9



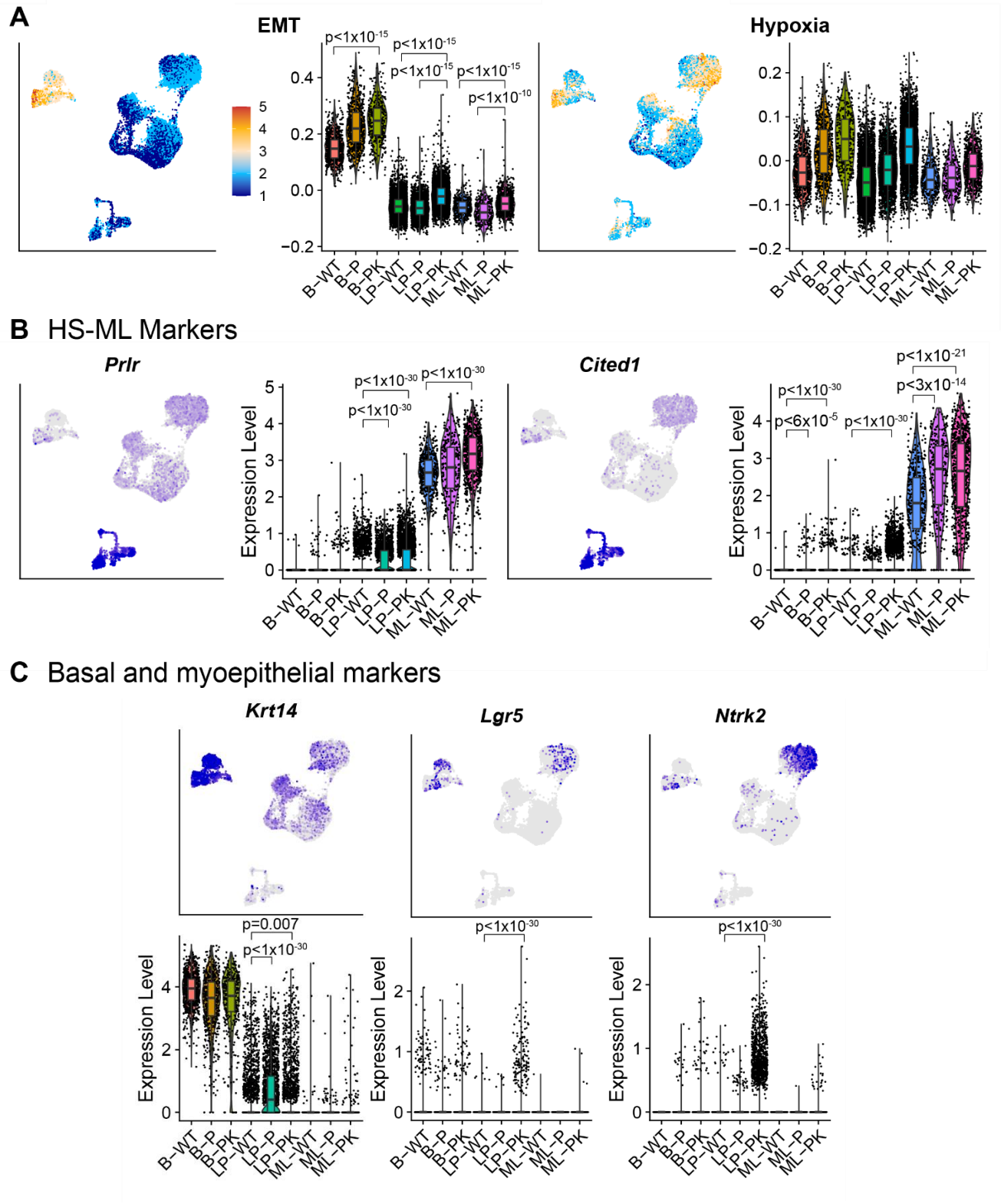
Supplementary Fig. S9. scRNAseq reveals alveogenic mimicry. A, Contribution the genotypes to each cell identity cluster. **B,** Unsupervised UMAP plot of scRNAseq profile colored by genotype and UMAP and violin plots depicting scRNAseq data of selected alveolar/lactation markers. Single gene expression on UMAP plots is shown with a maximum cutoff of q90. **C,** Unsupervised UMAP and violin plots depicting expression of the Lactation Progenitor and non-lactation Luminal Progenitor signatures described by Pervolarakis *et al.* (reference in main text).

Supplementary Fig. S10



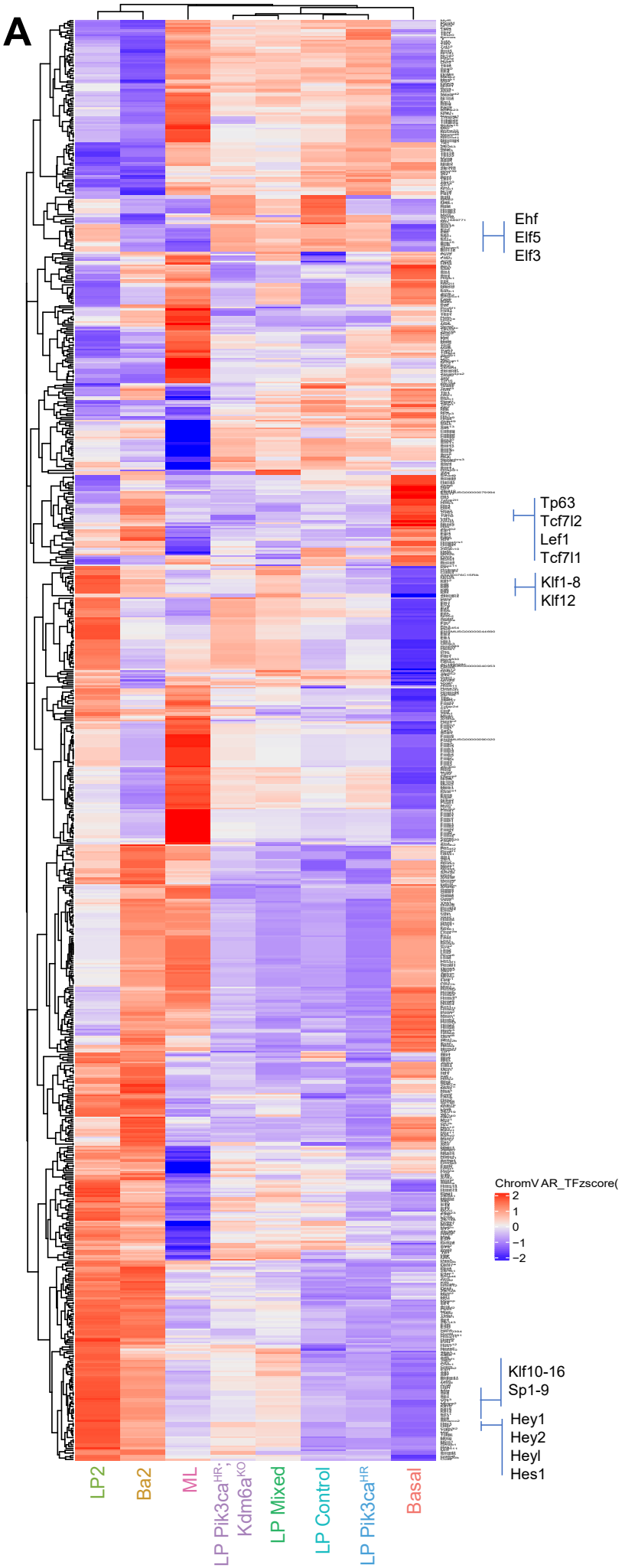
Supplementary Fig. S10. Staining of other epidrivers reveals signs of alveogenic mimicry. A, b-Casein staining in matched sgTigre and sgKmt2c mammary glands. Scale bar = 100 μm. **B**, b-Casein staining in sgAsxl2 tumors. **C,** UMAP and violin plot showing involution signature.

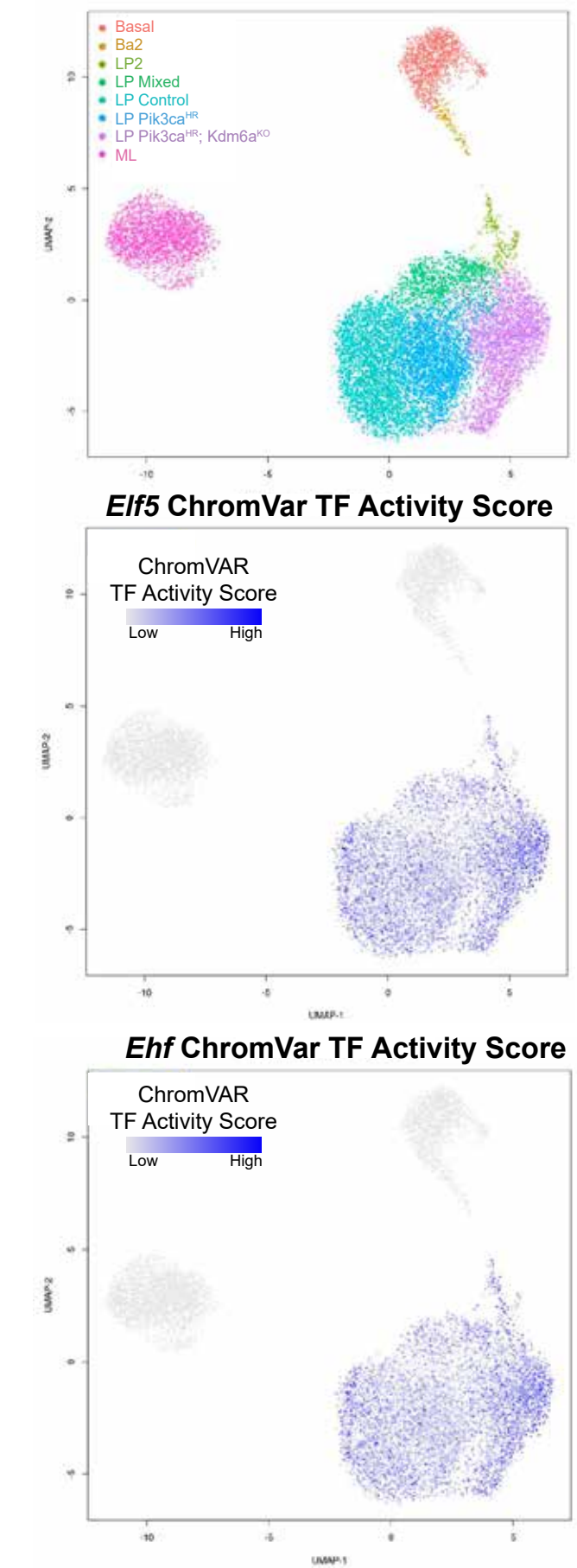
Supplementary Fig. S11



Supplementary Fig. S11. scRNAseq reveals lineage plasticity. A, UMAP and violin plot showing EMT and hypoxia signatures. **B and C**, UMAP and violin plots depicting scRNAseq data of selected HS-ML (**B**) and basal (**C**) markers. Single gene expression on UMAP plots is shown with a maximum cutoff of q90.

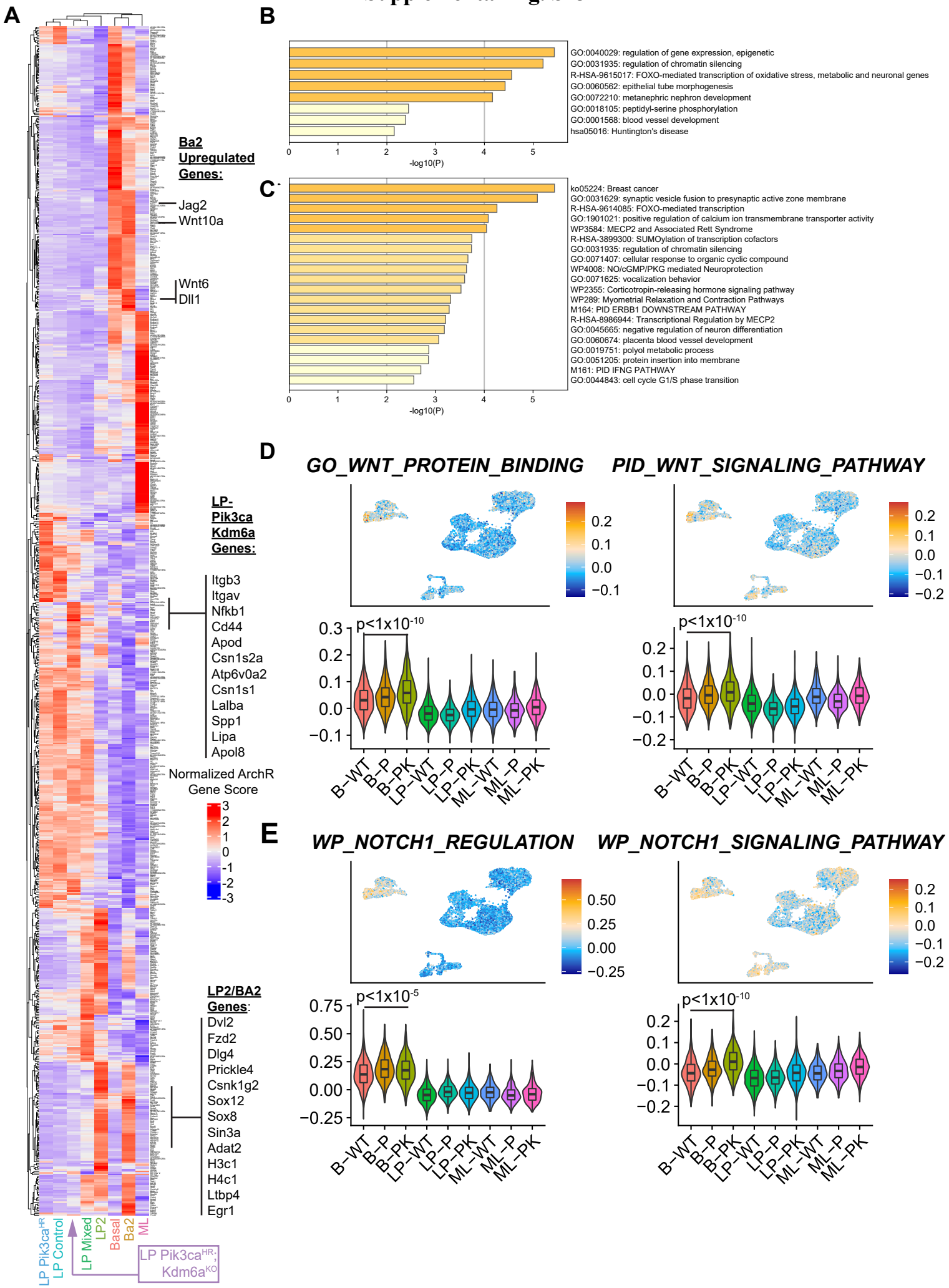
Supplemental Fig. S12





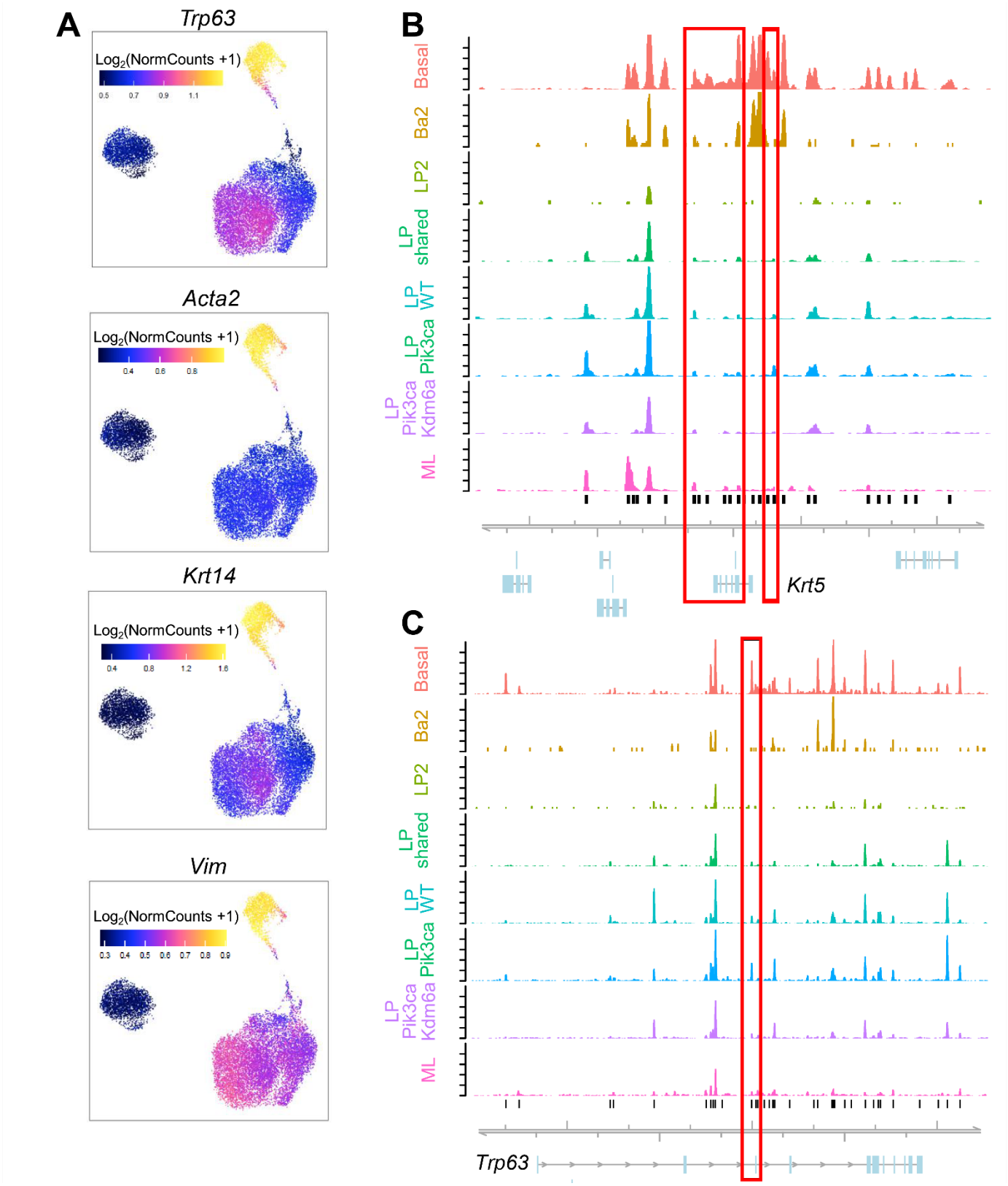
Supplementary Fig. S12. snATACseq reveals significant changes in chromatin accessibility upon Kdm6a loss. **A,** Unsupervised clustering of chromVAR TF activity scores for 788 transcription factors in the major clusters obtained by snATACseq UMAP clustering. **B,** UMAP plot of snATACseq profile colored by genotype and cluster (top) and UMAP plots depicting Ehf (middle) and Elf5 (bottom) ChromVAR TF activity score.

Supplemental Fig. S13



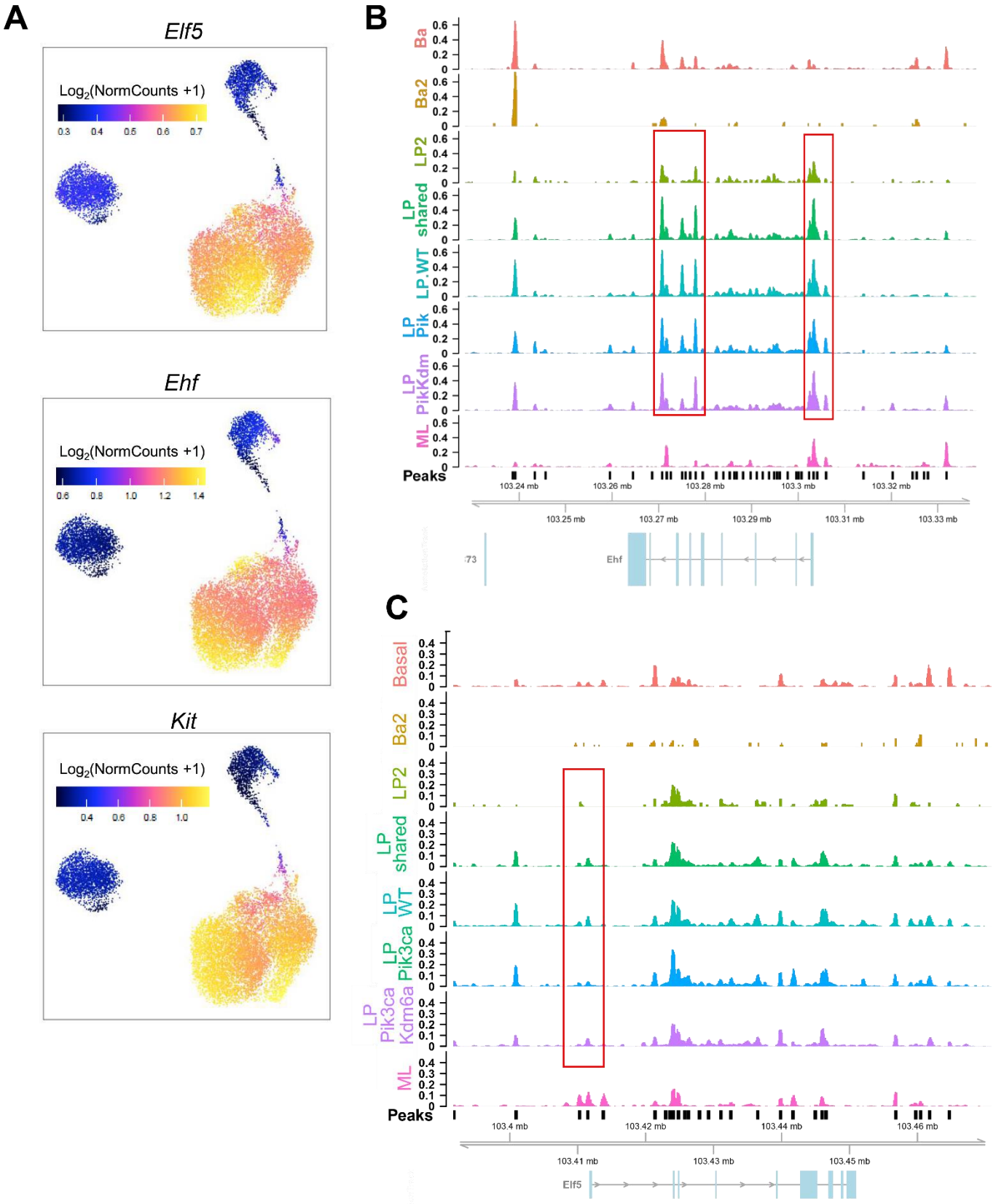
Supplementary Fig. S13. snATACseq confirms alveogenic mimicry and reveals WNT as well as NOTCH signaling in BA2 cluster. A, Unsupervised clustering of ArchR gene accessibility score across the major clusters obtained by snATACseq. **B and C**, METASCAPE pathway enrichment analysis showing enriched pathways shared between BA2 and LP2 clusters compared to all other clusters (**B**) and between BA2 compared to the basal clusters (**C**) as inferred by gene accessibility scores obtained by snATACseq. **D and E,** UMAP and violin plots showing WNT protein binding and signaling (**D**) and NOTCH regulation and signaling (**E**) signatures by scRNAseq.

Supplementary Fig. S14



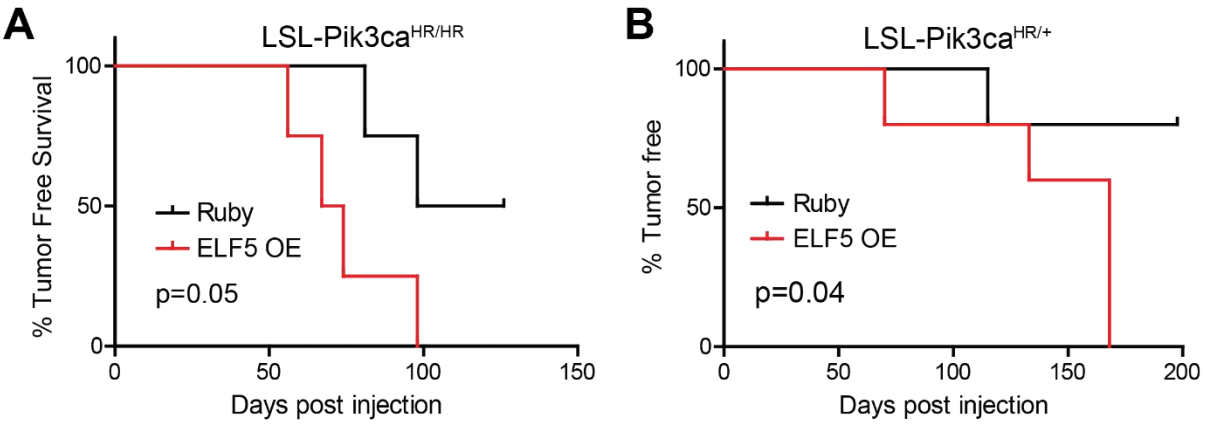
Supplementary Fig. S14. BA2 cluster exhibits downregulated basal markers. A, Unsupervised UMAP plot of snATACseq profile showing single-cell gene accessibility scores of the indicated basal cell markers revealing downregulation in the BA2 cluster. **B and C**, Signal tracks of the aggregate snATAC profile for Krt5 (**B**) and Trp63 (**C**) for each the indicated cluster.

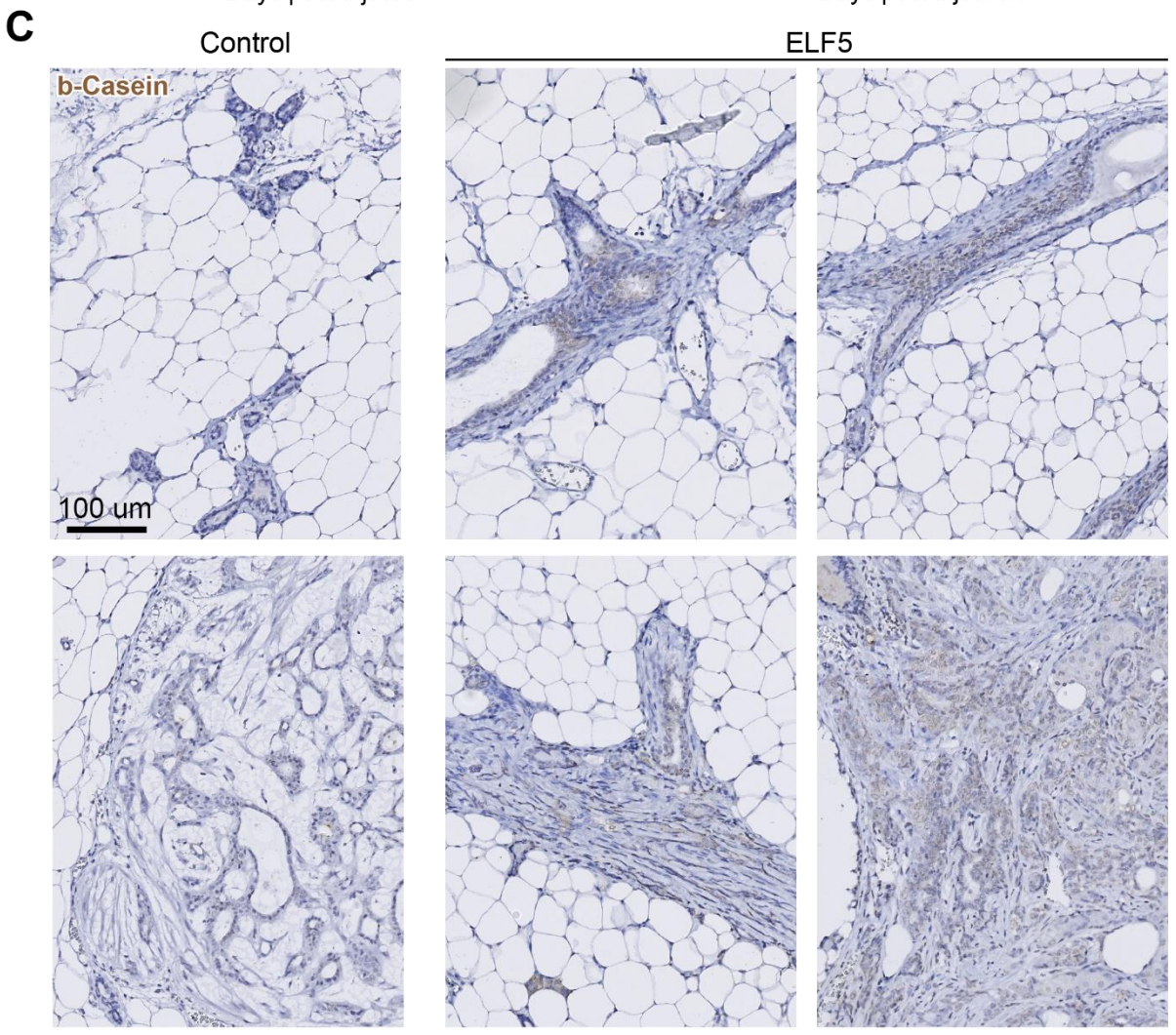
Supplementary Fig. S15



Supplementary Fig. S15. LP2 cluster exhibits downregulated luminal markers. A, Unsupervised UMAP plot of snATACseq profile showing single-cell gene accessibility scores of the indicated luminal cell markers revealing downregulation in the LA2 cluster. **B and C**, Signal tracks of the aggregate snATAC profile for Ehf (**B**) and Elf5 (**C**) for each the indicated cluster.

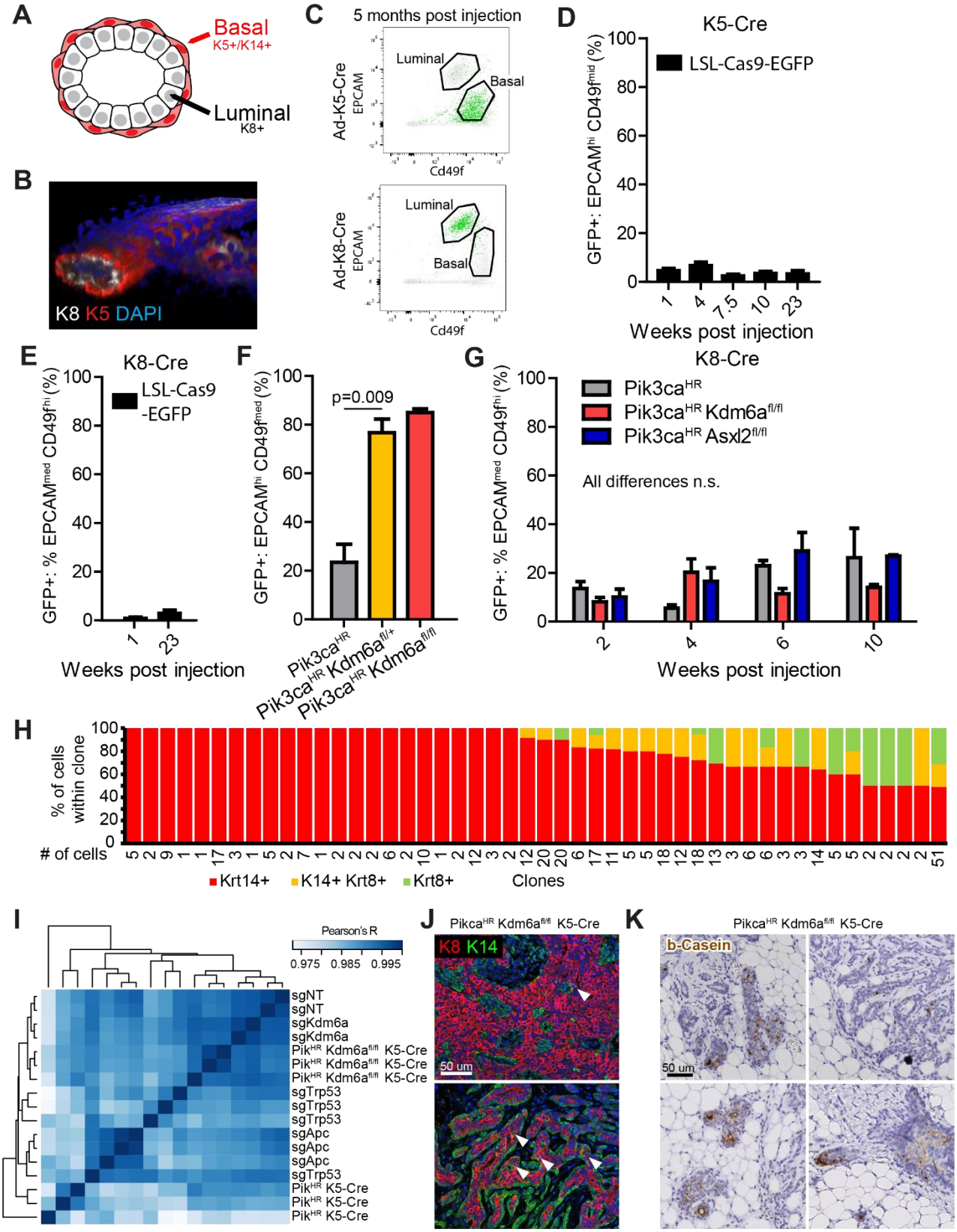
Supplementary Fig. S16





Supplementary Fig. S16. ELF5 over-expression accelerates tumorigenesis. A and B, Tumor-free survival for *Pik3ca*$^{H1047R/H1047R}$ (**A**) or *Pik3ca*^{H1047R};Cas9 (**B**) mice transduced with lentiviral over-expression of ELF5 or Ruby control (n≥4 per group). **C**, B-Casein expression in matched mammary glands for either control or overexpression of ELF5.

Supplementary Fig. S17

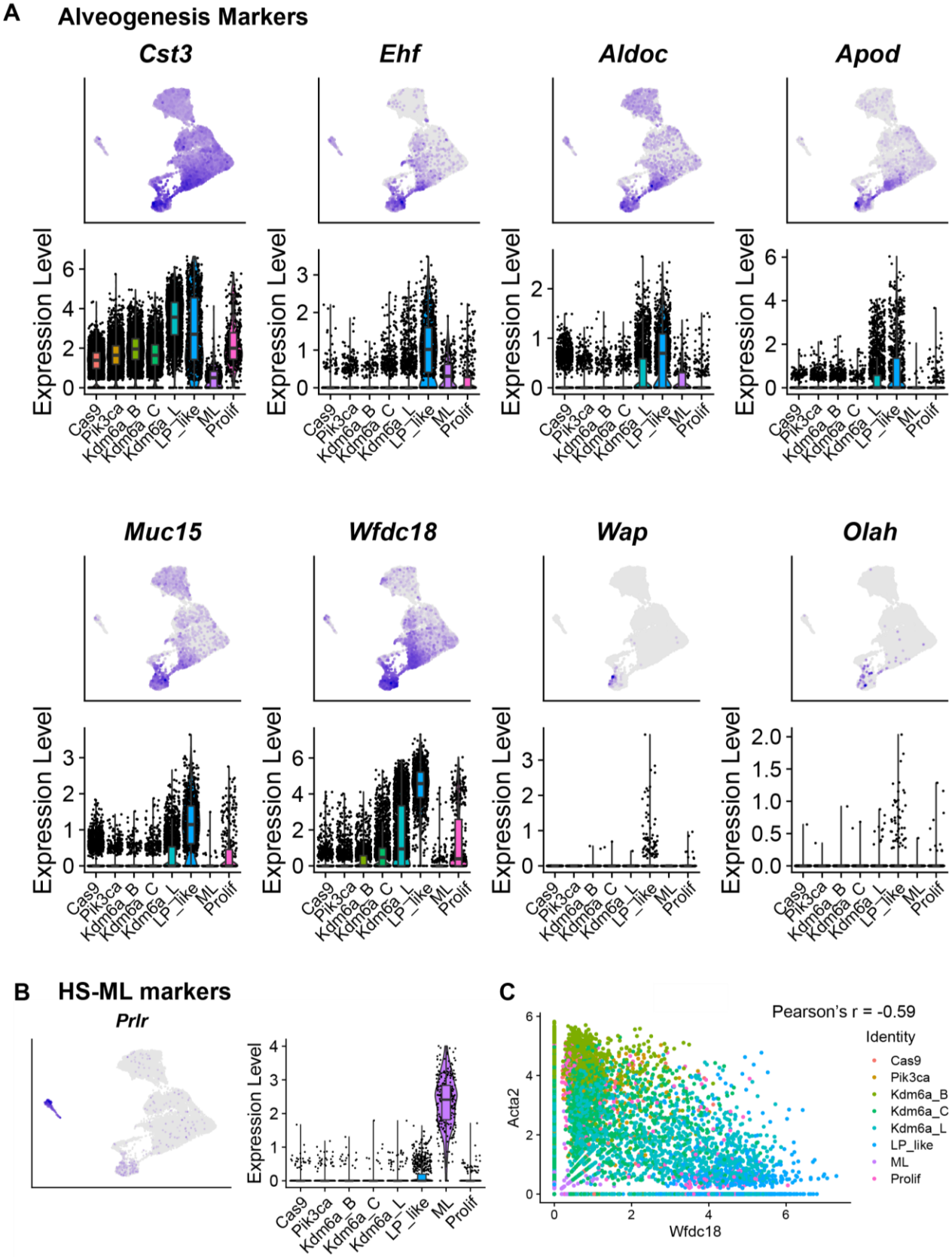


Supplementary Fig. S17. Lineage tracing of basal and luminal cells. A and B, Schematic (**A**) and representative immunofluorescent whole mount image (**B**) illustrating the location of basal K5+/K14+ and luminal K8+/K18+ mammary epithelial cells. **C,** Representative FACS blot depicting lineage tracing strategy and lineage restriction of Ad-K5-Cre and Ad-K8-Cre viral particles. **D,** Percent of GFP+ $EPCAM^{hi}CD49f^{mid}$ luminal cells at different time points after Ad-K5-Cre injection into mammary epithelium of R26-LSL-Cas9-GFP mice as determined by flow cytometry. **E,** Percent of GFP+ $EPCAM^{med}CD49f^{hi}$ basal cells at different time points after Ad-K8-Cre injection into mammary epithelium of R26-LSL-Cas9-GFP mice as determined by flow cytometry. **F,** Percent of GFP+ $EPCAM^{hi}CD49f^{mid}$ luminal cells at 5.5 weeks after Ad-K5-Cre injection into mammary epithelium of heterozygous $Kdm6a^{fl/+}$;$Pik3ca^{H1047R}$;R26-LSL-GFP mice as determined by flow cytometry. **G,** Percent of GFP+ $EPCAM^{mid}CD49f^{hi}$ basal cells at different time points after Ad-K8-Cre injection into mammary epithelium of $Pik3ca^{H1047R}$, $Pik3ca^{H1047R}$;$Kdm6a^{fl/fl}$ and $Pik3ca^{H1047R}$;$Asxl2^{fl/fl}$ mice. **H**, Clone composition 4 weeks post IP injected of 2 mg of tamoxifen into Krt5-CreERT2; $Pik3ca^{H1047R}$; LSL-Cas9-EGFP; $Kdm6a^{fl/fl}$ mice. Clones were identified by immunofluorescence of GFP, anti-Krt8 and anti-Krt14. Clone size in cell number is listed on the x axis. **I,** Clustering of transcriptional profiles of purified tumor cells from either lentiviral sgRNA mediated knockout in $Pik3ca^{H1047R}$; LSL-Cas9-EGFP mice (sgNT, sgKdm6a, sgTrp53, and sgApc), or Ad-K5-Cre injection into either $Pik3ca^{H1047R}$; LSL-Cas9-EGFP or $Pik3ca^{H1047R}$; LSL-Cas9-EGFP; $Kdm6a^{fl/fl}$ mice. **J,** Immunofluorescence staining of Krt8 and Krt14 in tumors from $Pik3ca^{H1047R}$; LSL-Cas9-EGFP; $Kdm6a^{fl/fl}$ mice injected with Ad-K5-Cre. Arrowheads indicate double-positive cells. **K**, Casein expression in mammary tumors and hyperplasia in $Pik3ca^{H1047R}$; LSL-Cas9-EGFP; $Kdm6a^{fl/fl}$ mice injected with Ad-K5-Cre.

Supplementary Fig. S18

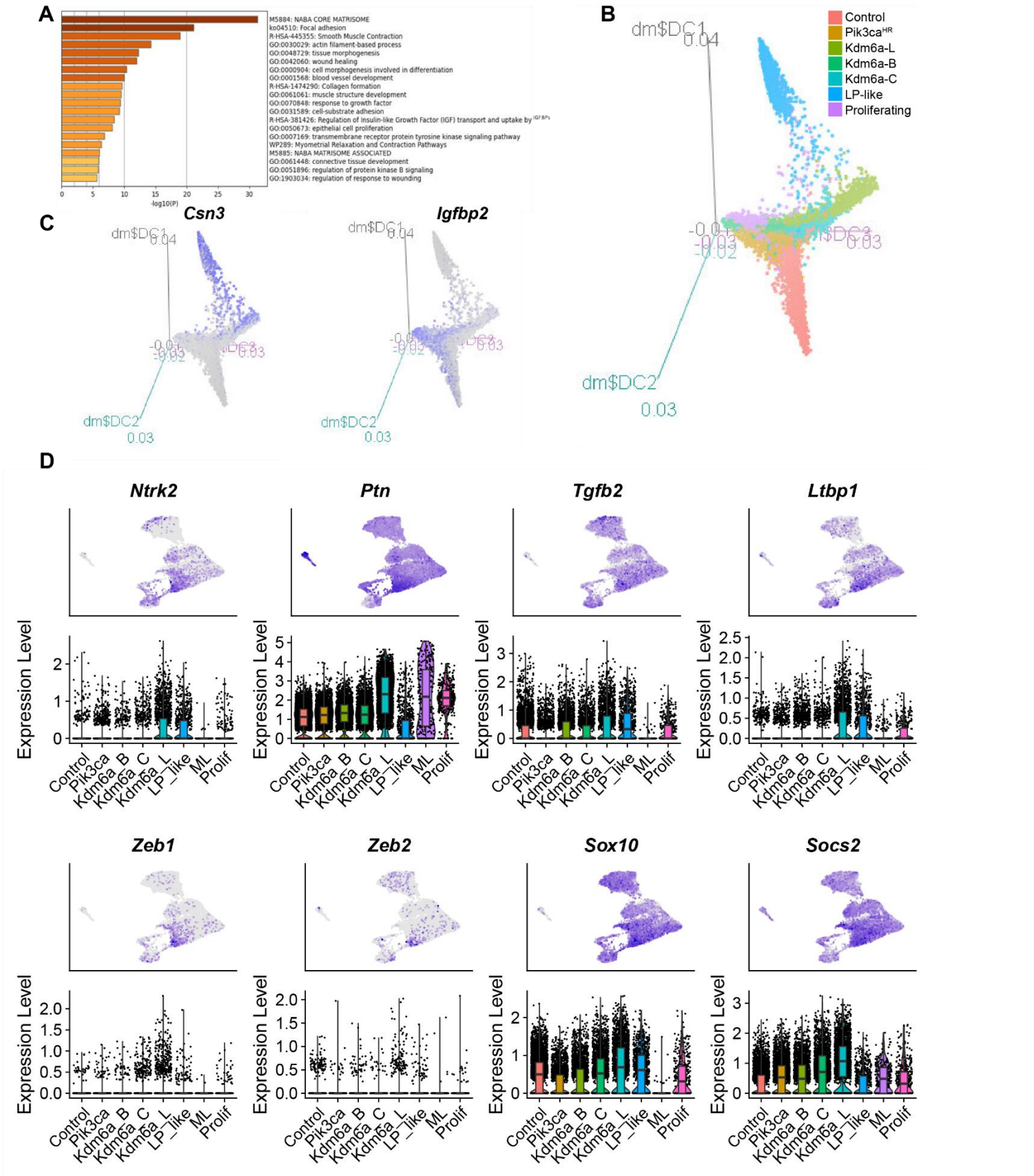
Supplementary Fig. S18. Expression of mammary epithelial lineage markers in cells labelled by K5-Cre. A, Contribution the genotypes to each cell clusters. **B-C,** UMAP and violin plots of K5-Cre scRNAseq data of the indicated basal cell markers (B) and luminal progenitor markers (C).

Supplementary Fig. S19



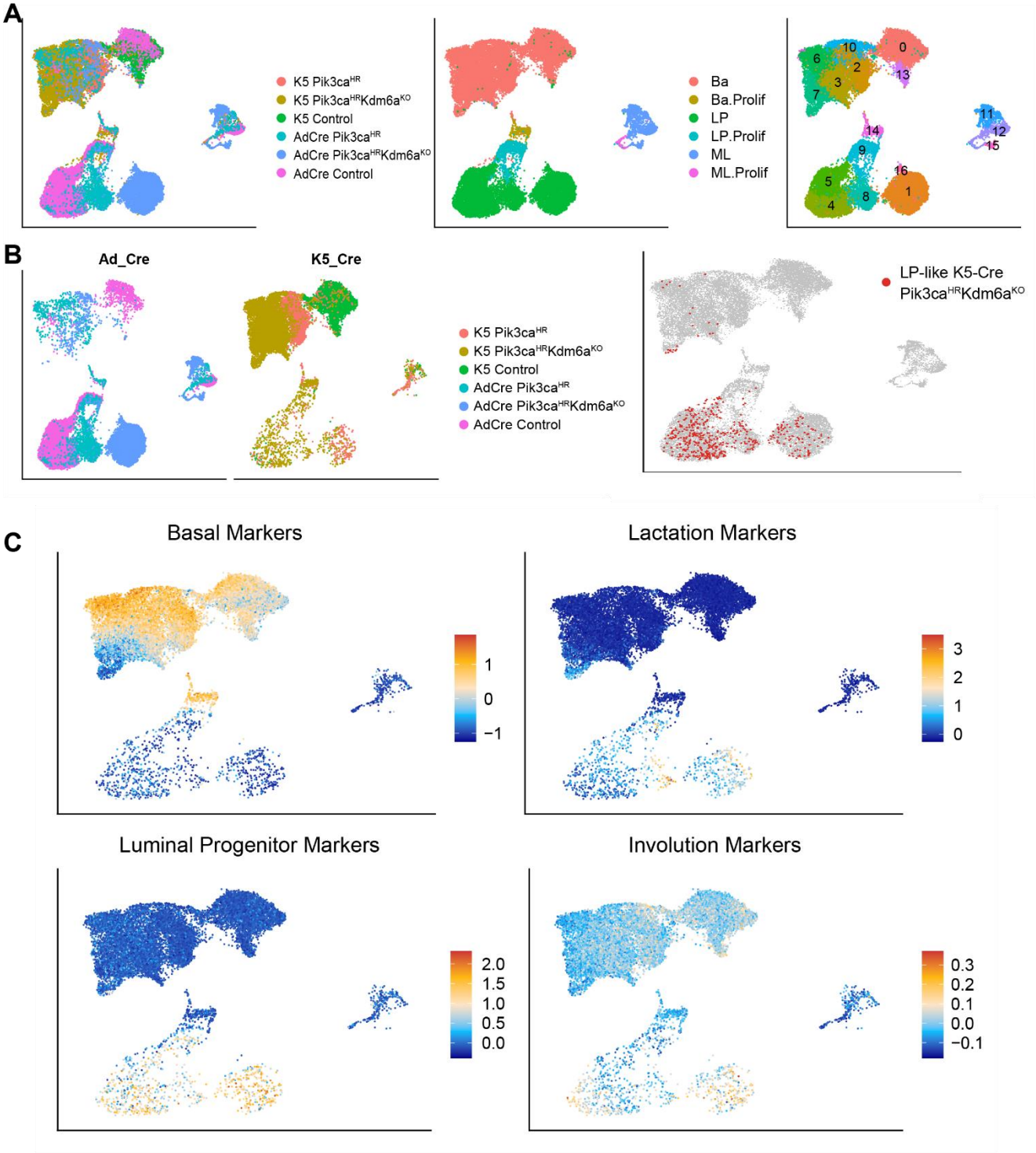
Supplementary Fig. S19. scRNAseq reveals expression of alveolar markers in basal cells. A and B, UMAP and violin plots of K5-Cre scRNAseq data of the indicated alveogenesis markers (**A**) and HS-ML markers (**B**). **C,** Plot showing expression of alveolar marker Wfdc18 and basal marker Acta2 in scRNAseq data. Of note, many Pik3ca;Kdm6a^{KO} express both markers.

Supplementary Fig. S20



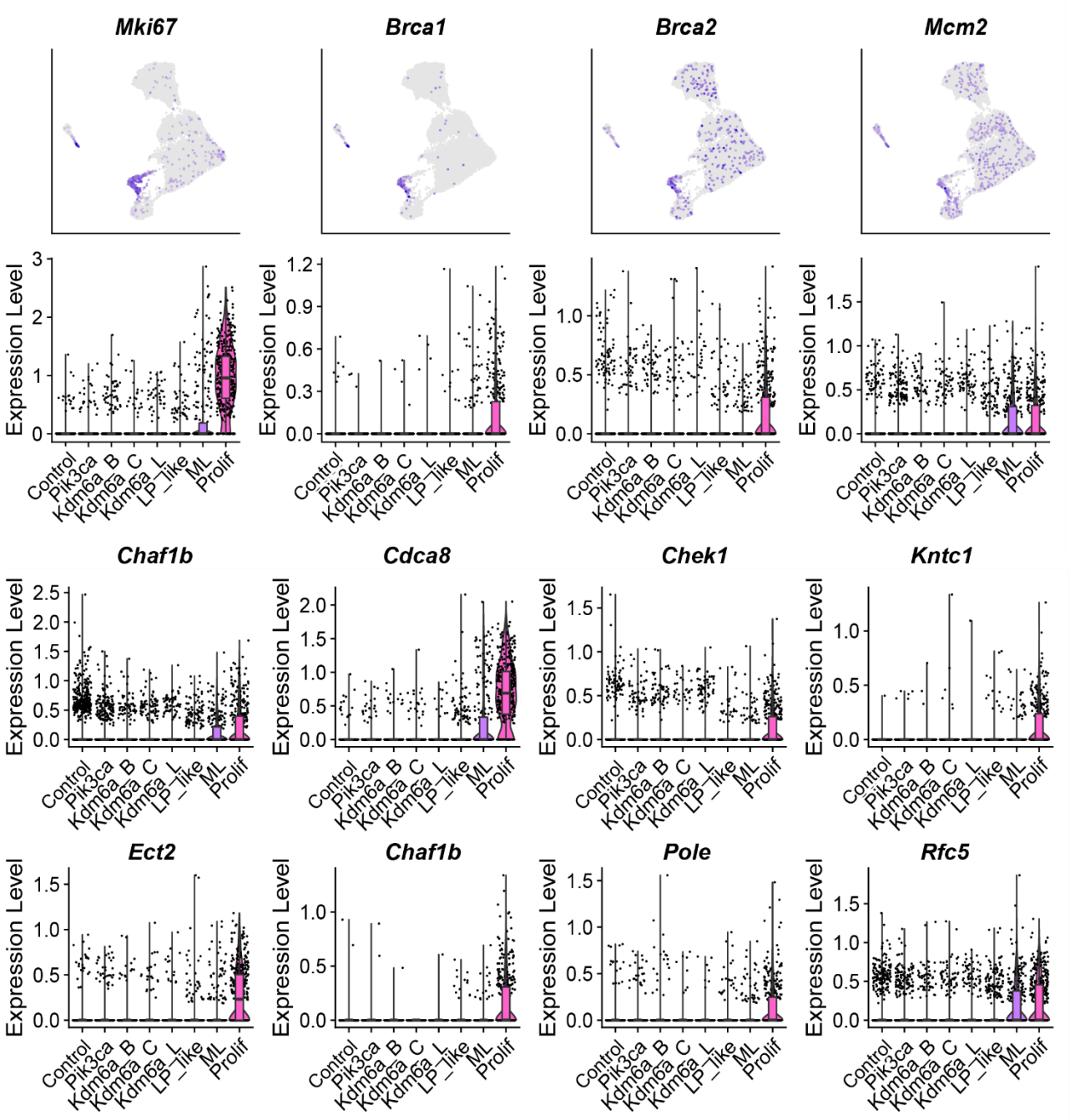
Supplementary Fig. S20. scRNAseq reveals heterogeneity along Pik3ca^{H1047R};Kdm6a^{KO} cells. A, Metascape pathway enrichment analysis showing enriched pathways in Kdm6a_B versus Kdm6a_L clusters. **B,** Diffusion mapping of scRNAseq data displayed by a 3D plot of the first 3 eigenvectors depicting branching of Kdm6a^{KO} cells between Kdm6a_B and Kdm6a_L clusters. **C,** Diffusion mapping of scRNAseq data displayed by a 3D plot of the first 3 eigenvectors depicting alveolar marker *Csn3* and the basal marker *Igfbp2* showing separation of the Kdm6a clusters. **D,** UMAP and violin plot of scRNAseq data showing expression of the genes that are upregulated in the Kdm6a_L cluster.

Supplementary Fig. S21



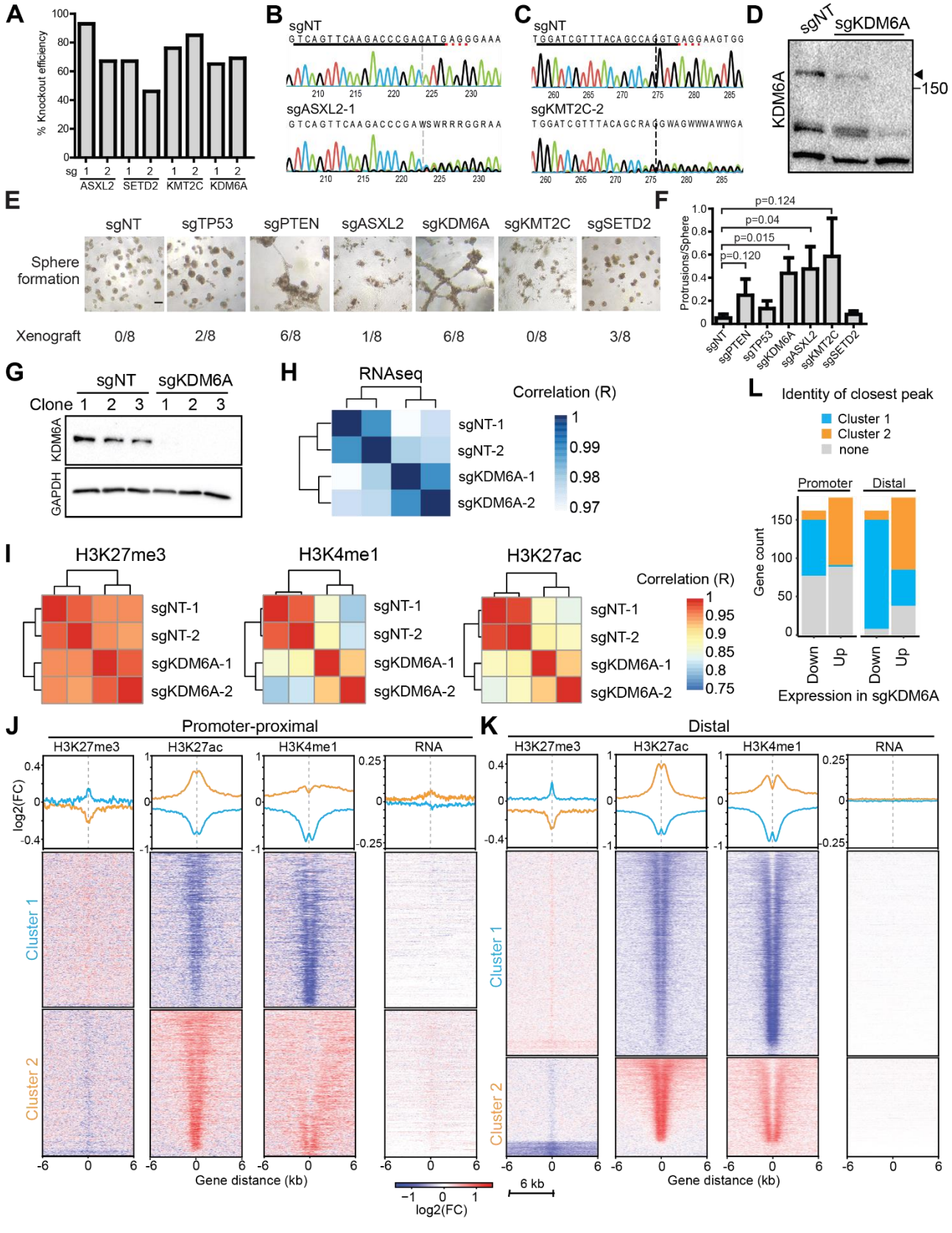
Supplementary Fig. S21. Integration of Ad-Cre and Ad-K5Cre scRNAseq. A, UMAP plots of scRNAseq data of Ad-Cre and Ad-K5Cre scRNAseq colored by genotype (left), by epithelial lineage (middle) and identified clusters (right). **B,** UMAP plots of the combined scRNAseq data showing only the Ad-Cre data (left), only the Ad-K5-Cre data (middle) and with Ad-K5-Cre Pik3ca;Kdm6a^{KO}-LP-like cluster highlighted in red (right). **C,** UMAP plots of the combined scRNAseq data showing only the Ad-Cre data showing basal (top left), alveolar/lactation (top right), luminal progenitor (bottom left), and involution (bottom right) marker signatures.

Supplementary Fig. S22



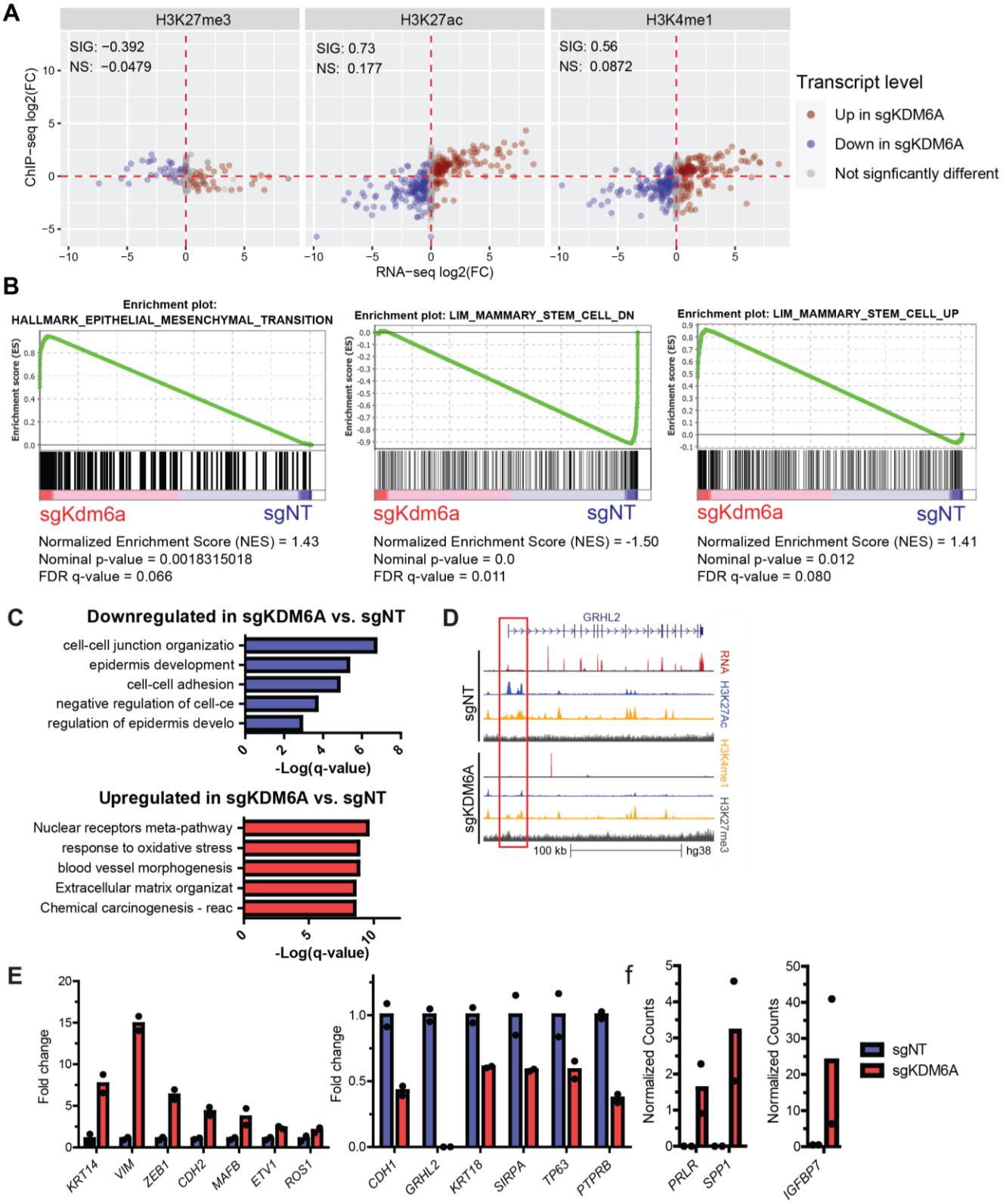
Supplementary Fig. S22. scRNAseq reveals Rb1/E2F pathway in basal proliferating cells and marker genes of BA-alveolar-LP transdifferentation. UMAP and violin plots of scRNAseq data showing expression of the proliferation marker Mki67 (KI67) and the indicated E2F/Rb1 target gene.

Supplementary Fig. S23



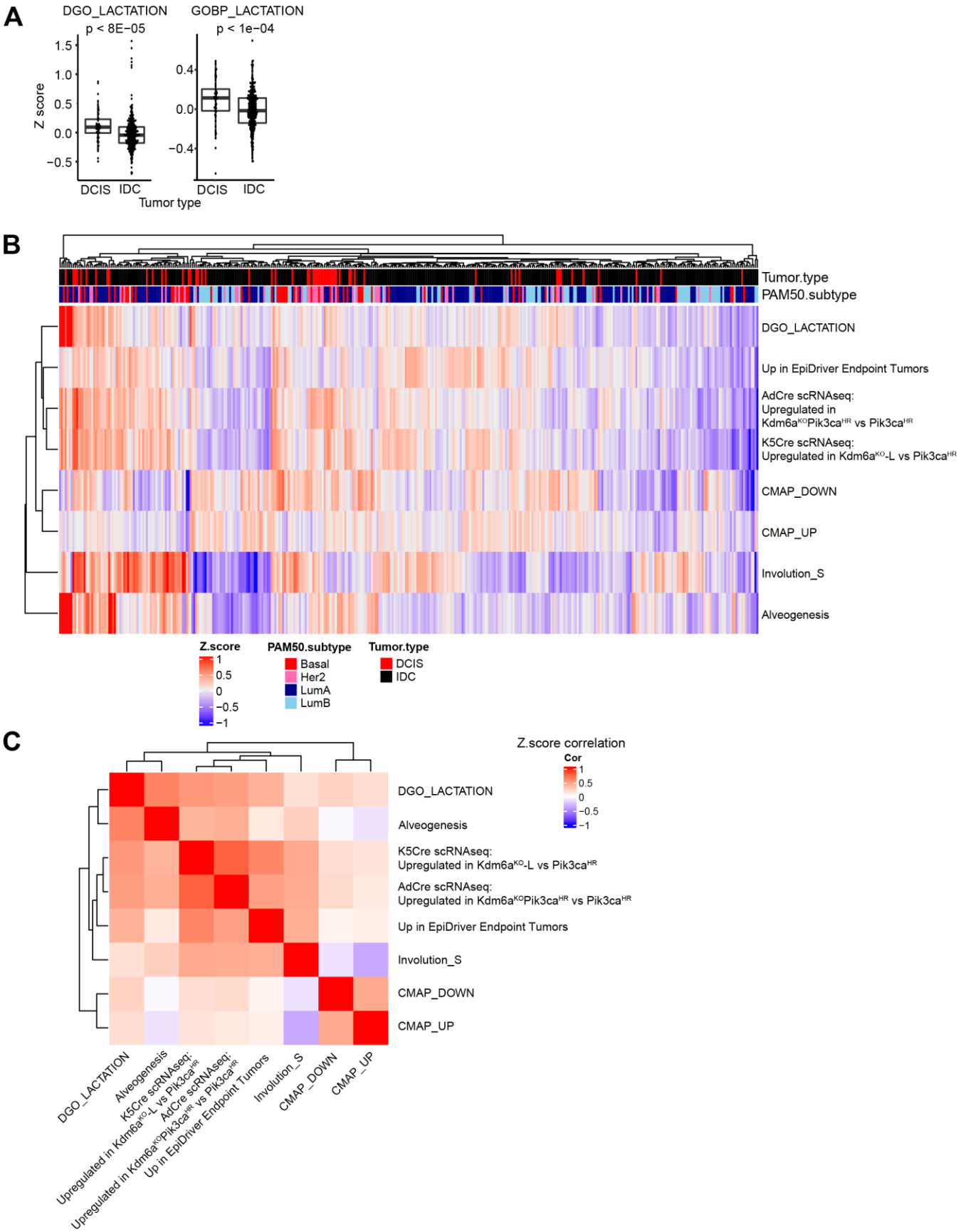
Supplementary Fig. S23. EpiDriver loss triggers transformation of MCF10A-PIK3CA^{H1047R} cells.
A-D, CRISPR/Cas9-mediated knock-out efficacy in MCF10A-PIK3CA^{H1047R} cells transduced with Cas9-Puro-sgRNA lentivirus targeting the indicated gene. TIDE analysis showing indel abundance (**A**), representative chromatograms (**B and C**) and western blot showing protein levels (**D**). **E**, Representative images of spheres formed by MCF10A-PIK3CA^{H1047R} cells with CRISPR/Cas9 knockout of the indicated genes (top) and their xenograft efficiency in NSG mice (bottom). **F**, Quantification of number of sphere protrusions of MCF10A-PIK3CA^{H1047R} cells with CRISPR/Cas9 knockout of the indicated genes. **G**, Western blot showing knockout efficiency in sgNT and sgKDM6A clones. **H**, Euclidean distance and unsupervised clustering of transcriptomes from sgNT and sgKDM6A clones. **I,** Pearson's correlation and unsupervised clustering of histone profiles from sgNT and sgKDM6A clones. **J-K,** K-means clustering of differential regions between sgNT and sgKDM6A clones. Regions were separated into two clusters based on H3K27me3, H3K27ac, and H3K4me1 profile. Fold-change in sgKDM6A vs. sgNT shown as an average across all regions (top), or color-coded for individual regions (bottom, FC legend pictured below). Fold-change in transcription for each region is displayed on the right. Regions were stratified based on their location, either promoter proximal (**J**) or distal (**K**). **L**, The closest peak within 2.5 kb (promoter-proximal) or 100 kb (distal) of each differentially expressed gene is labelled based on cluster.

Supplementary Fig. S24



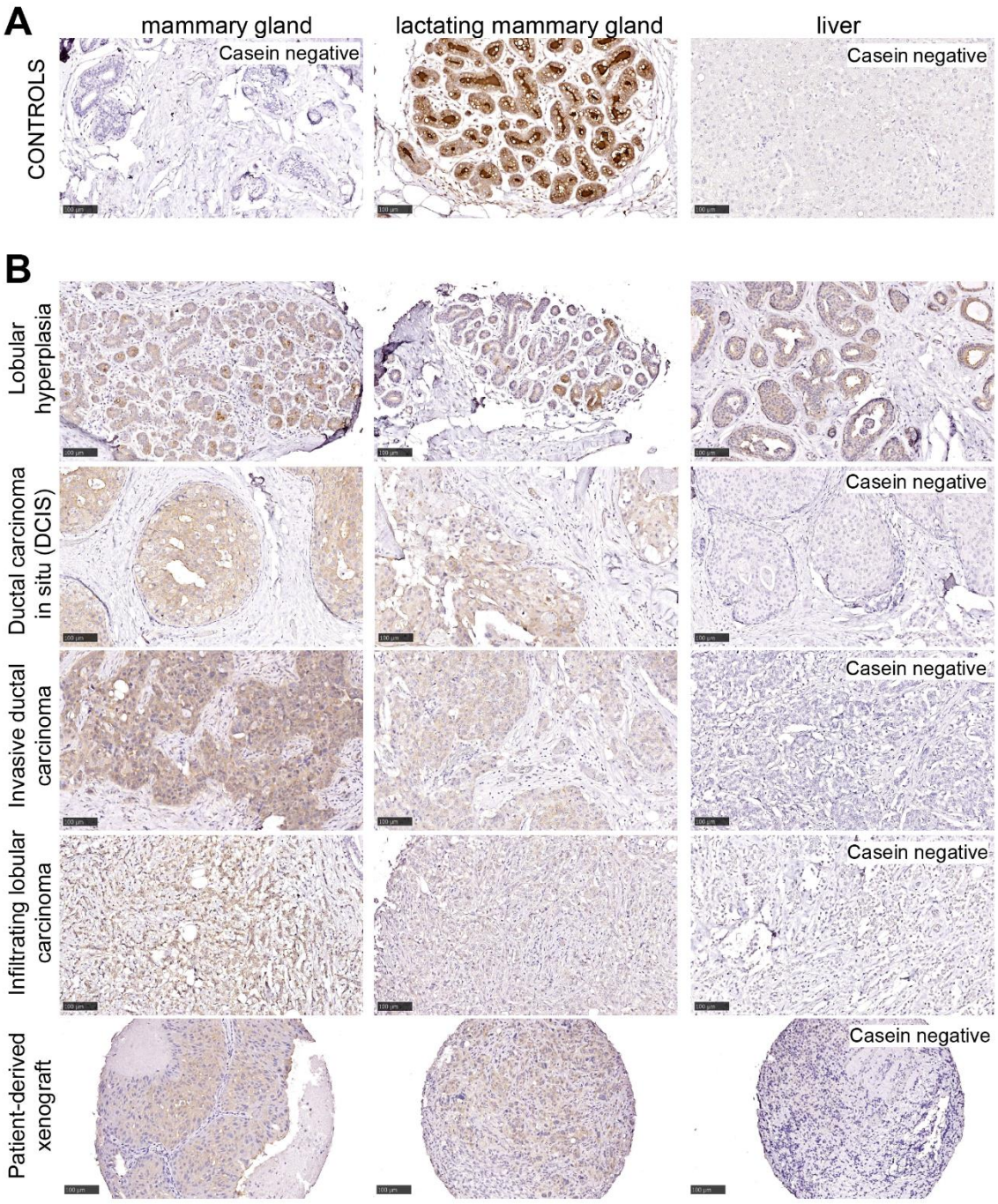
Supplementary Fig. S24. Knockout of KDM6A in MCF10A-PIK3CA^{H1047R} cells results in altered differentiation. **A,** Scatterplot showing gene expression log2FE (x-axis) and ChIP-seq log2FE at the ChIP-seq peak nearest to the TSS for each respective histone mark. Genes are color-coded based on the direction of gene expression change. Pearson correlation coefficients are reported for significantly (SIG) and non-significant (NS) differentially expressed genes. **B,** GSEA plots for sgKDM6A MCF10A-PIK3CA^{H1047R} cells compared to sgNT MCF10A-PIK3CA^{H1047R} cells. **C,** METASCAPE gene enrichment analysis for up- and down-regulated genes in sgKDM6A compared to sgNT control clones. **D**, UCSC genome browser tracks depicting histone ChIP-seq and RNA-seq signal at *GRHL2*, a repressor of EMT in sgKDM6A MCF10A-PIK3CA^{H1047R} cells compared to sgNT MCF10A-PIK3CA^{H1047R} cells. All ChIP-tracks are shown as fold-change over input; all RNA-seq tracks are normalized to total library size. All tracks are shown on the same scale within each experiment for sgKdm6a and sgNT. All tracks show merged signal from 2 replicates **E**, Expression of markers in RNAseq data depicted as fold-change over sgNT. $p^{adj}<0.05$ between sgNT and sgKdm6a for all genes shown. **F**, Expression of lactation-associated genes in RNAseq data depicted as normalized counts. Dots indicate values for each sample in **E** and **F.**

Supplementary Fig. S25



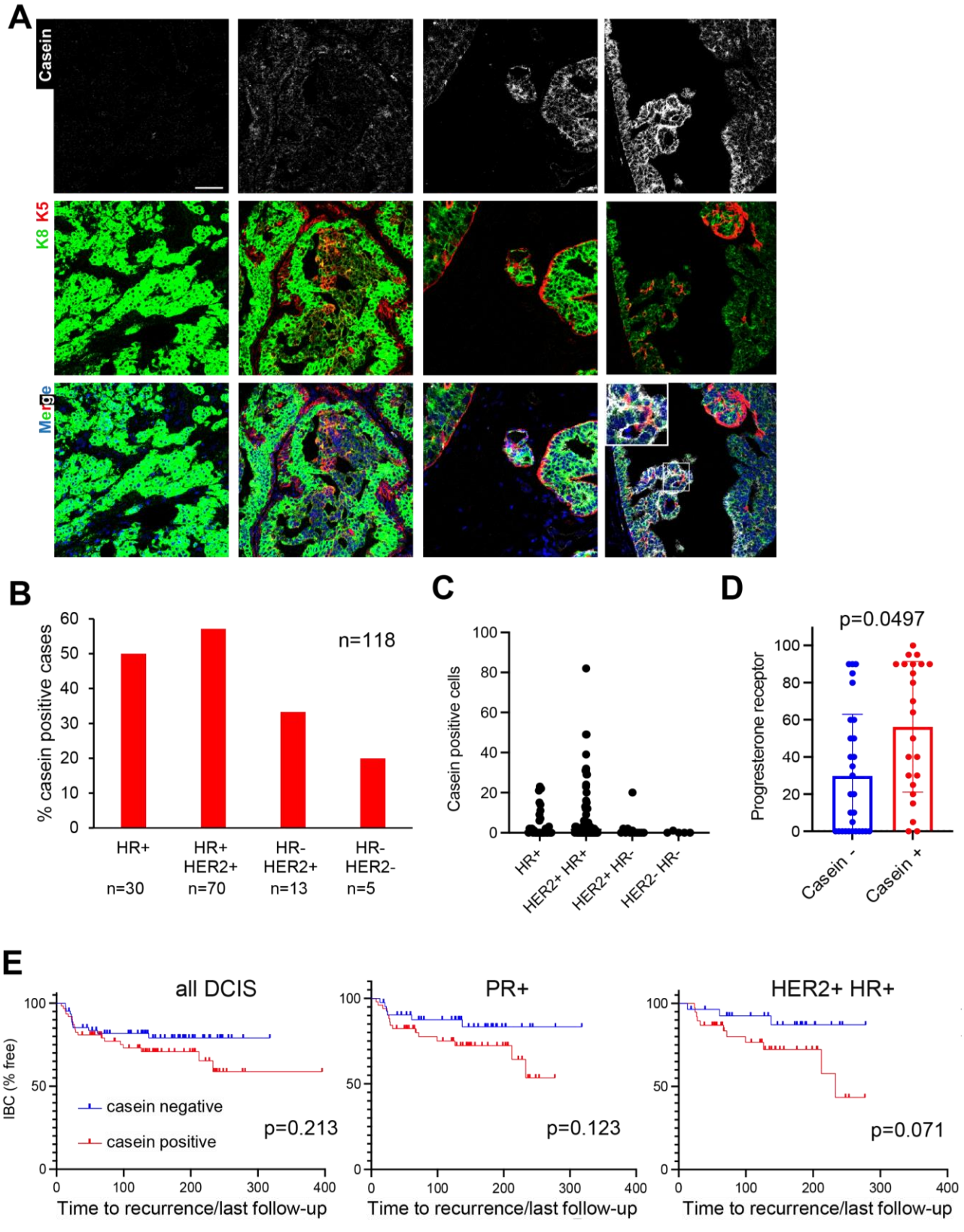
Supplementary Fig. S25. EpiDriver gene signatures co-occur with alveogenesis signatures in DCIS and IDC tumors. A, Average expression of DGO_Lactation and GOBP_lactation gene signatures from 57 DCIS and 313 invasive tumors**. B,** Heatmap of 57 DCIS and 313 IDC profiled by microarray. Mean z-score for each signature in each tumor is shown as vertical lines. Tumors are clustered using Euclidean distance. **C**, Pearson correlation between signature expression in DCIS and IDC tumors.

Supplementary Fig. S26



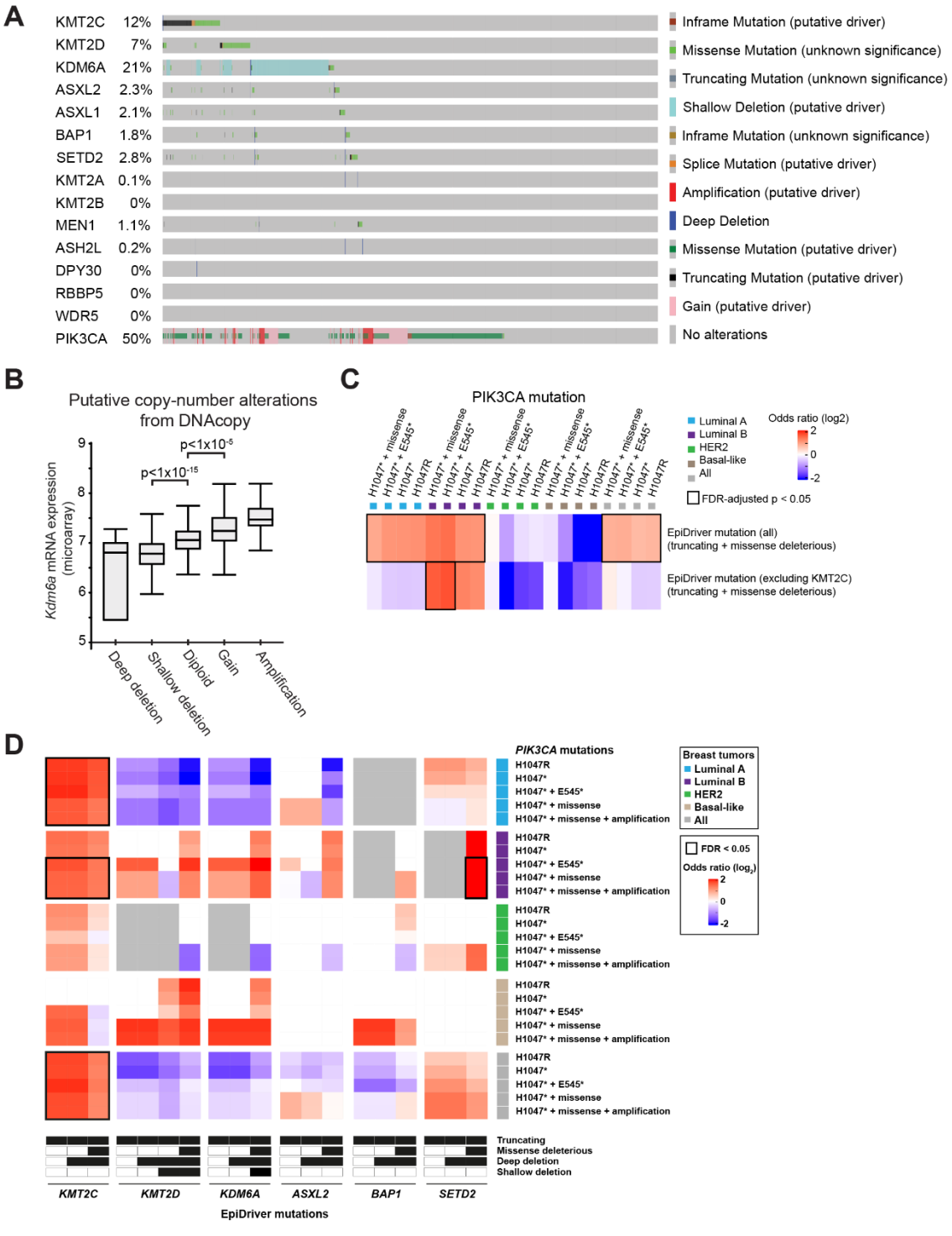
Supplementary Fig. S26. Casein is expressed in some human breast tumors but not normal breast or other organs. A, Representative immunohistochemistry images of normal breast and control tissues stained with casein CSN1S1 antibody (brown). **B,** Representative immunohistochemistry images of various breast cancers stained with casein CSN1S1 antibody (brown).

Supplementary Fig. S27



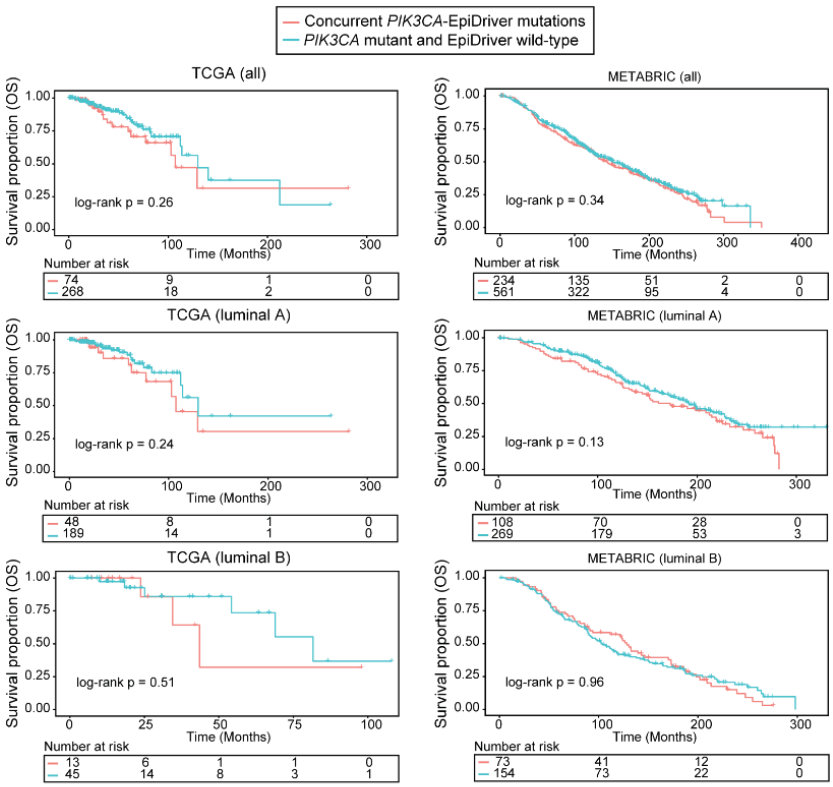
Supplementary Fig. S27. Recurrence-free survival and marker expression in DCIS tumors.
A, Representative images of casein, Krt8 and Krt5 expression in DCIS samples by IMC. Scale bar = 100 µM. **B,** Percentage of casein positive DCIS cases in an independent DCIS cohort stratified by breast cancer subtype. **C,** Percentage of casein positive cells within DCIS within this cohort by the indicated histological marker. **D,** Percentage of progesterone receptor positive DCIS cases stratified by casein expression. **E,** Kaplan-Meier plots showing recurrence-free survival as percentage in the indicated DCIS cohort.

Supplementary Fig. S28



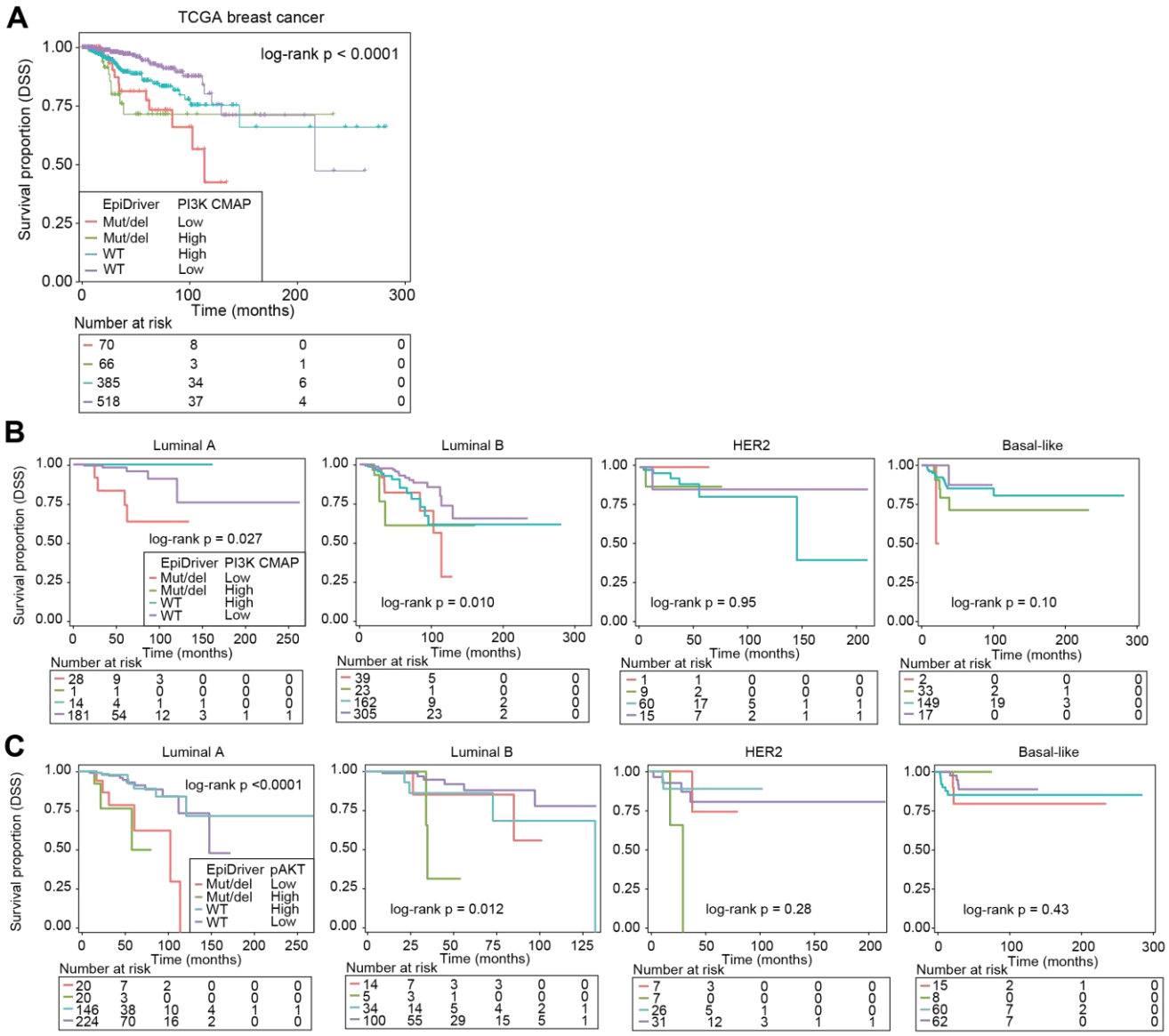
Supplementary Fig. S28. Epigenetic regulators are mutated in human tumors.
A, Oncoprint profiles of epigenetic genes identified as tumor suppressors in our screen compared to *KMT2A/B*, *MEN1*, and COMPASS-like core genes (*ASH2L*, *DPY30*, *RBBP5*, and WDR5), as well as alterations in *PIK3CA*. Shallow deletion shown only for *KDM6A*. Amplification and gain are shown only for *PIK3CA*. Interestingly, the methyltransferases *KMT2A/B* and the gene coding for an interaction partner, *MEN1*, and the genes coding for proteins common to all COMPASS complexes are not mutated in breast cancer. **B,** *KDM6A* expression stratified by *KDM6A* copy-number levels. Data for **a,** and **b,** were obtained from METABRIC and accessed through cBioPortal. **C and D,** Co-occurrence analysis of *PIK3CA* and EpiDriver (single gene analysis **(C)** or pooled together **(D)**) mutations in the combined breast cancer dataset of TCGA and METABRIC (total tumors n = 3,494). The heatmap shows the co-occurrence odds ratios (log_2) across breast cancer subtypes or all tumors considered, and significant (FDR-adjusted $p < 0.05$) associations are highlighted by black dashed rectangles. This analysis considered the following categories according to *PIK3CA* and EpiDriver status: presence of *PIK3CA* H1047R mutation; any H1047 amino acid change; any H1047 or E545 amino acid change; any *PIK3CA* missense mutation; any missense mutation and genomic amplification; or wild-type; and for each EpiDriver (*ASXL2*, *BAP1*, *KMT2C*, *KMT2D*, *KDM6A*, and *SETD2*), presence of a truncating mutation; missense change classified as deleterious; or wild-type. Examination the complete space of *PIK3CA*-EpiDriver co-occurrences, regardless of the number of tumors in each category identified significant (FDR-adjusted $p < 0.05$) associations between *PIK3CA* and *KMT2C* or *SETD2* mutations.

Supplementary Fig. S29



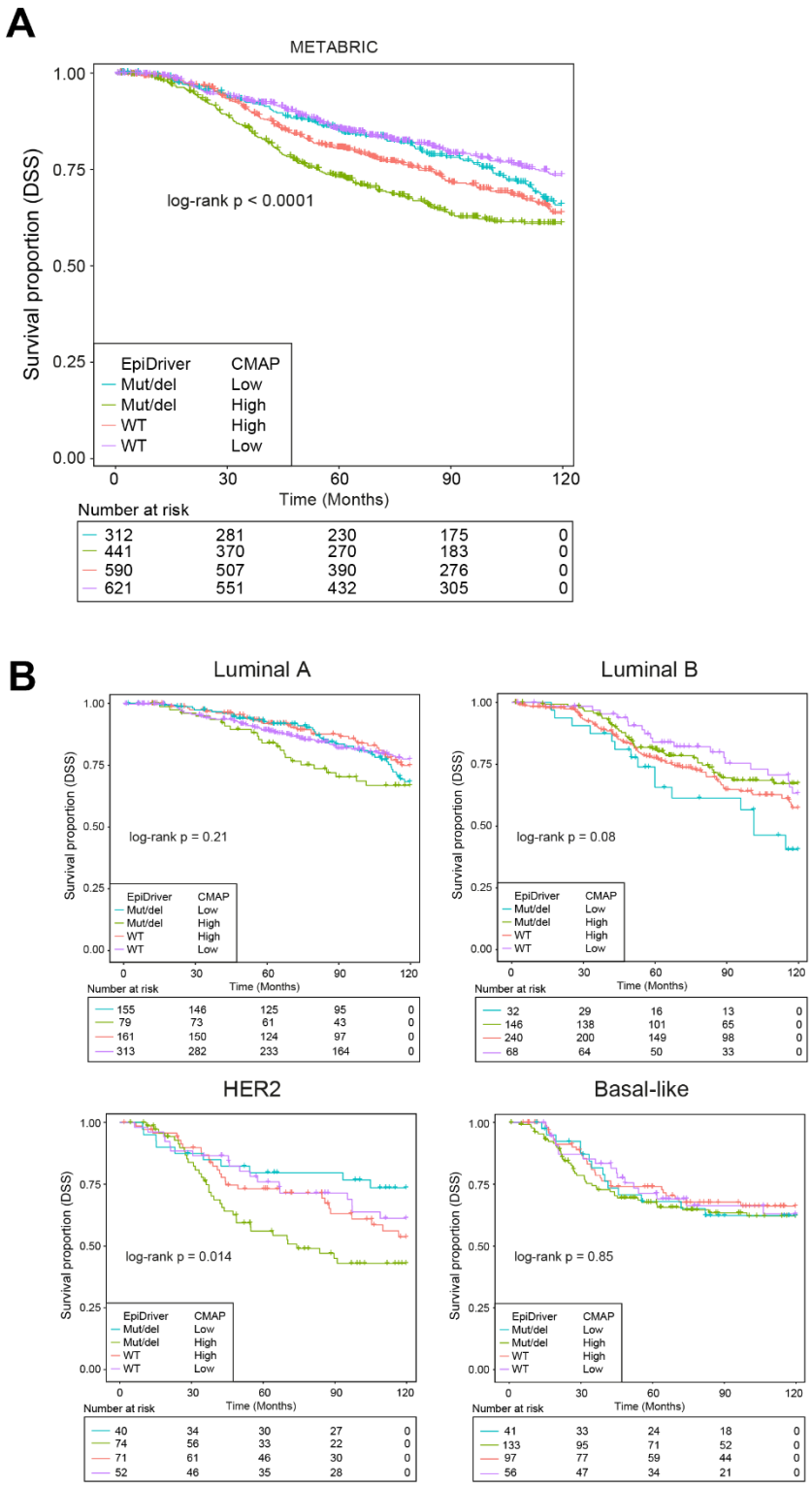
Supplementary Fig. S29. Co-occurrent *PIK3CA* and EpiDriver mutations tend to be associated with poorer survival of luminal A tumors. Kaplan-Meier plots showing OS of TCGA (left panels) and METABRIC (right panels) breast cancers (all, luminal A and B; top, middle, and bottom panels, respectively) stratified by *PIK3CA* mutant and EpiDriver mutant or wild-type status. The number of patients at risk and log-rank test p values are shown.

Supplementary Fig. S30



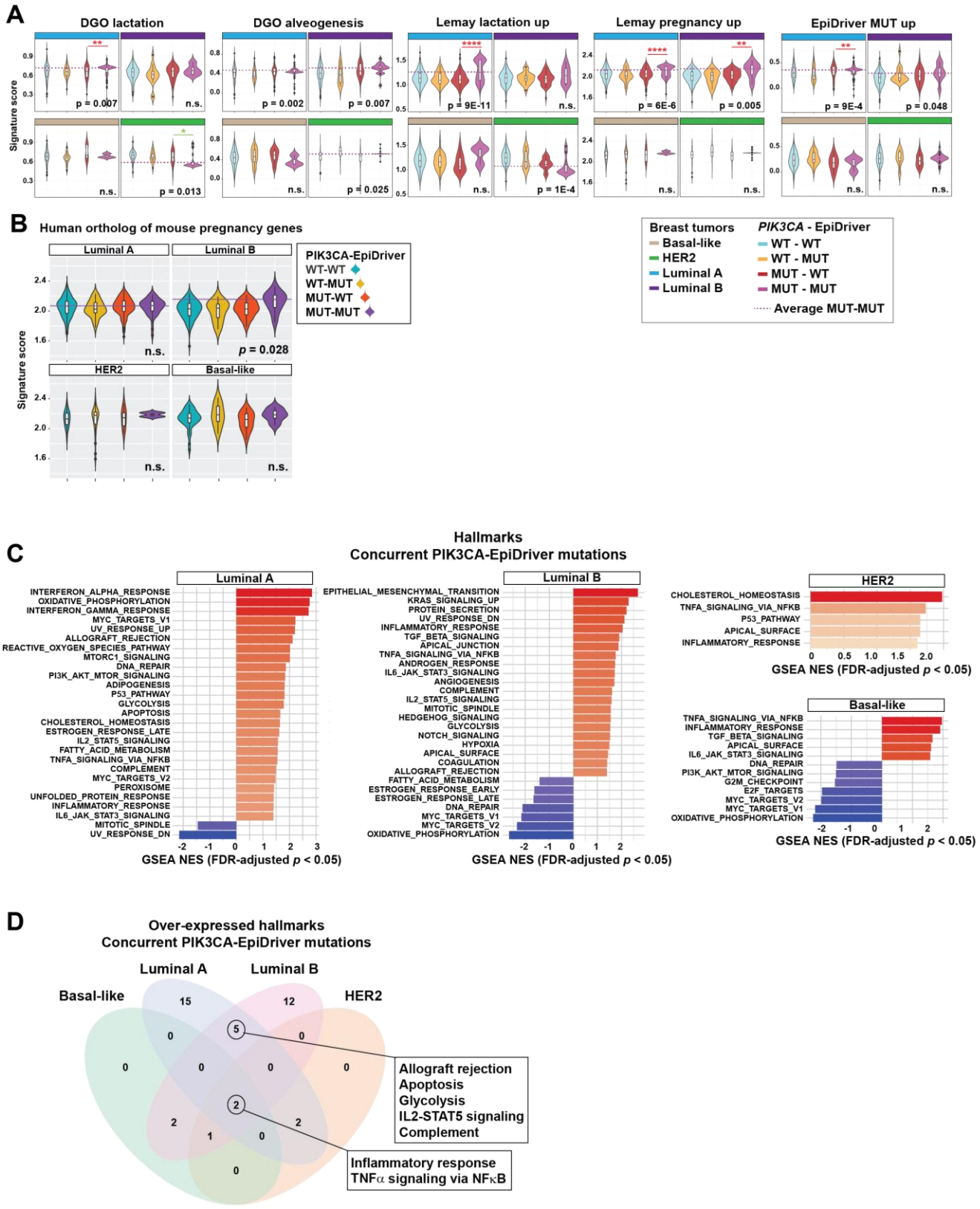
Supplementary Fig. S30. PI3K signaling and EpiDriver alterations stratifies in human breast tumors in the TCGA dataset. A, Kaplan-Meier plot of DSS of breast cancer patients from the TCGA cohort stratified by CMAP signature (Connectivity Map Analysis of PI3k, a signature of PI3K pathway activation) and EpiDriver mutations. The number of patients at risk and long-rank p value are shown. **B,** Kaplan-Meier plot of DSS of breast cancer cases stratified by tumor subtype, and PI3K CMAP transcriptional signature and EpiDriver mutations. There was a limited set of events in luminal A tumors to evaluate this association, but this subtype showed a significant higher proportion of tumors with high PI3K and EpiDriver mutation (odds ratio = 12.2; $p < 2x10\text{-}16$). **C,** Kaplan-Meier plot of DSS of breast cancer cases stratified by tumor subtype, and phospho-Ser473 AKT and EpiDriver mutations.

Supplementary Fig. S31



Supplementary Fig. S31. PI3K signaling and EpiDriver alterations stratifies in human breast tumors in the METABRIC. A, Kaplan-Meier plot of DSS of breast cancer patients from the METABRIC cohort stratified by PI3K CMAP transcriptional signature and EpiDriver mutations. The number of patients at risk and long-rank p value are shown. **B,** DSS analysis in tumors stratified by tumor subtype, and PI3K CMAP transcriptional signature and EpiDriver mutations.

Supplementary Fig. S32



Supplementary Fig. S32. Concurrent *PIK3CA*-EpiDriver mutations and impact on defined gene expression signatures. **A**, Violin plots showing the expression of indicated signatures in TCGA tumors with concurrent *PIK3CA*-EpiDriver mutations relative to other groups stratified by tumor subtype. The Kruskal-Wallis test p value is shown for each signature comparison among four *PIK3CA*-EpiDriver- groups in each subtype (n.s., not significant). DGO and Lemay signatures as well as the gene set homologous to upregulated genes in the EpiDriver/Pik3ca-mutant mouse mammary tumors are shown. The asterisks indicate significant differences between MUT-WT and MUT-MUT tumors (*PIK3CA*-EpiDriver status; Mann-Whitney test * $p < 0.05$; ** $p < 0.01$; **** $p < 0.0001$). Green and red asterisk indicates significant down and upregulated gene sets, respectively. The average value of the group with concurrent *PIK3CA-EpiDriver* mutations is depicted by a horizontal lilac line. **B**, Violin plots showing the overexpression of a mouse pregnancy up-regulated gene set in tumor subtypes with concurrent *PIK3CA* and EpiDriver mutations relative to other groups. Gene set was derived from Lemay *et al.*, 2009 (full reference in main text). The significance of the Kruskal-Wallis test (n.s., not significant) in shown in each subtype. The average value of the group with concurrent *PIK3CA* and EpiDriver mutations is depicted by a horizontal lilac line. Overexpression of this pregnancy up-regulated gene set is linked to concurrent *PIK3CA* and EpiDriver mutations in luminal B tumors. **C**, Hallmark gene sets positively (red) or negatively (blue) associated (FDR-adjusted $p < 0.05$) with concurrent *PIK3CA* and EpiDriver mutations in each tumor subtype. The normalized enrichment score (NES) of GSEA is shown in the X-axis. **D,** Venn diagram showing the numbers of over-expressed (FDR-adjusted $p < 0.05$) hallmark sets in tumors with concurrent *PIK3CA* and EpiDriver mutations across the subtypes. The common overexpressed sets between luminal A and B, and between the four major tumor subtypes are detailed (n = 5 and 2, respectively).