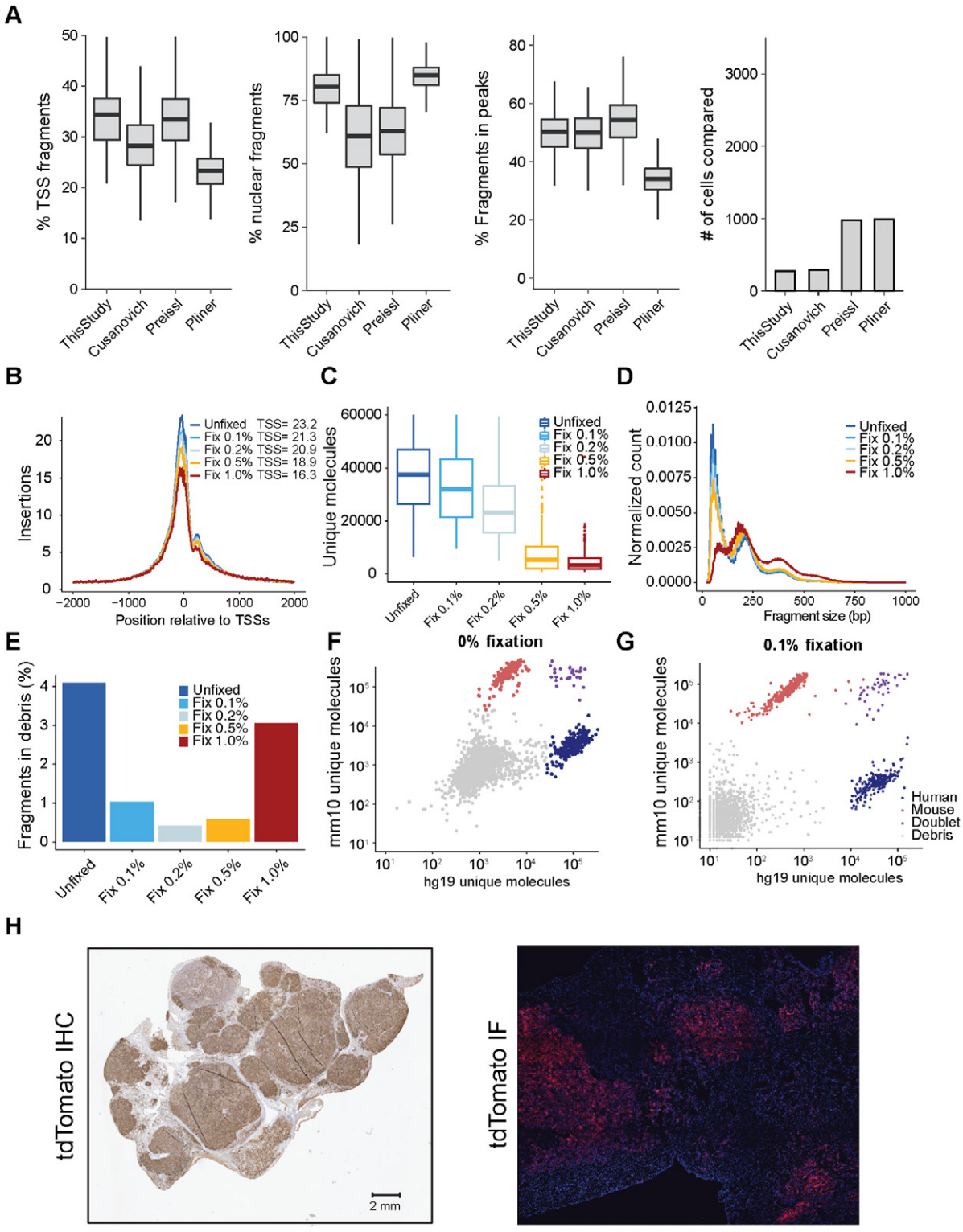

Figure S1. Related to Figure 1. (**A**) Quality control (QC) metrics for sciATAC-seq in 3T3/GM12878 species-mixing experiment compared to published data (Cusanovich et al., 2015; Pliner et al., 2018; Preissl et al., 2018). QC includes percent fragments in TSS, percent nuclear fragments, and percent fragments in peaks as assessed in the GM12878 fraction of experiment. Number of cells analyzed is plotted in bar graph (right). Box intervals represent 25% and 75% bounds. (**B**) Tn5 insertions at

positions relative to TSS across fixation conditions (unfixed to 1%) in GM12878 cells. (**C**) Number of fragments across fixation conditions (unfixed to 1%) in GM12878 cells. Box intervals represent 25% and 75% bounds. (**D**) Histogram of fragment sizes across fixation conditions (unfixed to 1%) in GM12878 cells. (**E**) Percentage of fragments associated with debris across fixation conditions (unfixed to 1%) in GM12878 cells. (**F**) Species-mixing plot demonstrating debris contamination in unfixed vs (**G**) fixed cells in pooled 3T3/GM12878 cells. (**H**) IHC for tdTomato, demonstrating that tumors maintain expression of tdTomato in late-stage tumors and immunofluorescence (IF) from frozen tumor section, scale bar 2 mm.

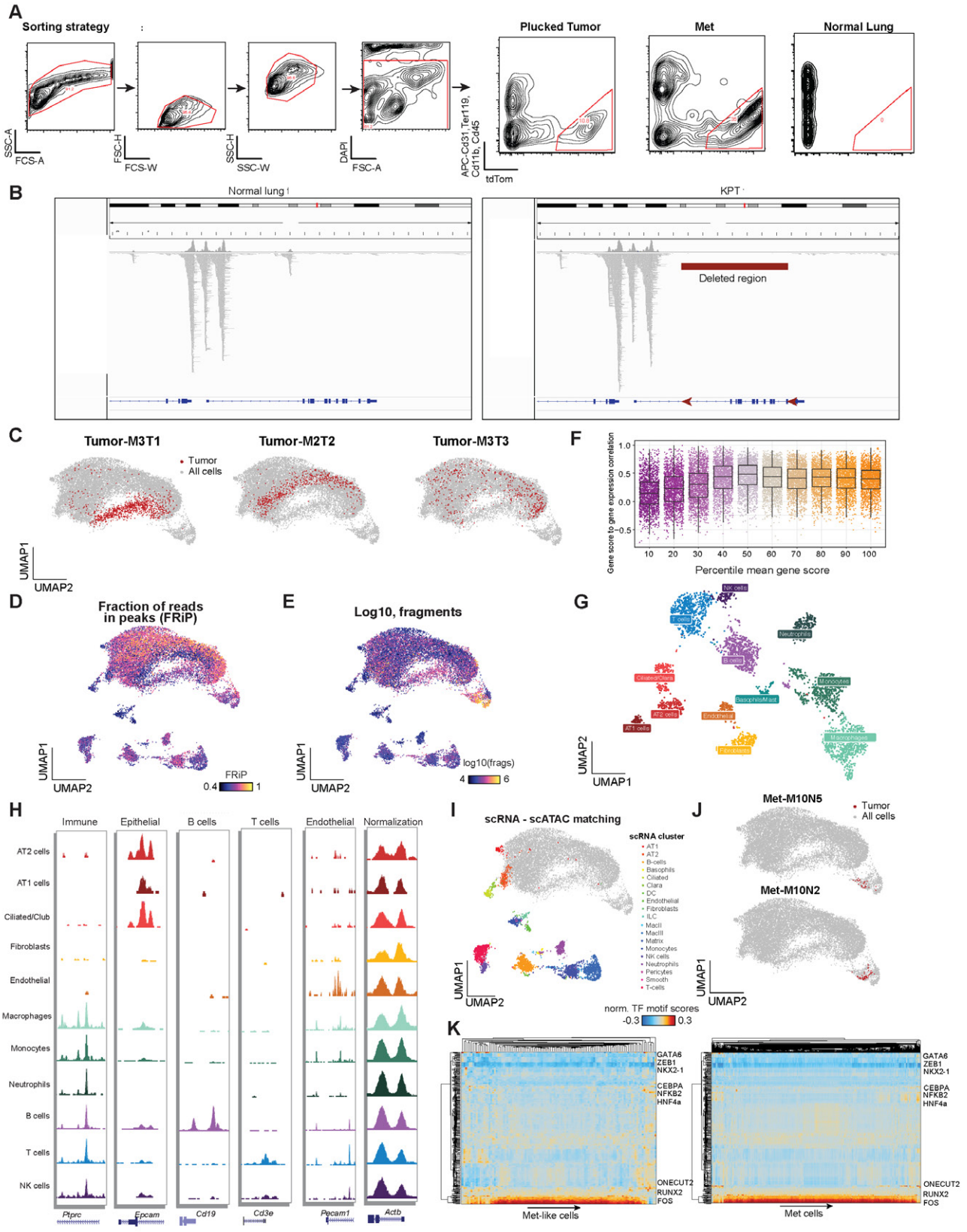

Figure S2. Related to Figure 2. (**A**) FACS plots demonstrating sorting strategy for KPT cancer cells and normal cells. Immune and endothelial FACS-depletion strategy is shown (TER119-APC, CD31-

APC, CD11b-APC, and CD19-APC). Enriched tdTom^{+} positive populations were utilized for downstream sciATAC-seq. (**B**) Reads mapped to *Trp53* locus in normal and KPT cancer cells to assess tumor purity. Red bar mapped to the expected *Trp53* deleted region. p53 track: region chr11:69,573,794-69,598,424; Downsampling window was 50,000 and # of reads per window was 4,000. (**C**) Exemplar cells sorted from individual KPT tumors profiled in this study. (**D**) FRIP overlaid on UMAP for all single-cells profiled (n = 17,274). (**E**) Library size as log10 fragments mapped to UMAP plot (library size) for all single-cells profiled (n = 17,274). (**F**) Percentile mean gene score associated with corresponding RNA expression in human LUAD patient data (Corces et al., 2018). (**G**) Normal lung populations identified by Louvain clustering, identified by gene scores and motifs (**H**) Aggregated bulk tracks across UMAP clusters, with peaks at genes associated with specific cell types. *Actb* tracks included for normalization across aggregated tracks. (**I**) Single-cell RNA-seq matching to normal sciATAC-seq populations to confirm cell type identification (Cohen et al., 2018). (**J**) Liver (Met-10N1) and lymph node (Met-10N2) metastatic cells derived from the same mouse (in red). (**K**) Hierarchical clustering of individual primary met-like cancer (left) or metastatic cells (isolated from lymph nodes) (right) with TF motif scores across samples.

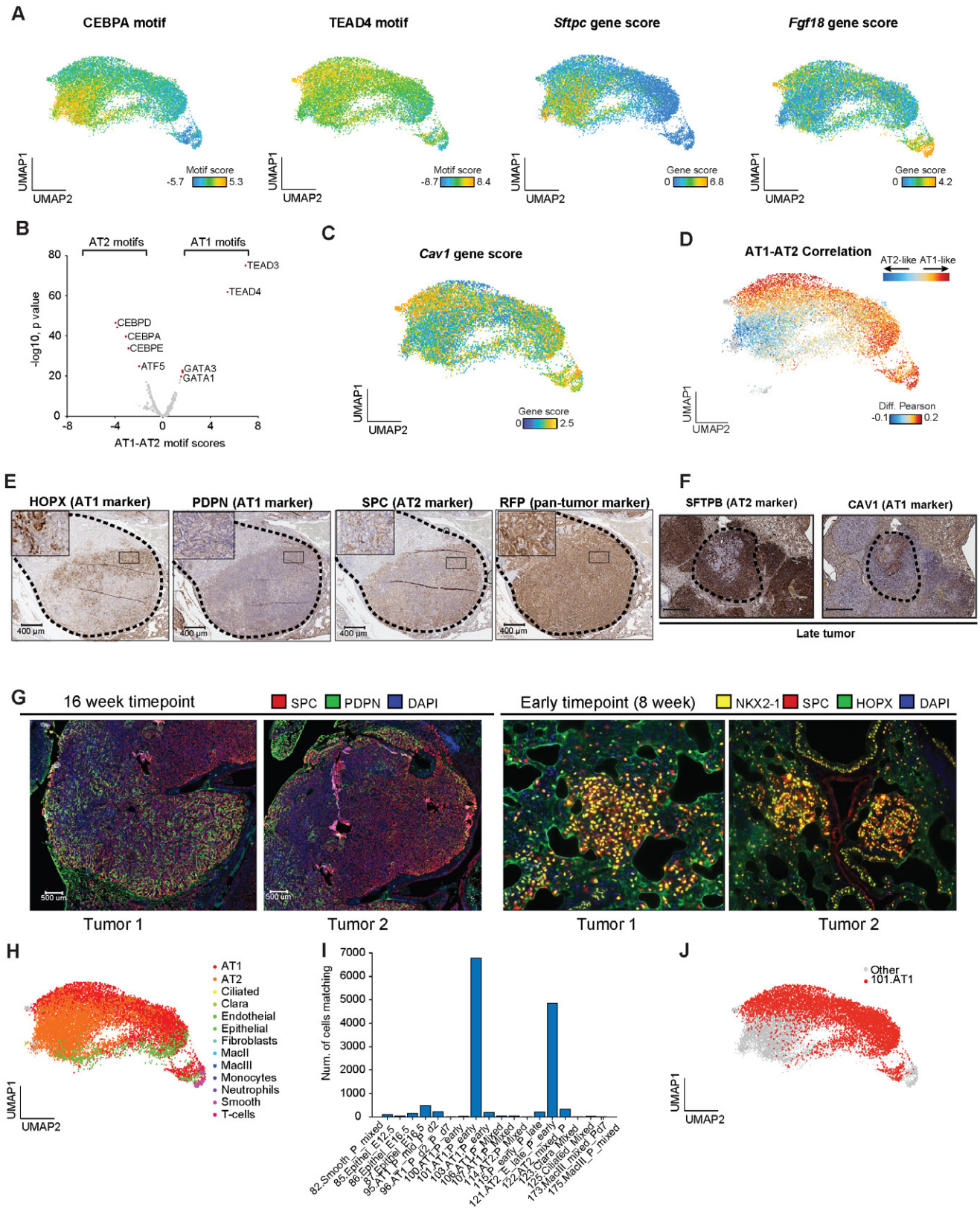

Figure S3. Related to Figure 3. (**A**) Top motif and gene scores associated with AT2 and AT1 cells (CEBPA, *Sftpb*; AT2, TEAD4, *Fgf18*; AT1). (**B**) Volcano plot of differential motif accessibility in normal AT2 and AT1 cells. (**C**) Cav1 gene score associated with AT1 cells. (**D**) Correlation of KPT cancer cells

to AT2 and AT1 gene scores. (**E**) IHC of outlined KPT tumor; staining for HOPX (AT1 marker), PDPN (AT1 marker), SPC (AT2 marker), and RFP (pan-tumor marker). (zoom 400 μm; inset 50 μm). (**F**) IHC of outline late-stage tumor for SFTPB (AT2 marker) and CAV1 (AT1 marker): CAV1 (500 μm), SFTPB (500 μm). (**G**) Multiplexed IHC staining of KPT tumors at 16 weeks and 8 weeks; 16 week time point (PDPN: green; SFTPC: red; DAPI: blue) (scale bar; (left) Tumor 1: 500 μm; Tumor 2: 500 μm; (right) Tumor 1; 25x, 50 μm; Tumor 2: 20x, 50 μm). 8-week time point (NKX2-1: yellow; HOPX: green; SFTPC: red; DAPI: blue). (**H**) Matching of scATAC-seq single-cell identities to published scRNA-seq (Cohen et al., 2018). Assignment of closest associated cell type in the tumor space. (**I**) Matching of lung development data to scATAC-seq cells. Time point labels were assigned based on most associated cells from specified time points (embryonic (E): E12.5, E16.5, E_late; postnatal (P): P_early, P_mid, P_day2, P_day7). Populations found across multiple time points were termed mixed (all timepoints), P_mixed (for mostly postnatal timepoints), or E_mixed (mixed embryonic timepoints) (Cohen et al., 2018). (**J**) Matching of 101.AT1.P_early overlaid on UMAP clustering data to most similar KP cancer cell.

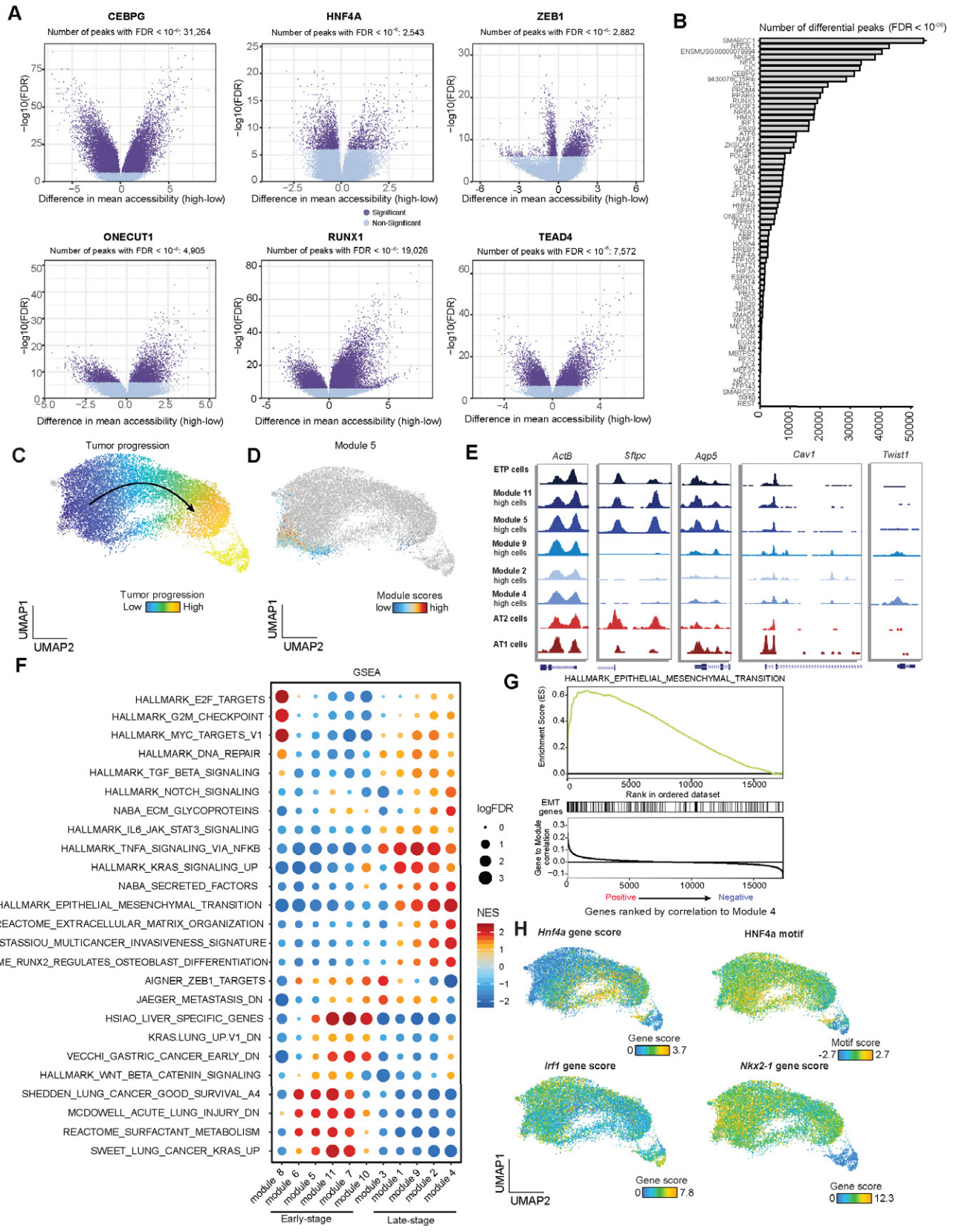

Figure S4. Related to Figure 4. (**A**) Scatter plots across TFs identifying top differential peaks from chromatin accessibility data. Significant peaks are highlighted in dark purple (FDR < 10^{-6}). Example TFs

plotted including CEBPG, HNF4A, ZEB1, ONECUT1, RUNX1, and TEAD4. (**B**) Number of differential peaks for each motif assessed in the defined modules (FDR < 10^{-06}). (**C**) UMAP representation of single cells demonstrating derived tumor progression score used for motif heatmap, Figure 4A, with a line fit indicating tumor progression from early to late/metastatic cells. (**D**) UMAP projection of ETP cells, with cells colored by module 5 accessibility. (**E**) Tracks of lung identity genes across module high cells and 8-week time point (ETP; early time point) cells compared to normal AT2 and AT1 cell types. (**F**) GSEA of genes ranked by correlation of gene scores to accessibility scores for each module, highlighting enrichments of relevant gene sets. Dots colored for positive (red) or negative (blue) enrichment of gene set among the corresponding module-associated gene ranking; NES: Normalized Enrichment Score. (**G**) GSEA enrichment plot highlighting positive enrichment of EMT signature among genes correlated with module 4 accessibility (FDR q $\leq$ 0.001). (**H**) Gene scores and motifs for select gene markers overlaid on UMAP clustering.

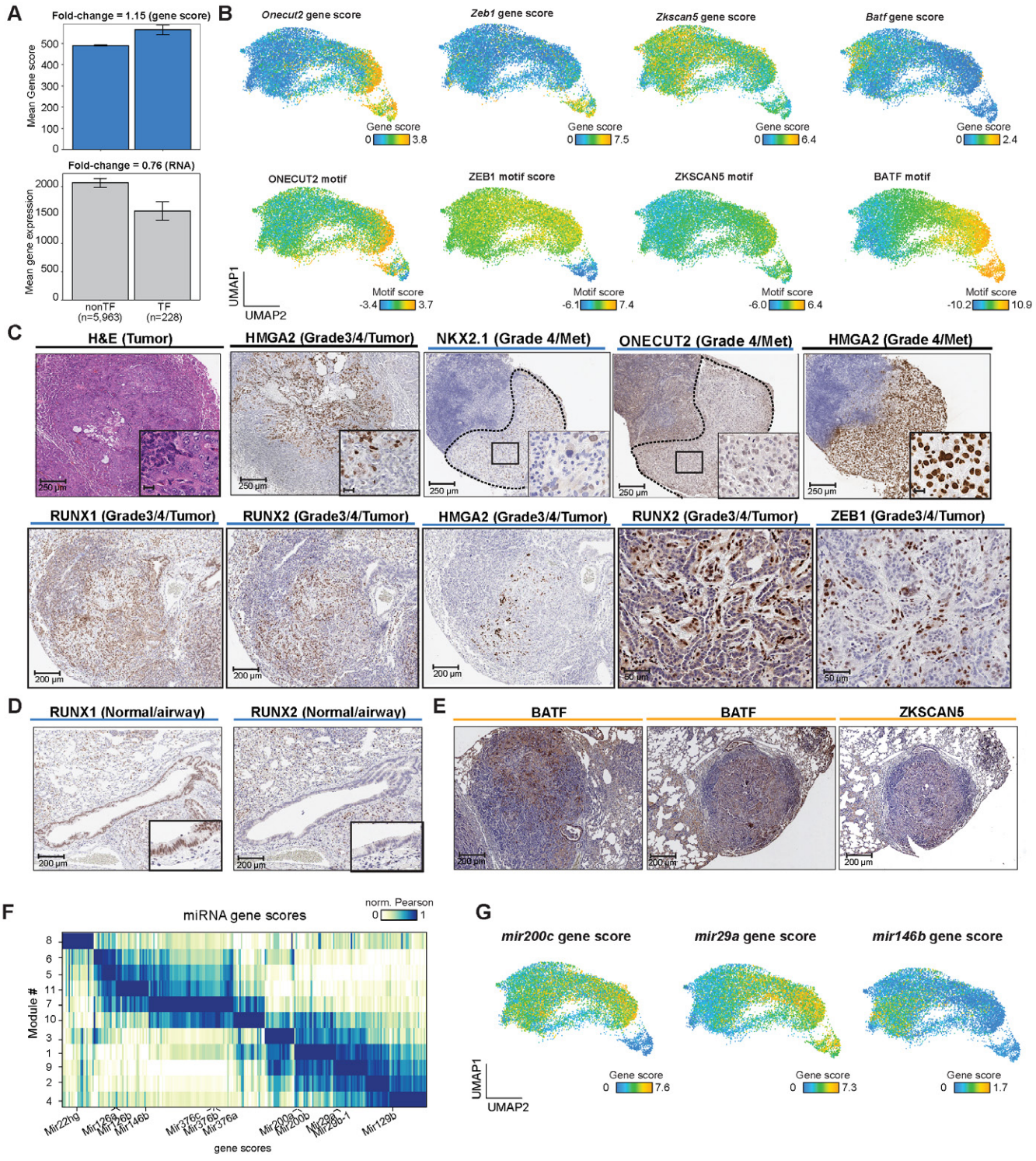

Figure S5. Related to Figure 5. (**A**) Fold-change of TF- and non-TF-encoding genes, comparing gene expression to gene scores in TCGA LUADs (n = 21). Bar plots show the mean (± s.e.) gene score or gene expression levels for either gene category. (**B**) Gene scores (top row) and motif scores (bottom row) across TF activators and repressors. Exemplary TFs were selected due to enrichment across modules. (**C**) Additional IHC from KP tumors presented in the main text, scale bar 250 μm, inset scale bar 50 μm. (**D**) RUNX1 and RUNX2 expression in the airway, scale bar 200 μm, inset 50 μm. (**E**) IHC for repressor BATF, ZKSCAN5 (present mostly in early-stage tumors), scale bar 200 μm. (**F**) Panel of differential gene scores for miRNAs. (**G**) Exemplar miRNAs gene scores painted on UMAP.

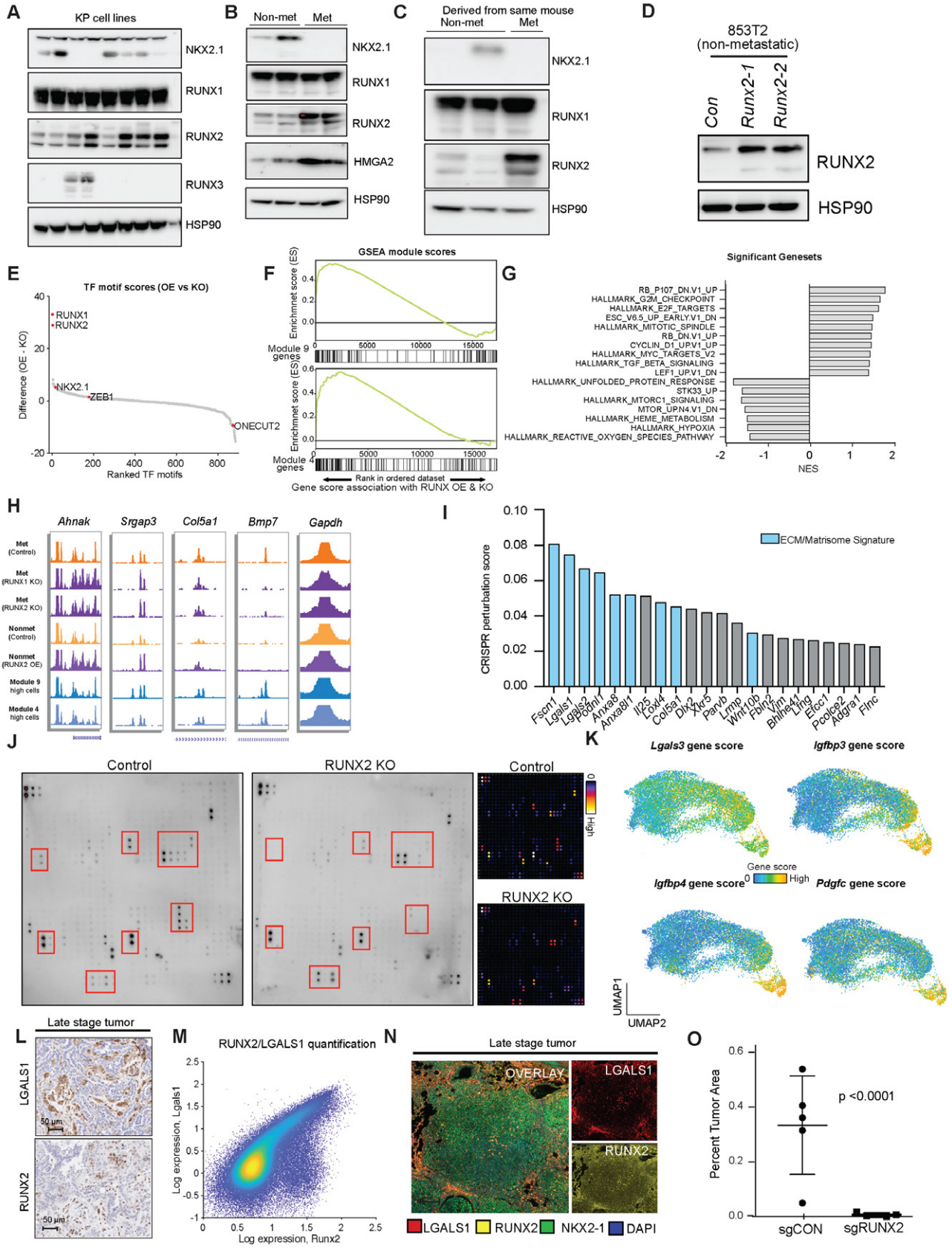

Figure S6. Related to Figure 6. (**A**) Expression of RUNX family members (RUNX1, RUNX2, RUNX3) in KP tumor-derived cell lines (853T2, 860T1, 860T3, 1183T3, 1183T4, 932T2, 932T3, 932LN). NKX2.1

staining delineates cell lines as having a more “non-metastatic” or “metastatic” phenotype. (**B**) RUNX family expression in cell lines used for CRISPR and CRISPRa experiments; (non-met: 853T2, 860T1; met: 1183T3, 860T3). (**C**) RUNX expression in cell lines isolated from plucked tumors (779T1, 779T2) and a lymph node (779LN). (**D**) Western blot demonstrating activation of *Runx2* in 853T2 Cas9 cell line. (**E**) TF motif score comparison in OE and KO cells. (**F**) GSEA with respect to changes in RUNX motif accessibility from RUNX2 perturbation experiments *in vitro* reveals enrichment of module 4 and module 9-associated single-cell gene signatures. (**G**) Normalized enrichment scores (NES) from GSEA for select gene sets and gene ontology (GO) terms that are associated with RUNX2 OE and KO cells. (**H**) Tracks highlighting differential chromatin accessibility at target genes identified by association with RUNX motif and module genes. (**I**) Genes plotted in decreasing order of perturbation score for top RUNX-regulated genes associated with module 2 (late-stage) using gene set enrichment analysis (GSEA). Genes associated with a matrisome signature are highlighted in blue. (**J**) Extracellular protein secretion array in 1183T3 Cas9 control and RUNX2 KO cell lines. Arrays quantified using Image J protein array plugin analysis. (**K**) UMAP of KPT tumor and normal lung cells highlighting gene scores identified in extracellular array computed using sciATAC-seq data for relevant cluster-specific genes. (**L**) IHC showing LGALS1 and RUNX2 expression in a late-stage KPT tumor, scale bar 50 μm. (**M**) RUNX2 and LGALS1 quantification of single-cells from multiplexed IHC image in **Figure 6G**. (**N**) Multiplexed IHC with rare $RUNX2^{+}$ cells that overlap with LGALS1 localization (scale bar; 5.5x, 200 μm). (**O**) Tumor area as a percentage of total lung area as quantified by Aiforia in an exemplar tail vein experiment with n = 5 mice per arm.

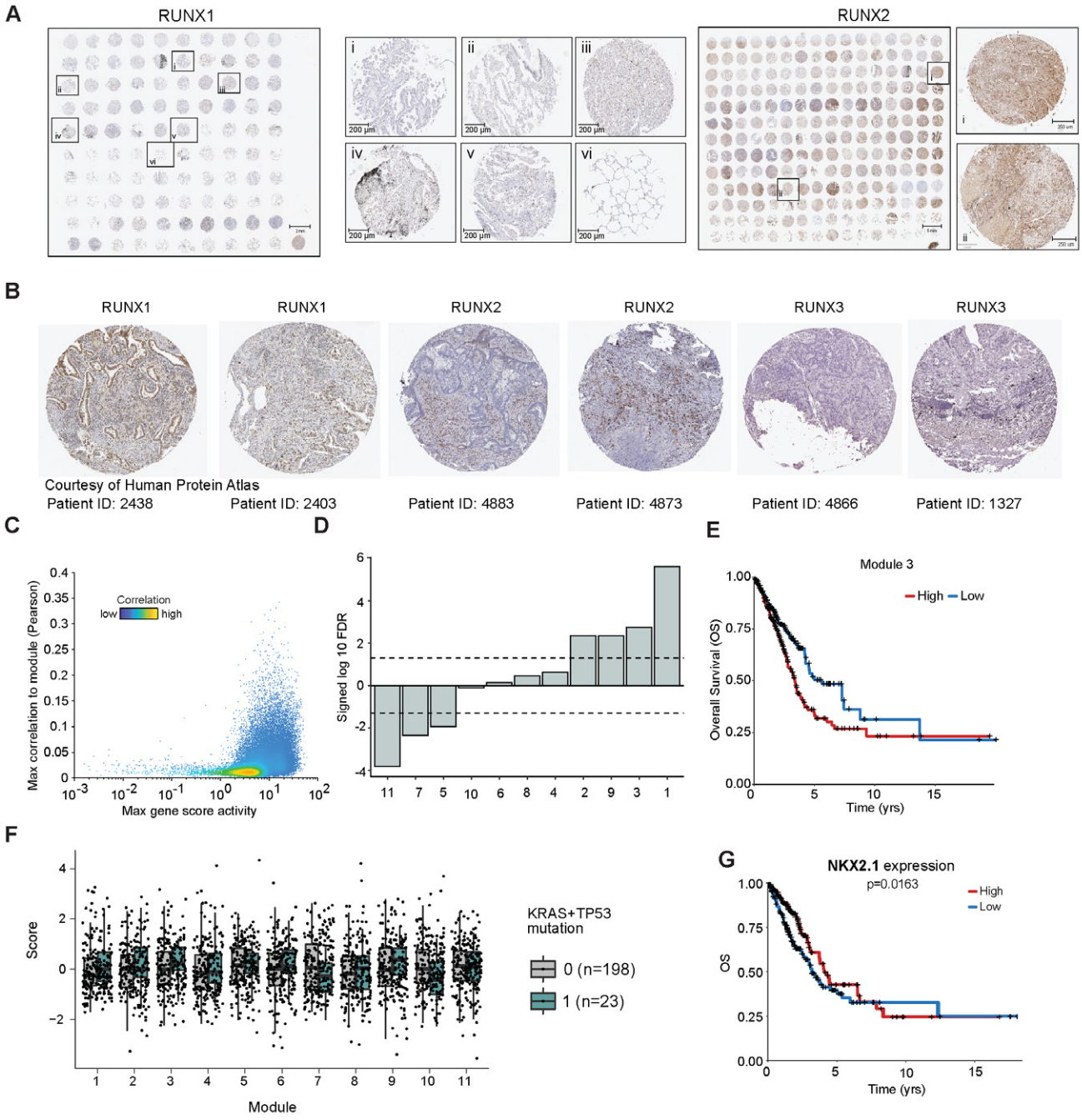

Figure S7. Related to Figure 7. (**A**) LUAD tumor microarray (TMA) map stained with RUNX1 (left, center). LUAD TMA map from RUNX2 stain, highlighted RUNX2 positive (i) primary tumor and (ii) metastatic tumor (right). (**B**) RUNX tumor staining from the Human Protein Atlas (Uhlén et al., 2015). Files can be accessed at the links found in **STAR Methods**. (**C**) Gene-module assignments based on maximum Pearson correlation coefficient per gene across modules (from module accessibility score to gene score correlations, Fig. 7B) versus the mean gene score. The top 200 genes per module were used to define module-specific genes (**D**) Significance of module-associated gene expression with overall survival (OS) using a Cox proportional hazard (coxph) test in TCGA LUADs (n=506). Positive values denote decreased survival, negative values denote increased patient survival. (**E**) Kaplan-Meier plot for overall survival (OS) with respect to module 3-associated gene expression in TCGA LUADs.

(**F**) Comparison of standardized module scores in patients with (group 1) and without (group 0) *KRAS* and *TP53* somatic mutations in TCGA LUADs. Median plotted and box intervals represent 25% and 75% bounds. (**G**) Kaplan-Meier plot for OS based on NKX2-1 RNA expression in TCGA LUADs. Patients split into High or Low groups based on median NKX2-1 expression.