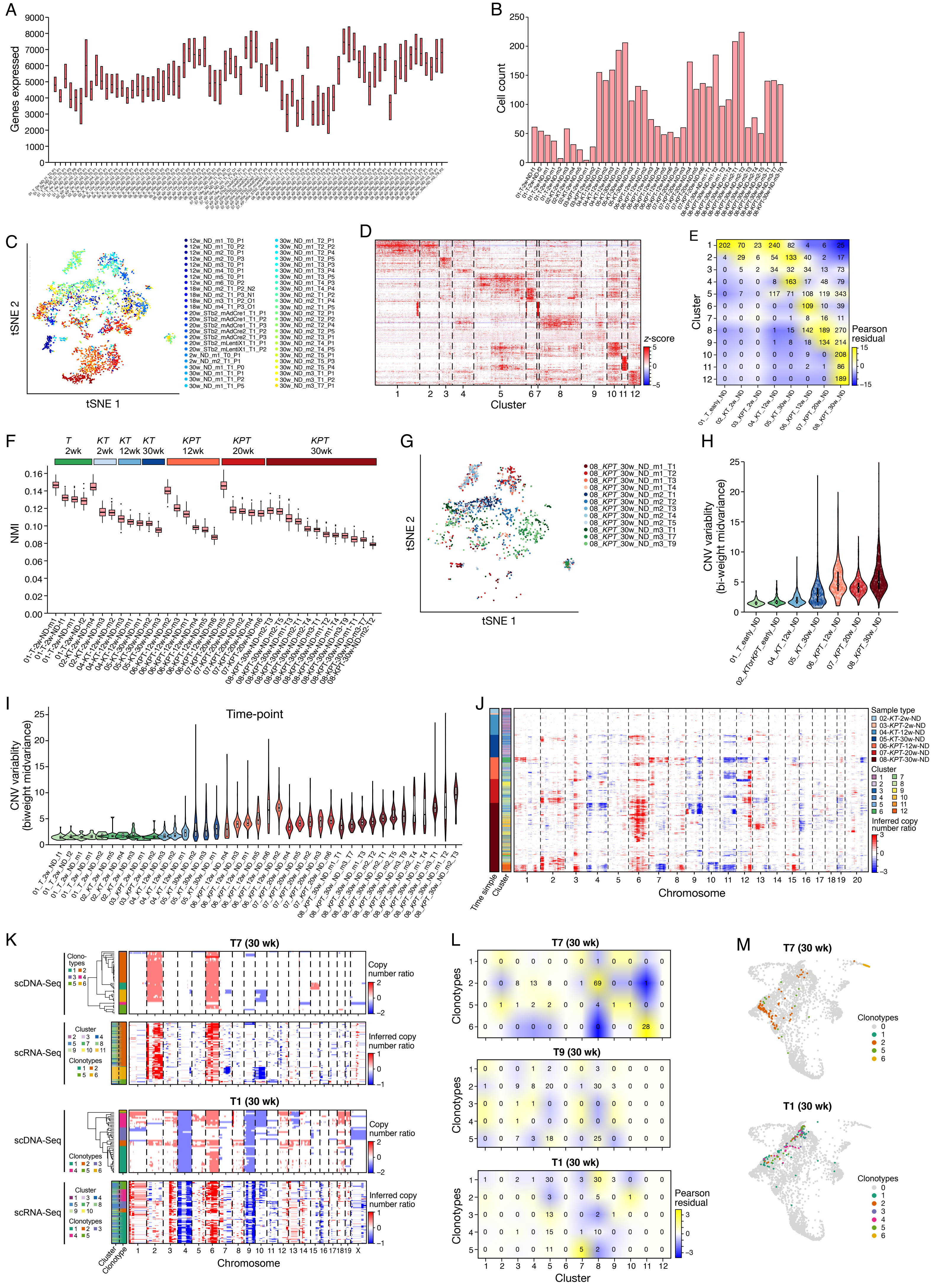

A
Genes expressed
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
B
Cell count
250
200
150
100
50
0
C
tSNE 2
tSNE 1
12w_ND_m1_T0_P1
12w_ND_m1_T0_P2
12w_ND_m2_T0_P1
12w_ND_m2_T0_P3
12w_ND_m3_T0_P1
12w_ND_m4_T0_P1
12w_ND_m5_T0_P1
12w_ND_m6_T0_P2
18w_ND_m2_T1_P2_N2
18w_ND_m2_T1_P3_N1
18w_ND_m3_T1_P2_O1
18w_ND_m4_T1_P3_O1
20w_STb2_mAdCre1_T1_P1
20w_STb2_mAdCre1_T1_P2
20w_STb2_mAdCre1_T1_P3
20w_STb2_mAdCre2_T1_P2
20w_STb2_mAdCre2_T1_P3
20w_STb2_mLentiX1_T1_P1
20w_STb2_mLentiX1_T1_P2
2w_ND_m1_T0_P1
2w_ND_m2_T1_P1
30w_ND_m1_T1_P0
30w_ND_m1_T1_P1
30w_ND_m1_T1_P2
30w_ND_m1_T1_P5
30w_ND_m1_T2_P1
30w_ND_m1_T2_P2
30w_ND_m1_T2_P4
30w_ND_m1_T2_P5
30w_ND_m1_T3_P1
30w_ND_m1_T3_P3
30w_ND_m1_T3_P4
30w_ND_m1_T3_P5
30w_ND_m1_T4_P3
30w_ND_m1_T4_P4
30w_ND_m2_T1_P1
30w_ND_m2_T1_P2
30w_ND_m2_T1_P4
30w_ND_m2_T1_P5
30w_ND_m2_T2_P1
30w_ND_m2_T2_P2
30w_ND_m2_T2_P4
30w_ND_m2_T2_P5
30w_ND_m2_T3_P1
30w_ND_m2_T3_P2
30w_ND_m2_T4_P1
30w_ND_m2_T4_P2
30w_ND_m2_T4_P5
30w_ND_m2_T5_P1
30w_ND_m2_T5_P3
30w_ND_m2_T5_P4
30w_ND_m3_T1_P1
30w_ND_m3_T7_P2
30w_ND_m3_T7_P1
D
z-score
5
0
−5
1 2 3 4 5 6 7 8 9 10 11 12
Cluster
E
Cluster
1 202 70 23 240 82 4 6 25
2 4 29 6 54 133 40 2 17
3 0 5 2 34 32 34 13 73
4 0 0 0 8 163 17 48 79
5 0 7 0 117 71 108 119 343
6 0 0 0 0 0 109 10 39
7 0 0 0 0 1 8 16 11
8 0 0 0 1 15 142 189 270
9 0 0 0 1 8 18 134 214
10 0 0 0 0 0 9 0 208
11 0 0 0 0 0 2 1 86
12 0 0 0 0 0 0 0 189
Pearson residual
15
0
−15
01_T_early_ND
02_KT_2w_ND
03_KPT_2w_ND
04_KT_12w_ND
05_KT_30w_ND
06_KPT_12w_ND
07_KPT_20w_ND
08_KPT_30w_ND
F
T 2wk
KT 2wk
KT 12wk
KT 30wk
KPT 12wk
KPT 20wk
KPT 30wk
NMI
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
G
tSNE 2
tSNE 1
08_KPT_30w_ND_m1_T1
08_KPT_30w_ND_m1_T2
08_KPT_30w_ND_m1_T3
08_KPT_30w_ND_m1_T4
08_KPT_30w_ND_m2_T1
08_KPT_30w_ND_m2_T2
08_KPT_30w_ND_m2_T3
08_KPT_30w_ND_m2_T4
08_KPT_30w_ND_m2_T5
08_KPT_30w_ND_m3_T1
08_KPT_30w_ND_m3_T7
08_KPT_30w_ND_m3_T9
H
CNV variability (bi-weight midvariance)
25
20
15
10
5
0
01_T_early_ND
02_KTorKPT_early_ND
04_KT_12w_ND
05_KT_30w_ND
06_KPT_12w_ND
07_KPT_20w_ND
08_KPT_30w_ND
I
Time-point
CNV variability (biweight midvariance)
25
20
15
10
5
0
J
Sample type
02-KT-2w-ND
03-KPT-2w-ND
04-KT-12w-ND
05-KT-30w-ND
06-KPT-12w-ND
07-KPT-20w-ND
08-KPT-30w-ND
Cluster
1 2 3 4 5 6 7 8 9 10 11 12
Inferred copy number ratio
3
0
−3
Time simple
Cluster
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome
K
T7 (30 wk)
scDNA-Seq
Clonotypes
1 2 3 4 5 6
Copy number ratio
2
0
−2
scRNA-Seq
Cluster
2 3 4 5 7 8 9 10 11
Clonotypes
1 2 5 6
Inferred copy number ratio
1
0
−1
T1 (30 wk)
scDNA-Seq
Clonotypes
1 2 3 4 5 6
Copy number ratio
2
0
−2
scRNA-Seq
Cluster
1 3 4 5 7 8 9 10
Clonotypes
1 2 3 4 5
Inferred copy number ratio
1
0
−1
Cluster
Clonotype
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X
Chromosome
L
T7 (30 wk)
Clonotypes
1 0 0 0 0 0 0 0 1 0 0 0 0
2 0 0 8 13 8 0 1 69 0 0 1 0
5 0 1 1 2 2 0 0 4 1 1 0 0
6 0 0 0 0 0 0 0 0 0 0 28 0
T9 (30 wk)
Clonotypes
1 0 0 0 1 2 0 0 3 0 0 0 0
2 1 0 9 8 20 0 1 30 3 0 0 0
3 0 0 1 1 0 0 0 0 0 0 0 0
4 0 0 1 0 0 0 0 0 0 0 0 0
5 0 0 7 3 18 0 0 25 0 0 0 0
T1 (30 wk)
Clonotypes
1 1 0 2 1 30 0 3 30 3 0 0 0
2 0 0 0 0 3 0 0 5 0 1 0 0
3 0 0 0 0 13 0 0 2 1 0 0 0
4 0 0 1 0 15 0 1 10 0 0 0 0
5 0 0 0 0 11 0 5 2 0 0 0 0
Pearson residual
3
0
−3
1 2 3 4 5 6 7 8 9 10 11 12
Cluster
M
T7 (30 wk)
Clonotypes
0 1 2 5 6
T1 (30 wk)
Clonotypes
0 1 2 3 4 5 6

Figure S1, related to Figure 1. Quality assessment and characterization of scRNA-Seq atlas of mouse LUAD progression. (**A**) Substantial number of detected genes across samples. The distribution of the number of expressed genes post-QC filtering (*y* axis) in the cells from each 96-well plate processed (*x* axis). *T*, *KT* and *KPT*: tumor genotypes as described in the text; T#: tumor identification number; P#: plate number. (**B**) Number of cells (*y* axis) in each sample (*x* axis) that passed quality control (*y* axis, **STAR Methods**) and were retained for all further analyses. (**C**) Reproducibility across plates within a sample type. tSNE (as in **Figure 1D**), with cells (dots) colored by plate ID. (**D**) Top marker genes identified for each cluster. Top 50 genes sorted by area under the ROC-curve (rows) that are differentially expressed in each cluster (column) in **Figure 1D**. **(E,F)** Growing heterogeneity with tumor progression. (**E**) The number of cells and relative enrichment (Pearson's residual, calculated as $(obseved\ number\ of\ cells - expected\ number\ of\ cells)/\sqrt{expected\ number\ of\ cells}$, where the expected value is calculated as the product of row and column marginal probabilities by total cells, color bar) from each cluster (rows, ordered by time of emergence) in each sample type (genotype/time point combination, columns), ordered by progression. (**F**) Transcriptional homogeneity decreases with tumor progression. Normalized Mutual Information (NMI, *y* axis) between cells within each individual sample, ordered temporally within a replicate experiment. Box plots: upper, median, lower quartile of 1,000 bootstrap samples, of 50 cells each, from the indicated time point; whiskers: 1.5 interquartile range. (**G**) Reproducibility of cell states across tumors and mice. tSNE of cell profiles (dots) as in **Figure 1D**, but showing only cells from *KPT* tumors that were individually microdissected at 30 weeks, with cells colored by tumor of origin. (**H,I**) Increased variability of CNVs with tumor progression. Distributions of the variability of CNVs (*y* axis) quantified by the biweight midvariance (a robust measure of scale), as inferred from scRNA-Seq data either across all the cells from each time point/genotype combination (H, *x* axis) or within individual biological replicates or dissected tumors (I, *x* axis), ordered by progression. Time point color code as in **Figure 1C**. (**J**) CNVs and cell clusters do not naively align. CNVs (amplification: red; deletion: blue) across chromosomal positions (columns) inferred for each cell (row) from scRNA-Seq data, marked for time point and cell cluster (as in **Figure 1D**). (**K, L**) More examples of transcriptional heterogeneity not simply following genetic clonotype (**K**) Congruence between CNV profiles inferred from scDNA-Seq and scRNA-Seq. CNVs shown (as in **Figure 1H,I)** for single cells (rows) of two individually microdissected *KPT* tumors at 30 weeks profiled by scDNA-Seq (top) or scRNA-Seq (bottom). Left color bar: Predominant clonotypes identified from scDNA-Seq (top) and assigned to scRNA-Seq cells (bottom). Far left color bar in scRNA-Seq panels: cell cluster membership as in **Figure 1G**. (**L**) Matrices showing clonotype distribution across the transcriptionally distinct clusters. Relative enrichment (Pearson's residual) of clonotypes in different clusters (*x* axis) in *KPT* LUAD tumors. **(M)** Clonotypes match multiple transcriptional states. PHATE maps as in **Figure 1D**, colored by clonotype for each of the two tumors.

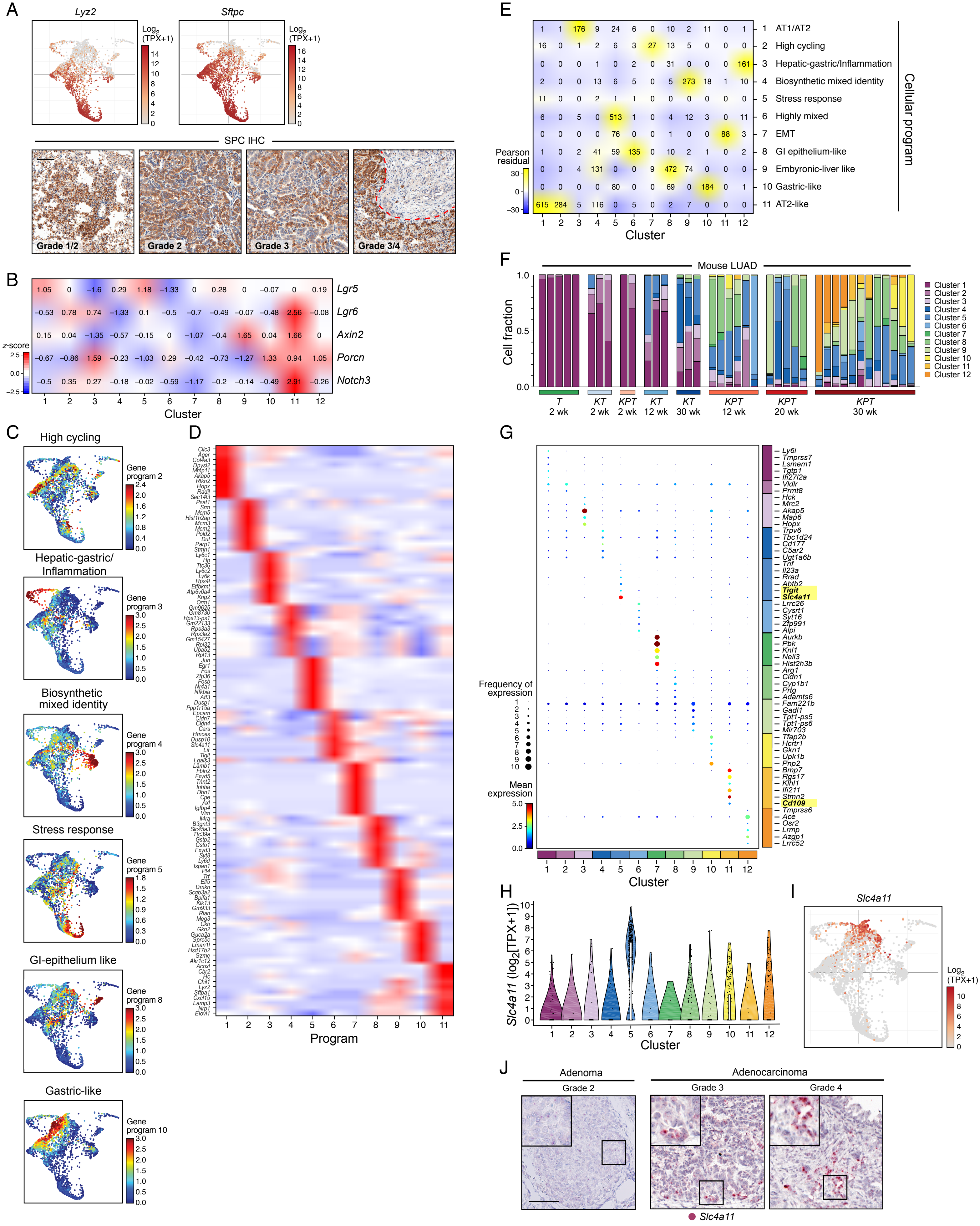

A
Lyz2
Sftpc
Log2 (TPX+1)
14 12 10 8 6 4 2 0
16 14 12 10 8 6 4 2 0
SPC IHC
Grade 1/2
Grade 2
Grade 3
Grade 3/4
B
z-score
2.5 0.0 −2.5
1.05 0 −1.6 0.29 1.18 −1.33 0 0.28 0 −0.07 0 0.19 Lgr5
−0.53 0.78 0.74 −1.33 0.1 −0.5 −0.7 −0.49 −0.07 −0.48 2.56 −0.08 Lgr6
0.15 0.04 −1.35 −0.57 −0.15 0 −1.07 −0.4 1.65 0.04 1.66 0 Axin2
−0.67 −0.86 1.59 −0.23 −1.03 0.29 −0.42 −0.73 −1.27 1.33 0.94 1.05 Porcn
−0.5 0.35 0.27 −0.18 −0.02 −0.59 −1.17 −0.2 −0.14 −0.49 2.91 −0.26 Notch3
1 2 3 4 5 6 7 8 9 10 11 12
Cluster
C
High cycling
Gene program 2
Hepatic-gastric/ Inflammation
Gene program 3
Biosynthetic mixed identity
Gene program 4
Stress response
Gene program 5
GI-epithelium like
Gene program 8
Gastric-like
Gene program 10
D
1 2 3 4 5 6 7 8 9 10 11
Program
E
1 1 176 9 24 6 0 10 2 11 0 1 1 AT1/AT2
16 0 1 2 6 3 27 13 5 0 0 0 2 High cycling
0 0 0 1 0 2 0 31 0 0 0 161 3 Hepatic-gastric/Inflammation
2 0 0 13 6 5 0 5 273 18 1 10 4 Biosynthetic mixed identity
11 0 0 2 1 1 0 0 0 0 0 0 5 Stress response
6 0 5 0 513 1 0 4 12 3 0 11 6 Highly mixed
0 0 0 0 76 0 0 1 0 0 88 3 7 EMT
1 0 2 41 59 135 0 10 2 1 0 2 8 GI epithelium-like
0 0 4 131 0 0 9 472 74 0 0 0 9 Embryonic-liver like
0 0 0 0 80 0 0 69 0 184 0 1 10 Gastric-like
615 284 5 116 0 5 0 2 7 0 0 0 11 AT2-like
Pearson residual
30 0 −30
1 2 3 4 5 6 7 8 9 10 11 12
Cluster
Cellular program
F
Mouse LUAD
Cell fraction
1.0 0.5 0.0
T 2 wk
KT 2 wk
KPT 2 wk
KT 12 wk
KT 30 wk
KPT 12 wk
KPT 20 wk
KPT 30 wk
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cluster 12
G
Ly6i
Tmprss7
Lsmem1
Tgtp1
Ifi27l2a
Vldlr
Prmt8
Hck
Mrc2
Akap5
Map6
Hopx
Trpv6
Tbc1d24
Cd177
C5ar2
Ugt1a6b
Tnf
Il23a
Rrad
Abtb2
Tigit
Slc4a11
Lrrc26
Cysrt1
Syt16
Zfp991
Alpi
Aurkb
Pbk
Knl1
Neil3
Hist2h3b
Arg1
Cldn1
Cyp1b1
Prlg
Adamts6
Fam221b
Gadl1
Tpt1-ps5
Tpt1-ps6
Mir703
Tfap2b
Hcrtr1
Gkn1
Upk1b
Pnp2
Bmp7
Rgs17
Klhl1
Ifi211
Stmn2
Cd109
Tmprss6
Ace
Osr2
Lrmp
Azgp1
Lrrc52
Frequency of expression
1 2 3 4 5 6 7 8 9 10
Mean expression
5.0 2.5 0.0
1 2 3 4 5 6 7 8 9 10 11 12
Cluster
H
Slc4a11 (log2[TPX+1])
0 1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10 11 12
Cluster
I
Slc4a11
Log2 (TPX+1)
10 8 6 4 2 0
J
Adenoma
Adenocarcinoma
Grade 2
Grade 3
Grade 4
Slc4a11

Figure S2, related to Figure 2 and Table S2. Alternative cell type and cell transition programs arise during tumor progression. (**A**) Loss of AT2 cell marker expression in a subset of LUAD cells during progression. Top: PHATE map (as in **Figure 1D**) of scRNA-seq profiles with cells (dots) colored by expression ($\log_2$(TPX+1)) of the AT2 cell markers *Lyz2* and *Sftpc*. Bottom: Immunohistochemistry for surfactant protein-C (SPC), encoded by *Sftpc*, in mouse lung neoplasias. Red dashed line indicates boundary between grade 3 and grade 4 regions in right-most image. Scale bar: 100 µm. (**B**) Regulators of alternative cell fate are expressed in late-emerging clusters. Mean expression (z-score on log(TPX+1) of previously reported LUAD tumor cancer cell subpopulation markers (rows) across the clusters (columns). (**C**) Additional NMF programs arising from an analysis of the scRNA-Seq data (**STAR Methods**). (**D**) Activation weighted mean expression of top program genes. Mean expression (z-score value of expression weighted by the relative contribution to activation value for that program; color bar) of genes (rows) associated with one of the 11 programs (columns). (**E**) Comparison of cell cluster membership and top scoring NMF program. Shown is the count of each combination of cluster (columns) and program (rows, **Table S2**), colored by the deviation from expected (Pearson's residual). (**F**) Cell membership in clusters over time. Fraction of cells (*y* axis) from each cluster (color code, as in **Figure 1D**), in each individual sample (*x* axis). (**G**) Selected gene markers for clusters 5 and 11. Top five uniquely expressed genes (rows, **STAR Methods**) in each cluster (columns) (color bar, as in **Figure 1D**). Bold and highlighted: Markers chosen for cluster 5 (*Slc4a11*, *Tigit*) and 11 (*Cd109*). (**H**) Expression of *Slc4a11* across clusters ($\log_2$ (TPX+1)). **(I)** Expression of *Slc4a11* across cells in a PHATE map ($\log_2$ (TPX+1)). (**J**) RNA *in situ* hybridization for *Slc4a11* (purple) in mouse adenoma and adenocarcinoma tissues. Scale bar: 100 µm

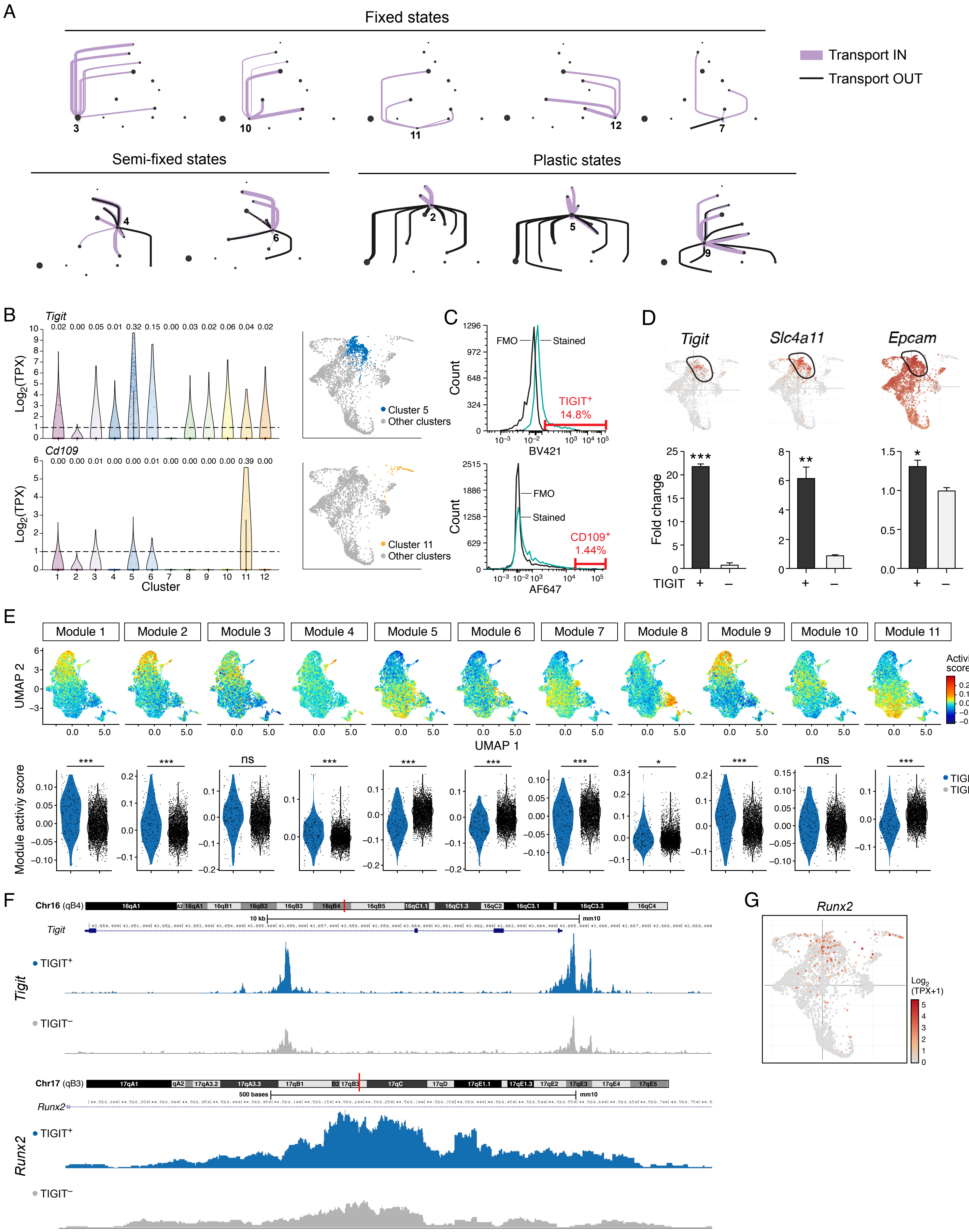

A
Fixed states
3
10
11
12
7
Transport IN
Transport OUT
Semi-fixed states
4
6
Plastic states
2
5
9
B
Tigit
0.02 0.00 0.05 0.01 0.32 0.15 0.00 0.03 0.02 0.06 0.04 0.02
Log2(TPX)
Cluster 5
Other clusters
Cd109
0.00 0.00 0.01 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.39 0.00
Log2(TPX)
1 2 3 4 5 6 7 8 9 10 11 12
Cluster
Cluster 11
Other clusters
C
Count
FMO
Stained
TIGIT+
14.8%
BV421
Count
FMO
Stained
CD109+
1.44%
AF647
D
Tigit
Slc4a11
Epcam
Fold change
***
**
*
TIGIT + –
E
Module 1
Module 2
Module 3
Module 4
Module 5
Module 6
Module 7
Module 8
Module 9
Module 10
Module 11
UMAP 2
UMAP 1
Activity score
0.2
0.1
0.0
−0.1
−0.2
Module activiy score
*** *** ns *** *** *** *** * *** ns ***
TIGIT+
TIGIT−
F
Chr16 (qB4)
10 kb
mm10
Tigit
TIGIT+
TIGIT−
Chr17 (qB3)
500 bases
mm10
Runx2
TIGIT+
TIGIT−
G
Runx2
Log2 (TPX+1)
5
4
3
2
1
0

Figure S3, related to Figure 3. TIGIT$^+$ mouse LUAD HPCS cells are plastic. (**A**) Optimal transport model predicts fixed, semi-fixed, and plastic cell states. Optimal Transport graphs (as in **Figure 3A**), but showing all transitions (aggregate across all time points) to (black) and from (purple) the selected cluster's cells. (**B**) *Tigit* and *Cd109* expression mark cluster 5 and 11 cells, respectively. Left: distribution of expression levels ($\log_2$(TPX+1), *y* axis) of *Tigit* (top) and *Cd109* (bottom) across the cells in each cluster (*x* axis). Right: PHATE map of scRNA-seq profiles (as in **Figure 1D**) with cells (dots) colored by *Tigit* (top) and *Cd109* (bottom) expression ($\log_2$(TPX+1), color bar). (**C**) Isolation of TIGIT$^+$ and CD109$^+$ populations by FACS. FACS plot of the distribution of TIGIT (top) and CD109 (bottom) expression in either fluorescence minus-one (FMO, colored curves) or stained cells (black curves). Red: percentage of positive cells. (**D**) Prospective isolation of cluster 5 cells by TIGIT expression. Top: PHATE map embedding (as in **Figure 1D**) with cells (dots) colored by the colored by expression of each gene. Bottom: Mean RNA levels (*y* axis, quantitative PCR) of cluster 5 markers *Tigit*, *Slc4a11,* and the epithelial marker *Epcam* in TIGIT$^+$ *vs*. TIGIT$^-$ LUAD cells (*x* axis) isolated at 17-22 weeks following tumor initiation. Error bars: standard deviation, $n = 4$. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (unpaired *t* test). (**E**) scATAC-Seq data showing gene activity for modules from the accompanying LaFave et al. manuscript. (top) UMAP representation of TIGIT$^+$ and TIGIT$^-$ sorted cells subjected to scATAC-Seq with gene activity as depicted by the scale bar. (bottom) Violin plot representation for the same data with TIGIT$^+$ and TIGIT$^-$ separated. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ Wilcoxon rank sum test. (**F**) Representative bulk ATAC-Seq peak data showing enrichment of accessible peaks in the TIGIT$^+$ over the TIGIT$^-$ sample for representative genes *Tigit* and *Runx2.* (**G**) Expression level of *Runx2* projected onto the PHATE map as in **Figure 1D**.

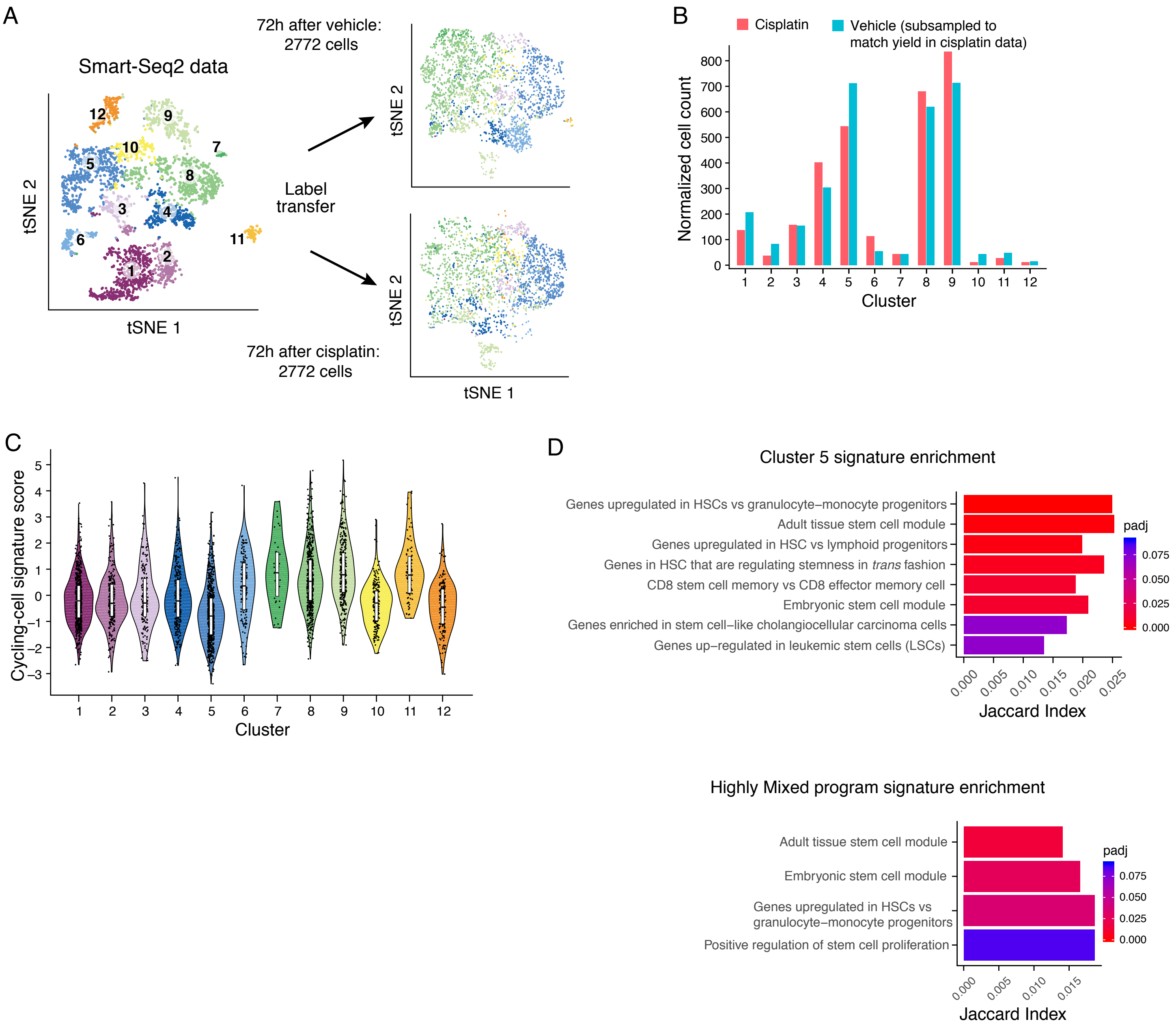

A
Smart-Seq2 data
tSNE 2
tSNE 1
12
9
10
7
5
8
3
4
6
11
1
2
72h after vehicle:
2772 cells
Label
transfer
72h after cisplatin:
2772 cells
tSNE 2
tSNE 2
tSNE 1
B
Cisplatin
Vehicle (subsampled to
match yield in cisplatin data)
Normalized cell count
0
100
200
300
400
500
600
700
800
1
2
3
4
5
6
7
8
9
10
11
12
Cluster
C
Cycling-cell signature score
5
4
3
2
1
0
−1
−2
−3
1
2
3
4
5
6
7
8
9
10
11
12
Cluster
D
Cluster 5 signature enrichment
Genes upregulated in HSCs vs granulocyte−monocyte progenitors
Adult tissue stem cell module
Genes upregulated in HSC vs lymphoid progenitors
Genes in HSC that are regulating stemness in *trans* fashion
CD8 stem cell memory vs CD8 effector memory cell
Embryonic stem cell module
Genes enriched in stem cell−like cholangiocellular carcinoma cells
Genes up−regulated in leukemic stem cells (LSCs)
padj
0.075
0.050
0.025
0.000
0.000
0.005
0.010
0.015
0.020
0.025
Jaccard Index
Highly Mixed program signature enrichment
Adult tissue stem cell module
Embryonic stem cell module
Genes upregulated in HSCs vs
granulocyte−monocyte progenitors
Positive regulation of stem cell proliferation
padj
0.075
0.050
0.025
0.000
0.000
0.005
0.010
0.015
Jaccard Index

Figure S4, related to Figure 5 and Table S4. Cells in the HPCS associate with chemoresistance, quiescence and metastasis. (**A**) Annotation of cell profiles collected after cisplatin treatment with cell cluster labels from the LUAD progression atlas. tSNE of scRNA-Seq profiles from the LUAD progression data (right), with cells (dots) colored by clusters (as in **Figure 1D**), showing the cells collected after vehicle (top) or cisplatin (bottom) treatment, with cells (dots) colored by classifier predicted of cluster labels from the LUAD progression data (**STAR Methods**). Vehicle cells were subsampled to 2,772 to match cisplatin treatment cell yield. Both vehicle and cisplatin treated cells were collected from microdissected *KPT* tumors at 20 weeks post-tumor initiation 72 hours following a single dose of vehicle and cisplatin (n = 2 mice per condition). (**B**) Normalized cell count of cells in each cluster after chemotherapy treatment. (**C**) Distribution of gene-set signature scores of cell-cycle related genes, z-score (*y* axis) per cluster (*x* axis). The "Cycling cell" signature from the Mouse Cell Atlas study (Han et al., 2018) was used. (**D**) MSigDB (Liberzon et al., 2011; Subramanian et al., 2005) stemness signatures significantly enriched (FDR adjusted p value <0.1) in cluster 5 (top) and the Highly mixed program (NMF 6, bottom). Signatures are ranked by p value (color bar) and the length of the bar corresponds to the Jaccard Index.

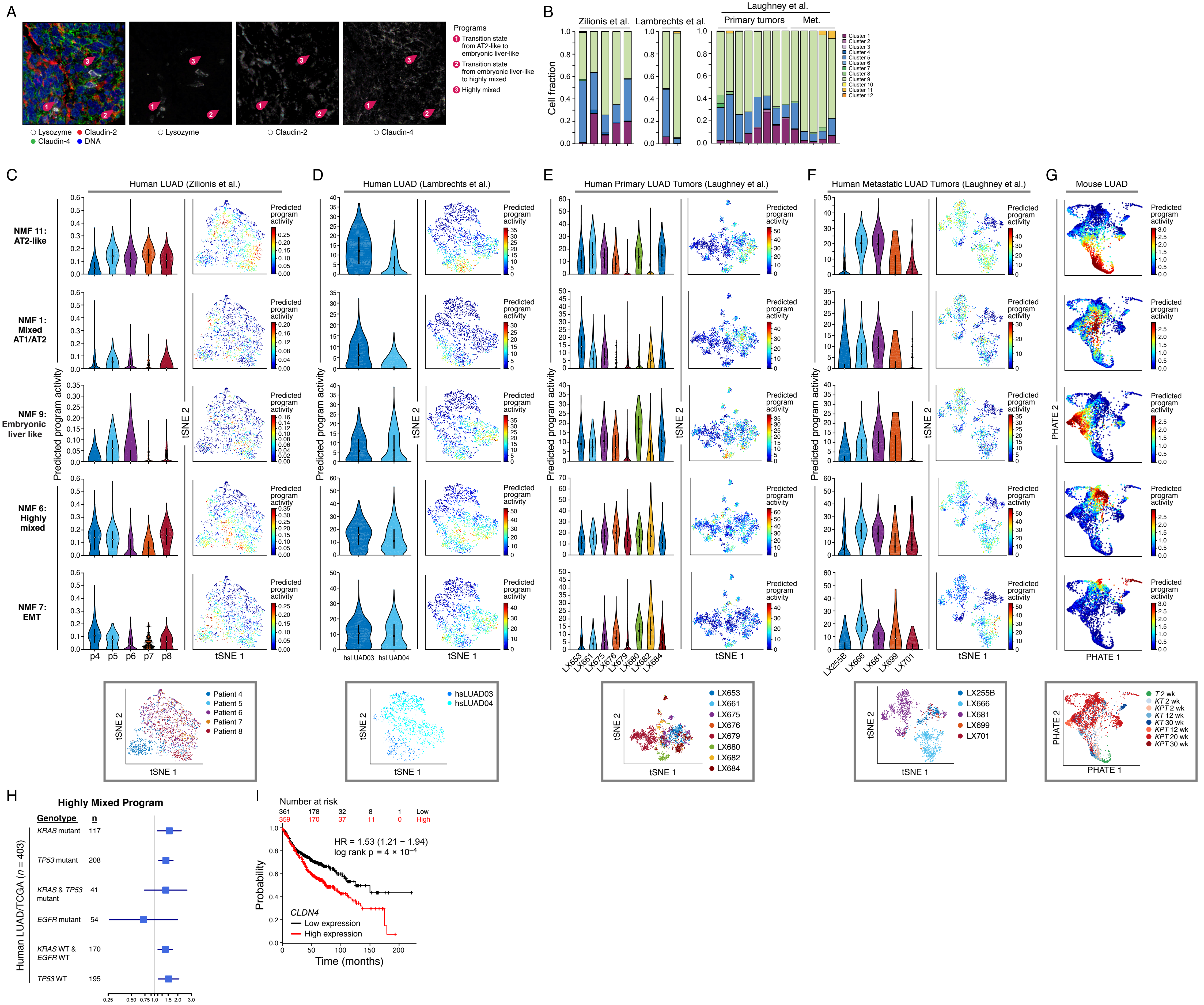
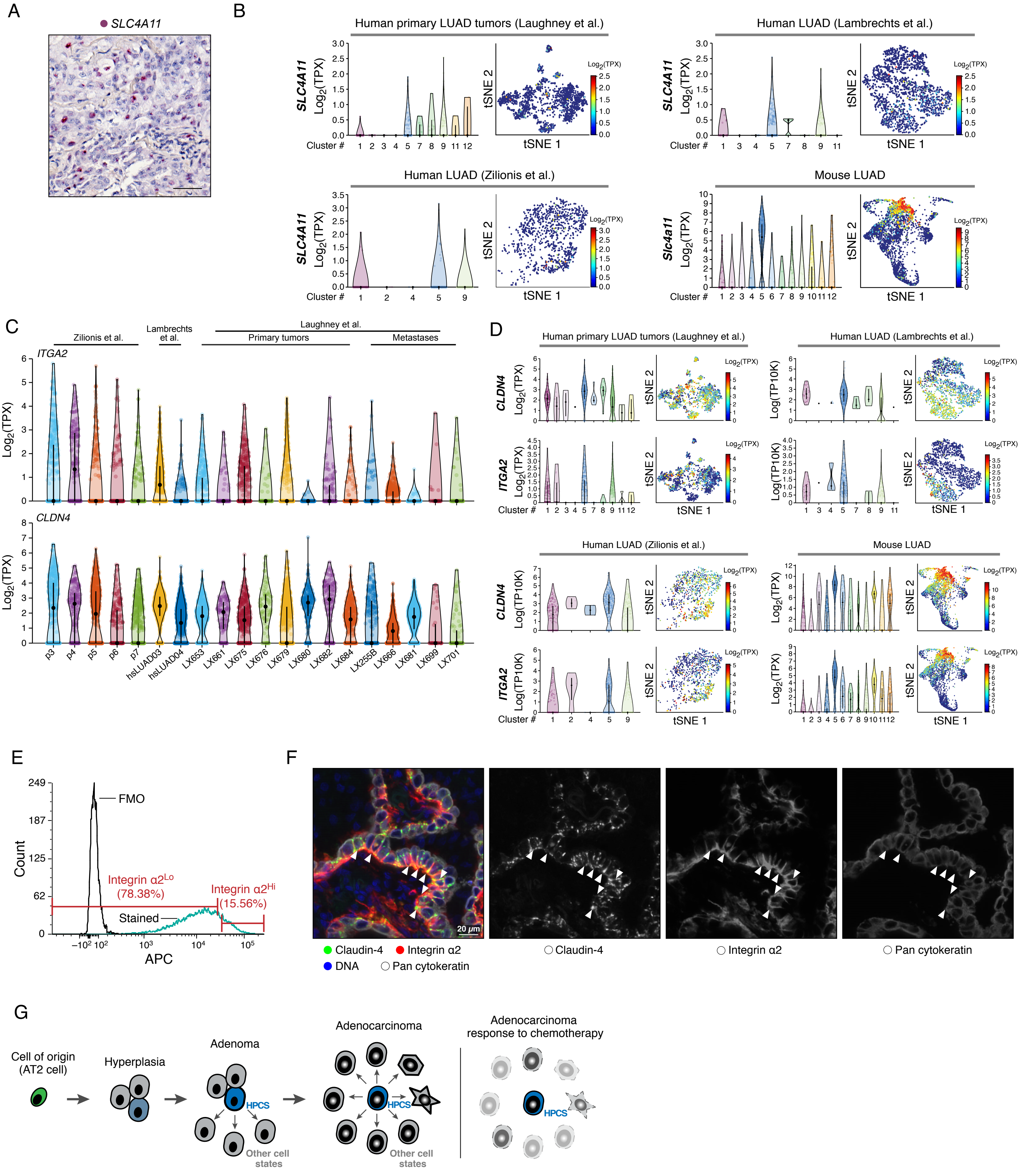

A
Lysozyme
Claudin-2
Claudin-4
DNA
Programs
1 Transition state from AT2-like to embryonic liver-like
2 Transition state from embryonic liver-like to highly mixed
3 Highly mixed
B
Zilionis et al.
Lambrechts et al.
Laughney et al.
Primary tumors
Met.
Cell fraction
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cluster 12
C
Human LUAD (Zilionis et al.)
D
Human LUAD (Lambrechts et al.)
E
Human Primary LUAD Tumors (Laughney et al.)
F
Human Metastatic LUAD Tumors (Laughney et al.)
G
Mouse LUAD
NMF 11: AT2-like
NMF 1: Mixed AT1/AT2
NMF 9: Embryonic liver like
NMF 6: Highly mixed
NMF 7: EMT
Predicted program activity
tSNE 1
tSNE 2
PHATE 1
PHATE 2
p4
p5
p6
p7
p8
hsLUAD03
hsLUAD04
LX653
LX661
LX675
LX676
LX679
LX680
LX682
LX684
LX255B
LX666
LX681
LX699
LX701
Patient 4
Patient 5
Patient 6
Patient 7
Patient 8
T 2 wk
KT 2 wk
KPT 2 wk
KT 12 wk
KT 30 wk
KPT 12 wk
KPT 20 wk
KPT 30 wk
H
Highly Mixed Program
Genotype
n
KRAS mutant 117
TP53 mutant 208
KRAS & TP53 mutant 41
EGFR mutant 54
KRAS WT & EGFR WT 170
TP53 WT 195
Human LUAD/TCGA (n = 403)
0.25
0.50
0.75
1.0
1.5
2.0
3.0
I
Number at risk
361 178 32 8 1 Low
359 170 37 11 0 High
HR = 1.53 (1.21 − 1.94)
log rank p = 4 × 10⁻⁴
Probability
CLDN4
Low expression
High expression
Time (months)
0
50
100
150
200

Figure S5, related to Figure 6. Human LUAD cancer cells express key transcriptional programs derived from the mouse model. (**A**) HPCS markers are expressed in human LUAD *in situ*. Immunofluorescence staining for Lysozyme (AT2-like program), Claudin-2 (Embryonic liver-like), and Claudin-4 (Highly mixed program), along with DNA visualized by DAPI staining. Pink numbered arrowheads indicate cell states or transitions predicted by the mouse model. Scale bar: 20 µm. (**Figure 2D-F**): 1 - AT2-like (lysozyme) to Embryonic liver-like (claudin-2) transition; 2 - Embryonic liver-like (claudin-2) to Highly mixed (claudin-4) transition; 3 - Highly mixed program (claudin-4). LUAD tissue from a 72-year old male operated at the University of Vanderbilt-Ingram Cancer Center (Stage I (pT1a)) (Amin et al., 2017)). (**B**) Some mouse LUAD cell cluster signatures are detected in human LUAD. Fraction of cells (*y* axis) from each tumor (*x* axis) that express the signature of each cluster (color legend, **STAR Methods**). For simplicity, unclassified cells are not displayed. (**C-G**) Activation program scores of mouse LUAD programs in individual malignant cells from human LUAD tumors. For each of the five main programs detected in mouse LUAD, shown are violin plots (left) of the distribution of program scores (*y* axis) of in the cancer cells in each tumor (*x* axis), and a tSNE of the cell profiles (right), with cells (dots) colored by their program scores, and by the patient identity (bottom framed panel) in each of three scRNA-Seq studies of cancer cells from human LUAD tumors [(**C**)**,** (**D**)**,** (**E**)**,** (**F**), (Lambrechts et al., 2018; Laughney et al., 2020; Zilionis et al., 2019)], as well as a PHATE map of mouse LUAD cells (**G**), colored by the program scores (as in **Figure 2D**), and by timepoint/genetic combination (bottom framed panel). (**H**) Analysis of TCGA data shows the Highly Mixed program is associated with worse prognosis for patients irrespective of *KRAS* and/or *TP53* mutations. WT: wild-type. (**I**) High expression of the HPCS marker *CLDN4* is associated with poor survival. Kaplan-Meier curve of patients stratified by high (red) *vs*. low (black) expression of *CLDN4* (Gyorffy et al., 2013).



A
SLC4A11
B
Human primary LUAD tumors (Laughney et al.)
Human LUAD (Lambrechts et al.)
Human LUAD (Zilionis et al.)
Mouse LUAD
SLC4A11
Slc4a11
Log2(TPX)
Cluster #
tSNE 1
tSNE 2
C
Zilionis et al.
Lambrechts et al.
Laughney et al.
Primary tumors
Metastases
ITGA2
CLDN4
Log2(TPX)
p3
p4
p5
p6
p7
hsLUAD03
hsLUAD04
LX653
LX661
LX675
LX676
LX679
LX680
LX682
LX684
LX255B
LX666
LX681
LX699
LX701
D
Human primary LUAD tumors (Laughney et al.)
Human LUAD (Lambrechts et al.)
Human LUAD (Zilionis et al.)
Mouse LUAD
CLDN4
ITGA2
Log(TP10K)
Cluster #
E
FMO
Stained
Count
APC
Integrin α2Lo (78.38%)
Integrin α2Hi (15.56%)
F
20 μm
Claudin-4
Integrin α2
DNA
Pan cytokeratin
G
Cell of origin (AT2 cell)
Hyperplasia
Adenoma
Adenocarcinoma
Adenocarcinoma response to chemotherapy
HPCS
Other cell states

Figure S6, related to Figure 7, Table S1, and Table S3. Markers for the HPCS in mouse LUAD identify a similar state in human LUAD. (**A**) RNA *in situ* hybridization of *SLC4A11* in a human LUAD tumor from an 81-year old female patient operated at Memorial Sloan Kettering Cancer Center (Stage IIB (pT2bN1) (Amin et al., 2017), genotype *KRASp.G12A; TP53p.P278S*). Scale bar: 50 µm. (**B**) Violin plots showing the expression levels of *SLC4A11* or *Slc4a11* in various clusters from primary human (left) and mouse (right) LUAD tumors. (**C**) Violin plots of expression levels for *ITGA2* and *CLDN4* in 15 primary human LUAD tumors and 8 metastases. (**D**) Expression of HPCS signature genes in human tumors. Distribution of expression levels (*y* axis, $\log_2$(TPX), violin plots), and in tSNE plots of two signature genes of mouse HPSC in tumor cells (*x* axis, **STAR Methods**) from three human data sets or from mouse. (**E**) FACS plot of the distribution of Integrin $\alpha2^{Hi}$ and Integrin $\alpha2^{Lo}$ expression in either fluorescence minus-one (FMO, black curve) or stained cells (colored curve). Red: percentage of positive cells. (**F**) Immunofluorescence for Claudin-4 and Integrin α2 (both markers of the Highly mixed cell state) along with pan-cytokeratin. DNA visualized by DAPI staining. White arrow: a Claudin-4 and Integrin α2 double positive cancer cell. **(G)** Model. LUAD initiates in AT2 cells of the lung, giving rise to adenomatous atypical hyperplasia (AAH), which retain alveolar identity. Upon transition of AAH to adenoma, a high-plasticity cell state (HPCS, blue) emerges, which gives rise to multiple surrogate cancer cell identities (gray cells). The HPCS gives rise to identities with features of the lung lineage in adenomas, whereas in adenocarcinomas the HPCS produces phenotypes that resemble other endoderm-derived cells (such as embryonic liver and gastric epithelium) and finally gives rise to epithelial-to-mesenchymal transition (EMT).