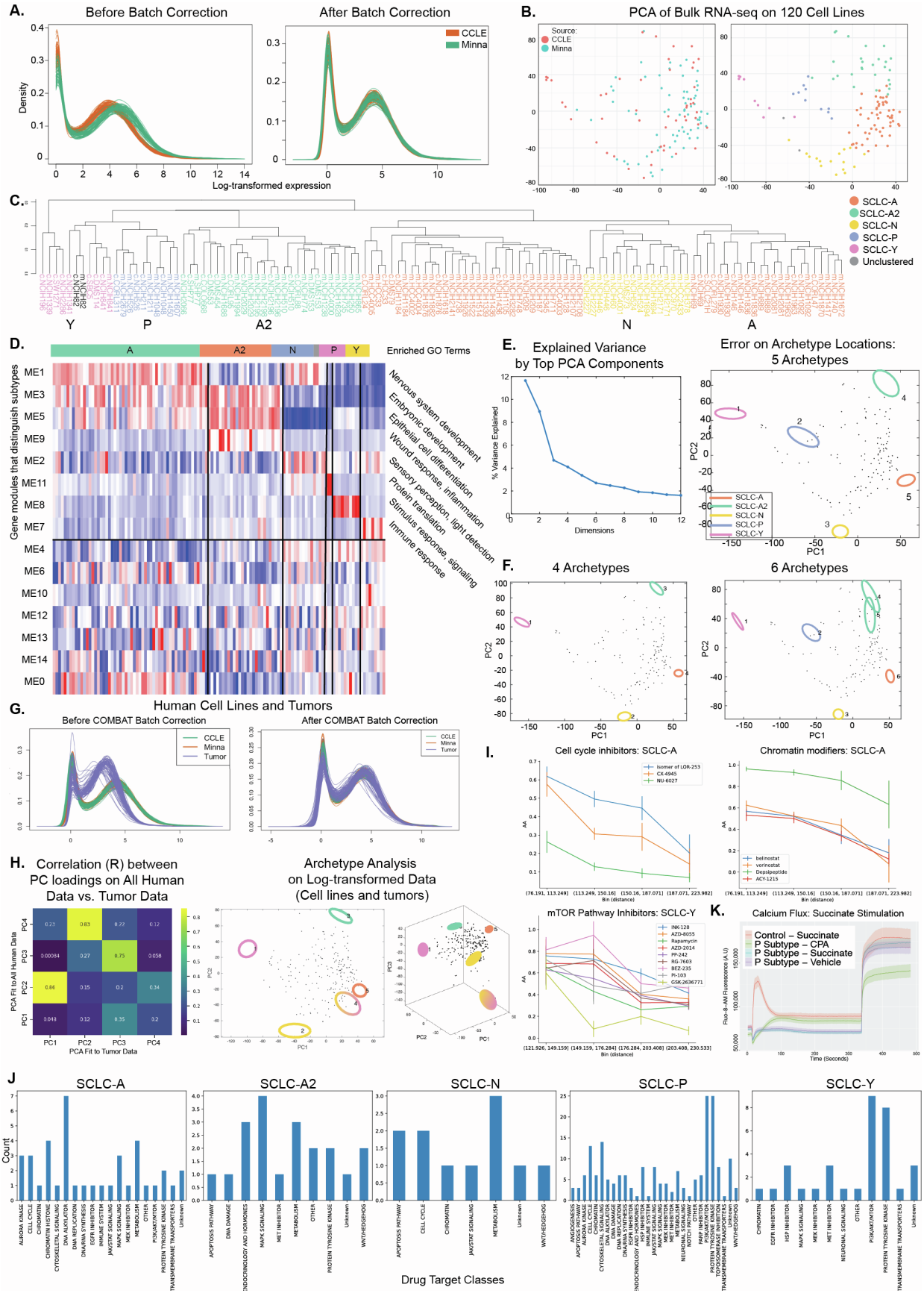
Cell Systems, Volume *13*

Supplemental information

Archetype tasks link intratumoral heterogeneity to plasticity and cancer hallmarks in small cell lung cancer

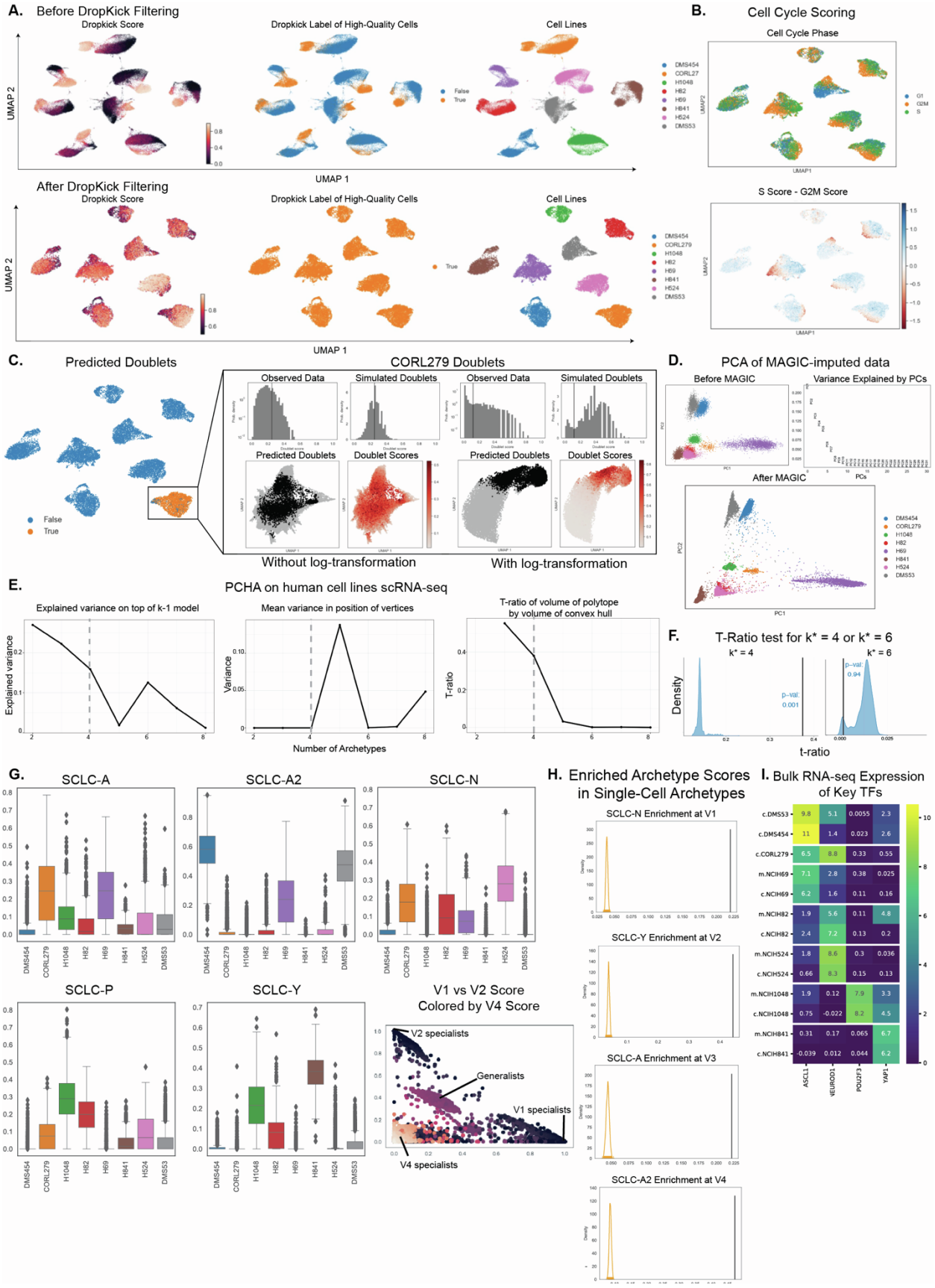
Sarah M. Groves, Geena V. Ildefonso, Caitlin O. McAtee, Patricia M.M. Ozawa, Abbie S. Ireland, Philip E. Stauffer, Perry T. Wasdin, Xiaomeng Huang, Yi Qiao, Jing Shan Lim, Jackie Bader, Qi Liu, Alan J. Simmons, Ken S. Lau, Wade T. Iams, Doug P. Hardin, Edward B. Saff, William R. Holmes, Darren R. Tyson, Christine M. Lovly, Jeffrey C. Rathmell, Gabor Marth, Julien Sage, Trudy G. Oliver, Alissa M. Weaver, and Vito Quaranta

Figure S1: Related to Figures 1 and 2



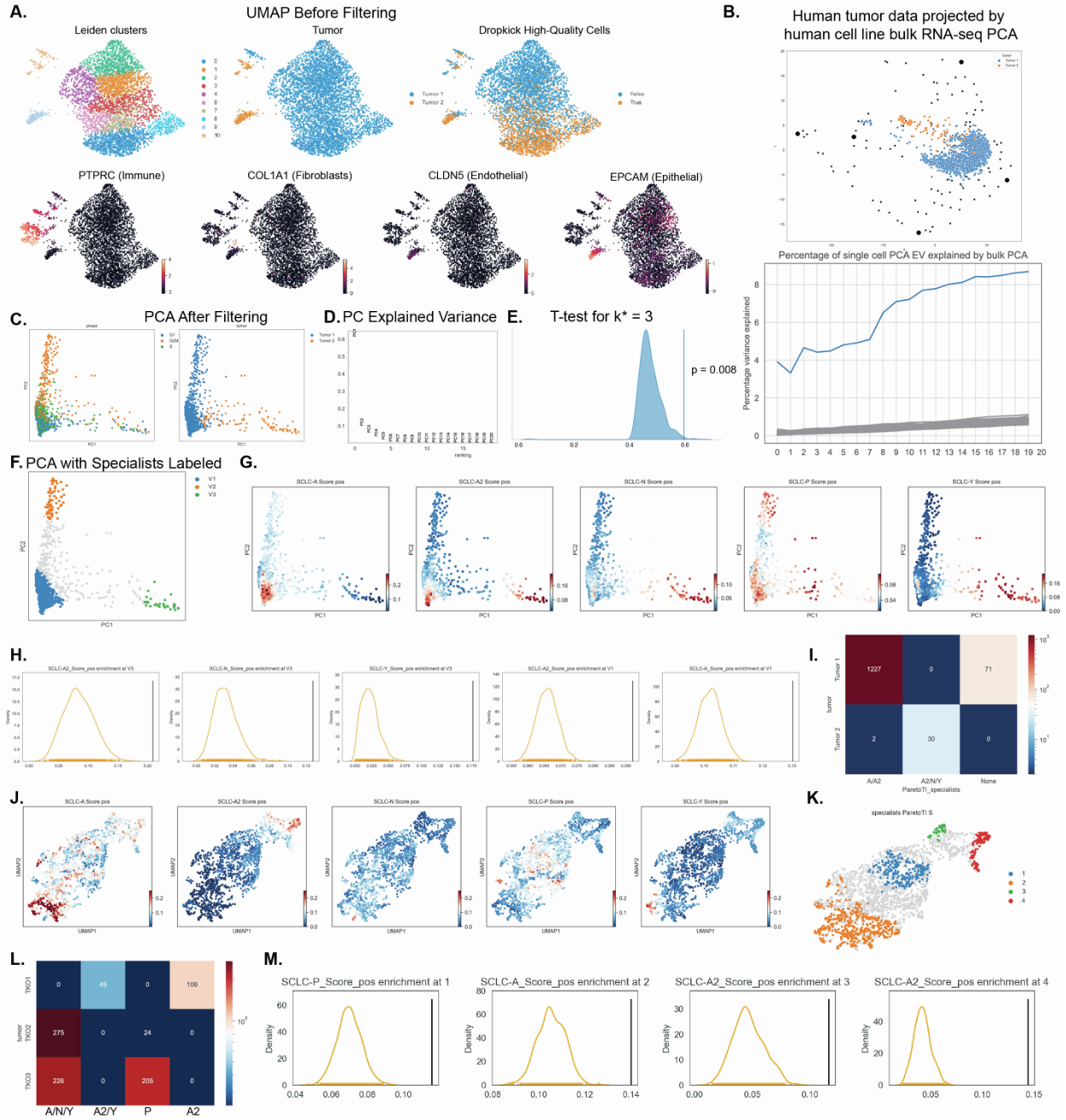
Supplemental Figure 1: Supplement to bulk RNA-seq preprocessing and PCA on human cell lines, related to Figure 1. A. Gene expression distribution before and after batch correction for CCLE and Minna datasets. **B.** Cell line source in PCA, and clustering shown by color on PCA. **C.** Hierarchical clustering on cell lines as shown in **B** with labels. "C" or "m" designates the source of each cell line. H82 is labeled as unclustered because it is known to express NEUROD1 but is clustered with other SCLC-Y cell lines. **D.** WGCNA on cell lines shows genes can be grouped into 15 coexpressed gene modules. Several of the modules (above black line) distinguish subtype clusters and are labeled with enriched gene ontology terms describing each gene program. **E.** Explained variance in human cell lines by top PCA components. Archetypal analysis on top 12 components gives 5 archetypes with low error upon bootstrapping. **F.** Four or six archetypes correspond to the five archetypes in **E**. **G.** Batch correction of human cell lines and human tumors (81 samples) using COMBAT. Distribution of log-transformed expression is shown before and after correction. **H.** PC loadings on all human data are highly correlated with PC loadings on tumor data only, with PC1 from tumor-data PCA matching PC3 from all-data PCA, PC2 (tumor) matching PC4 (all), and PC3 (tumor) matching (PC3). AA on all human data shows archetypes that match human cell line archetypes. **I-J.** Drug sensitivity to various drug classes. **I** shows select classes for SCLC-A and SCLC-Y, with the cell lines closest to each archetype showing the highest sensitivity (activity area). Response is binned by distance to archetype. **J** shows the significant number of drugs from each class according to an ANOVA (one-vs all) for bin closest to archetype. **K.** H1048 (SCLC-P) cells and a positive control cell line (known to respond to succinate with calcium signaling through SUCNR1) were treated with a concentration response curve of succinate at 10 seconds (maximum treatment, 2.4 mM, shown here) and calcium flux was evaluated. While the control (pink) shows response, the SCLC-P cell line does not show any response to succinate (blue), similar to vehicle (purple). Cyclopiazonic acid (CPA, green) was used to release ER calcium stores to ensure detection of calcium flux was possible. At 340 seconds, ionomycin is added to show maximal calcium signal. Error bars show standard error from 3 replicates for each cell line.

Figure S2: Related to Figure 3



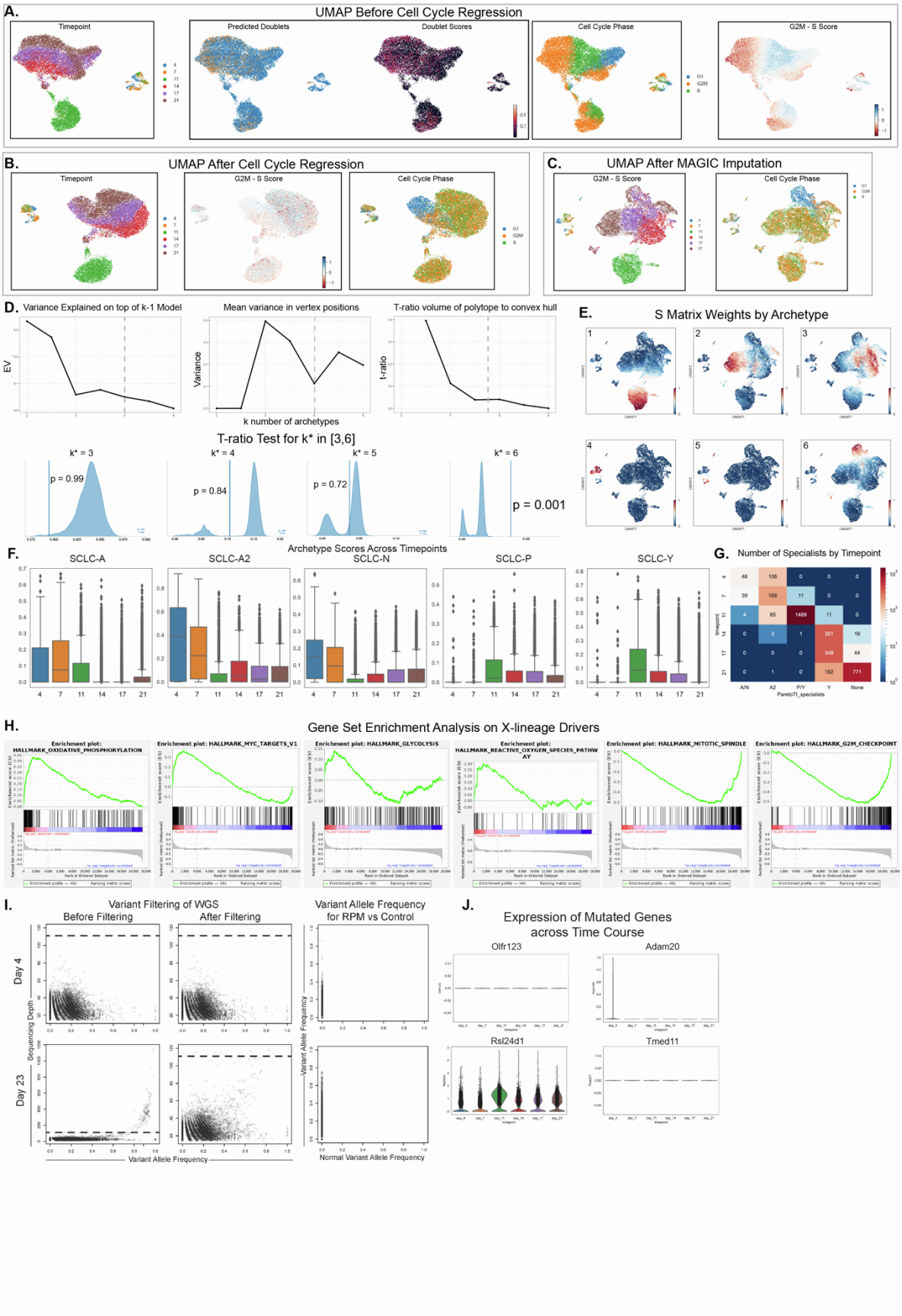
Supplemental Figure 2: Archetype signature generation and application to scRNA-seq on human cell lines, related to Figure 3. A. Before and after filtering by Dropkick to remove low quality cells. **B.** Cell cycle scoring by Scanpy shows slight dependence of UMAP location on cell cycle. **C.** Predicted doublets using Scrublet. Very few cells are predicted to be doublets except in cell line CORL279. Here, log-transformation shows a distinct region of cells with high likelihood to be doublets, which were removed from future analysis. **D.** PCA of data before and after MAGIC imputation. A high proportion of the variance in the cell lines was explained by the top 10 PCs after MAGIC imputation. **E.** Principal Convex Hull Analysis (PCHA) on single cell data. Explained variance per number of archetypes shows a large dip after 4 archetypes, with an increase for 6 archetypes. Mean variance in position of vertices greatly increases for 5 archetypes, and the T-ratio greatly decreases for 5 archetypes. **F.** T-ratio plots for k* = 4 or 6. Only k*=4 gives a significant t-ratio. **G.** Archetype signatures by cell line. Using the signature scoring method (see methods), we see that some cell lines are specialists for a particular archetype (enriched in bulk archetype score). A proportion of the cells across cell lines are generalists, not enriched in any archetype score. **H.** Enrichment of archetype scores at each single-cell vertex by permutation test (see methods). P = 0 for all tests shown, defined as the proportion of the background distribution above the score for archetype specialists at a particular vertex. **I.** Expression of four key SCLC transcription factors in bulk RNA-seq data. Subtype proportions for each cell line are consistent with expression of key TFs at bulk level. For example, H1048 expresses both POU2F3 and YAP1 and is predicted to be a mix of SCLC-P and SCLC-Y specialists and generalists. **e.** Generalist subtype proportions by cell line shown as bar plots.

Figure S3: Related to Figure 4

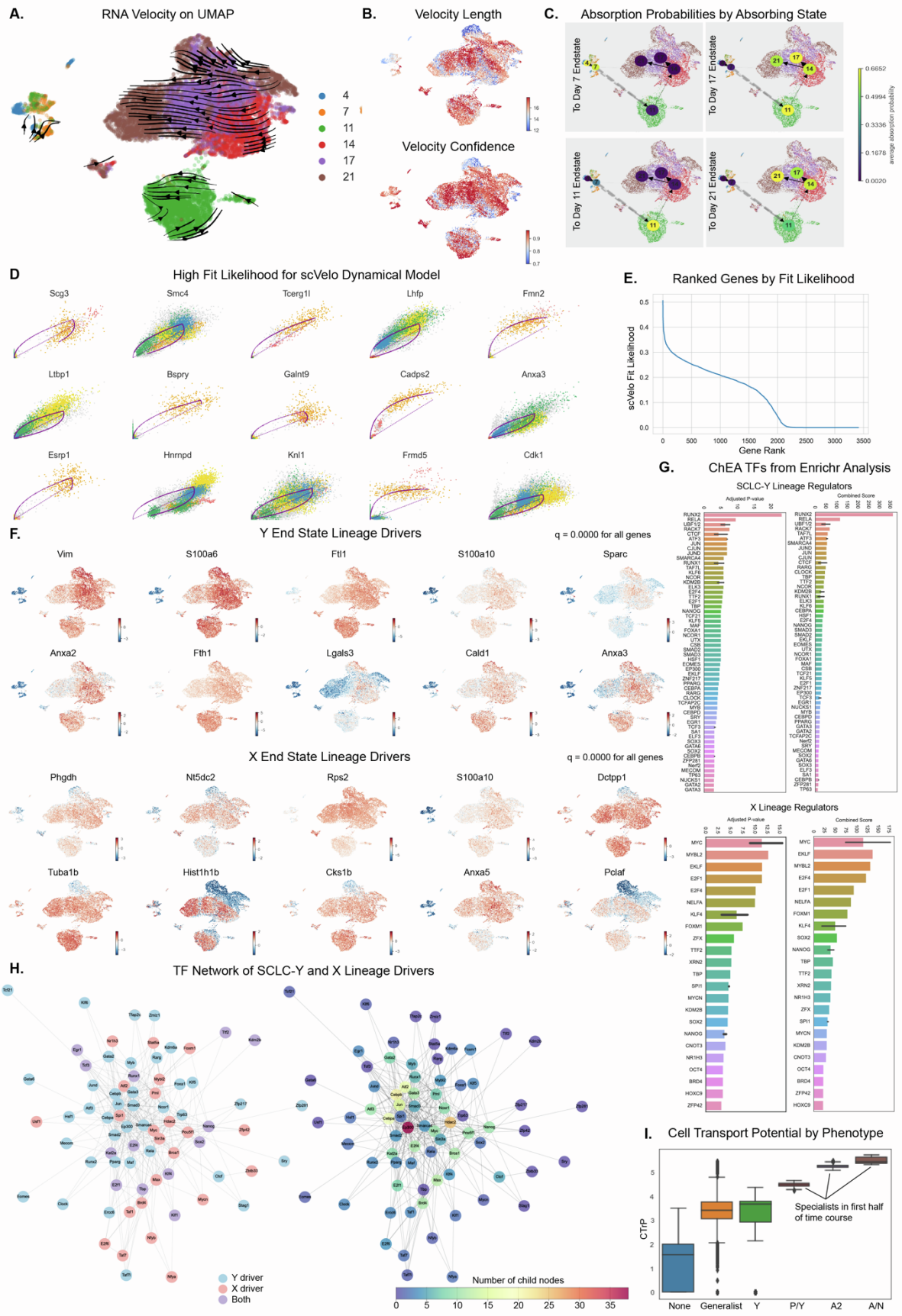


Supplemental Figure 3: Single cell preprocessing and archetype signature scoring on human and mouse tumors, related to Figure 4. A. UMAP before and after filtering cells. We removed low quality cells as determined by Dropkick. We also removed subpopulations that expressed immune, fibroblast, or endothelial markers. **B.** Human tumor single cell data projected into the bulk archetype space. About 6% of the variance in the single cell tumor data is explained by bulk data on cell lines. **C.** PCA of data after filtering in **A.** and before MAGIC imputation. **D.** Explained variance by top principal components. **E.** T-ratio test for k* = 3 on human tumor data. **F.** PCA with three archetype specialists labeled before MAGIC imputation. **G.** Scores for bulk archetype signatures on PCA of human tumor data. **H.** Enrichment of archetype signature scores at each single-cell archetype vertex. Only significant enrichments shown. **I.** Number of human tumor cells that are considered specialists for each of the three archetypes. **J.** UMAP of TKO data colored by archetype signature scores. **K.** Four groups of archetype specialists shown on TKO UMAP. **L.** Number of TKO cells from each tumor that are archetype specialists. Archetypes are labeled by all enriched signatures. **M.** Enrichment of archetype signature scores used to label **L.** Only most enriched for each single cell vertex is shown.

Figure S4: Related to Figure 5



Supplemental Figure 4: RPM time series scRNA-seq preprocessing and Archetypal Analysis, related to Figure 5. A. UMAP of RPM time series before cell cycle regression, showing timepoints, predicted doublets (removed for downstream analysis), and cell cycle phase and score. **B.** UMAP of data after regressing out cell cycle dependence (G2M-S score). **C.** UMAP after MAGIC imputation. **D.** Variance explained by different numbers of archetypes. Mean variance in vertices shows a decrease for 6 archetypes. T-ratio for different numbers of vertices shows the t-ratio levels off after 5-6 archetypes. Bottom plots show t-ratio tests for 3-6 archetypes. Only k*=6 is significant. **E.** Archetypal analysis S matrix scores for 6 archetypes. **F.** Bulk archetype signature scores by timepoint. NE (A, A2, and N) scores are higher in earlier timepoints. **G**. Number of archetype specialists across timepoints. **H.** Gene set enrichment analysis shows X specialists are enriched in oxidative phosphorylation, MYC targets, glycolysis, and ROS pathway, and depleted in mitotic spindle and G2M checkpoint gene sets **I. (Left)** Variant allele frequency (VAF) before and after filtering for two timepoints of independent RPM time series. We only consider the variants that have more than 15 reads and less than 110 reads coverage (depth, DP). **(Right)** VAF in normal control sample versus RPM timepoints. Somatic variants in RPM samples have no alternative alleles in the normal sample. **J.** Expression of four genes coded by genomic regions with somatic mutations. Three of the four genes have low to no expression across the time course.


A.
RNA Velocity on UMAP
4
7
11
14
17
21
B.
Velocity Length
Velocity Confidence
C.
Absorption Probabilities by Absorbing State
To Day 7 Endstate
To Day 17 Endstate
To Day 11 Endstate
To Day 21 Endstate
average absorption probability
D
High Fit Likelihood for scVelo Dynamical Model
Scg3
Smc4
Tcerg1l
Lhfp
Fmn2
Ltbp1
Bspry
Galnt9
Cadps2
Anxa3
Esrp1
Hnrnpd
Knl1
Frmd5
Cdk1
E.
Ranked Genes by Fit Likelihood
scVelo Fit Likelihood
Gene Rank
F.
Y End State Lineage Drivers
q = 0.0000 for all genes
Vim
S100a6
Ftl1
S100a10
Sparc
Anxa2
Fth1
Lgals3
Cald1
Anxa3
X End State Lineage Drivers
q = 0.0000 for all genes
Phgdh
Nt5dc2
Rps2
S100a10
Dctpp1
Tuba1b
Hist1h1b
Cks1b
Anxa5
Pclaf
G.
ChEA TFs from Enrichr Analysis
SCLC-Y Lineage Regulators
Adjusted P-value
Combined Score
X Lineage Regulators
H.
TF Network of SCLC-Y and X Lineage Drivers
Y driver
X driver
Both
Number of child nodes
I.
Cell Transport Potential by Phenotype
CTrP
Specialists in first half of time course
None
Generalist
Y
P/Y
A2
A/N

Supplemental Figure 5: Single cell velocity and plasticity for RPM time series, related to Figure 5. A. RNA velocity streams shown on UMAP across timepoints. **B.** Velocity length and confidence shows most cells have significant velocity vectors. **C.** Absorption probabilities by absorbing state. Each circle represents a timepoint, colored by probability of absorption at that end state. **D.** Top fit likelihood genes from scVelo dynamical model are shown as unspliced vs spliced RNA phase plots. Fit from scVelo model is overlaid. **E.** Plot of scVelo velocity genes show the distribution of model fit likelihoods. **F.** Lineage drivers for Y and X end states shown on UMAP. Q-value from CellRank analysis was 0 for all genes shown. **G.** TFs that are significant regulators of SCLC-Y and X lineages. Shown are TFs from ChEA only. **H.** TF network of lineage drivers. Left: TFs are colored by lineage; some TFs are shared between lineages and shown in purple. Only TFs that connect to this single main network are shown. Right: TFs are colored by number of child nodes they regulate in the network. P300 regulates the most child nodes at 38, making it the most central node to the lineage drivers in this time course. Other central TFs include MYC, CEBP family genes, JUN, and RUNX1/2. **I.** CTrP decreases from early-timepoint archetypes, including A/N, A2, and P/Y to the X archetype. While SCLC-Y is an absorbing state in the system, many of the Y specialists still have high plasticity.

TABLE S1: Enriched gene ontology sets for WGCNA gene modules, related to Figure 1 and Supplemental Figure 1.

	module	rank	BonferoniP	termName
GO:0010469	black	1	5.41916084864334E-11	regulation of signaling receptor activity
GO:0002376	black	2	1.36044649385637E-10	immune system process
GO:0019221	black	3	4.78960054486425E-09	cytokine-mediated signaling pathway
GO:0006952	black	4	2.42469343558444E-08	defense response
GO:0006955	black	5	4.69316153816772E-08	immune response
GO:0050896	black	6	8.49931302506631E-08	response to stimulus
GO:0071345	black	7	1.23794999100474E-07	cellular response to cytokine stimulus
GO:0010033	black	8	1.26245570838546E-07	response to organic substance
GO:0034097	black	9	2.01692523059288E-07	response to cytokine
GO:0071310	black	10	2.10423276086616E-07	cellular response to organic substance
GO:0006614	blue	1	1.83217154002747E-70	SRP-dependent cotranslational protein targeting to membrane
GO:0045047	blue	2	2.74659123121613E-70	protein targeting to ER
GO:0006613	blue	3	5.67127573760014E-70	cotranslational protein targeting to membrane
GO:0072599	blue	4	1.71906737898291E-67	establishment of protein localization to endoplasmic reticulum
GO:0006413	blue	5	6.60288167437702E-67	translational initiation
GO:0070972	blue	6	4.77255979775339E-59	protein localization to endoplasmic reticulum
GO:0000184	blue	7	1.43980928636161E-58	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
GO:0006412	blue	8	7.45975686394416E-51	translation
GO:0043043	blue	9	2.61702997425545E-49	peptide biosynthetic process
GO:0006612	blue	10	1.3760947336365E-48	protein targeting to membrane
GO:0007389	brown	1	9.00072374428462E-06	pattern specification process
GO:0009952	brown	2	1.40907099080107E-05	anterior/posterior pattern specification
GO:0009653	brown	3	1.97133081485968E-05	anatomical structure morphogenesis
GO:0048562	brown	4	0.000135596798135959	embryonic organ morphogenesis
GO:0009887	brown	5	0.00025335099713845	animal organ morphogenesis
GO:0035295	brown	6	0.000445762256451326	tube development
GO:0061448	brown	7	0.00104695406930455	connective tissue development
GO:0001501	brown	8	0.00149131167061171	skeletal system development
GO:0001503	brown	9	0.00189339053997136	ossification
GO:0048598	brown	10	0.00272237625947707	embryonic morphogenesis
GO:0030855	green	1	0.00524960878716573	epithelial cell differentiation
GO:0009913	green	2	0.00585302188687668	epidermal cell differentiation
GO:0007601	greenyellow	1	1.02307056861811E-23	visual perception
GO:0050953	greenyellow	2	1.58203099926516E-23	sensory perception of light stimulus
GO:0009583	greenyellow	3	4.20691807112603E-21	detection of light stimulus
GO:0009584	greenyellow	4	3.49056349597458E-18	detection of visible light
GO:0007600	greenyellow	5	5.78682847909344E-18	sensory perception
GO:0007602	greenyellow	6	1.92104592826975E-15	phototransduction
GO:0009582	greenyellow	7	2.9128012763726E-15	detection of abiotic stimulus
GO:0050877	greenyellow	8	7.5166345639531E-15	nervous system process
GO:0007603	greenyellow	9	2.68990946205477E-11	phototransduction, visible light
GO:0051606	greenyellow	10	4.76588097439568E-10	detection of stimulus
GO:0043062	magenta	1	7.7980511165155E-06	extracellular structure organization
GO:0016485	magenta	2	3.28489617884439E-05	protein processing
GO:0072378	magenta	3	0.00012511754875909	blood coagulation, fibrin clot formation

GO:0002526	magenta	4	0.000181335852420731	acute inflammatory response
GO:0007596	magenta	5	0.000399405014868462	blood coagulation
GO:0050817	magenta	6	0.000456295966454984	coagulation
GO:0042730	magenta	7	0.000475843568240935	fibrinolysis
GO:0007599	magenta	8	0.000520580010033955	hemostasis
GO:0009611	magenta	9	0.000616280707557106	response to wounding
GO:0030195	magenta	10	0.000623918863711097	negative regulation of blood coagulation
GO:0050909	pink	1	9.67488443467652E-06	sensory perception of taste
GO:0007606	pink	2	0.000449182758350224	sensory perception of chemical stimulus
GO:0050913	pink	3	0.00327207494497955	sensory perception of bitter taste
GO:00076001	pink	4	0.0231935191361072	sensory perception
GO:0007154	pink	5	0.0246589427099578	cell communication
GO:0023052	pink	6	0.0335258047170497	signaling
GO:0050907	pink	7	0.0433065934040735	detection of chemical stimulus involved in sensory perception
GO:0050912	pink	8	0.0433541770333006	detection of chemical stimulus involved in sensory perception of taste
GO:0030198	purple	1	5.42382880572989E-13	extracellular matrix organization
GO:00430621	purple	2	3.51409407746252E-12	extracellular structure organization
GO:00071551	purple	3	1.12511761810867E-10	cell adhesion
GO:0031589	purple	4	3.2945295161713E-06	cell-substrate adhesion
GO:0030199	purple	5	4.78999654317724E-06	collagen fibril organization
GO:0016477	purple	6	0.000343378507309123	cell migration
GO:0001568	purple	7	0.000683194027183539	blood vessel development
GO:0030334	purple	8	0.000890213885065437	regulation of cell migration
GO:0097435	purple	9	0.0011886153475409	supramolecular fiber organization
GO:0007275	purple	10	0.00160315082440526	multicellular organism development
GO:0000398	red	1	2.57738792798907E-06	mRNA splicing, via spliceosome
GO:0000375	red	2	3.5919199343137E-06	RNA splicing, via transesterification reactions
GO:0006396	red	3	9.12811484429963E-06	RNA processing
GO:0006397	red	4	1.10532705953669E-05	mRNA processing
GO:00512761	red	5	2.41219214764632E-05	chromosome organization
GO:0008380	red	6	2.62805829781647E-05	RNA splicing
GO:0016071	red	7	0.000934567581633522	mRNA metabolic process
GO:0006139	red	8	0.0013655298649833	nucleobase-containing compound metabolic process
GO:0010467	red	9	0.00432486805379817	gene expression
GO:0006725	red	10	0.00443057133065571	cellular aromatic compound metabolic process
GO:0007399	turquoise	1	3.9114341467648E-05	nervous system development
GO:0000902	turquoise	2	0.00402353158660299	cell morphogenesis
GO:0016192	turquoise	3	0.00604961042869958	vesicle-mediated transport
GO:0034341	turquoise	4	0.00859028231135073	response to interferon-gamma
GO:00023761	turquoise	5	0.00883676212382238	immune system process
GO:0017156	turquoise	6	0.0240986850760173	calcium ion regulated exocytosis
GO:00069551	turquoise	7	0.0302567580390949	immune response
GO:0031175	turquoise	8	0.0303206741552842	neuron projection development
GO:0007269	turquoise	9	0.0335084605142006	neurotransmitter secretion
GO:0051648	turquoise	10	0.0361443406404716	vesicle localization
GO:0022613	yellow	1	5.87050662329621E-31	ribonucleoprotein complex biogenesis
GO:0042254	yellow	2	2.23255954053755E-30	ribosome biogenesis
GO:00325431	yellow	3	6.63574196367001E-26	mitochondrial translation
GO:0016072	yellow	4	6.30220844813267E-25	rRNA metabolic process

GO:00063961	yellow	5	2.47108038978792E-23	RNA processing
GO:0006364	yellow	6	5.07588017100682E-23	rRNA processing
GO:00346411	yellow	7	8.32096755475593E-23	cellular nitrogen compound metabolic process
GO:00064151	yellow	8	4.06699051891866E-20	translational termination
GO:00701251	yellow	9	9.04338768501359E-20	mitochondrial translational elongation
GO:00701261	yellow	10	2.77432045935756E-19	mitochondrial translational termination

Table S2: T-ratios for Polytopes fit to Human Cell Lines, related to Figure 1

# of vertices	P-value	T-ratio
3	0.508	0.520
4	0.059	0.247
5	0.034	0.107
6	0.016	0.042
7	0.001	0.020

Table S3: Hypergeometric test for archetype enrichments from polytopes fit to human cell lines with different numbers of vertices, related to Figure 1

Original archetype (cell lines only) **k* = 5**	Matching archetypes from cell lines **k* = 4**	p-value of Hypergeometric test on enrichments	Matching archetypes from cell lines **k* = 6**	p-value of Hypergeometric test on enrichments
1 (Y)	1	$P_1 = 0$	1 & 2	$P_1 = 0$, $P_2 = 1.94 \text{ X } 10^{-27}$
2 (P)	1	$P_1 = 3.73 \text{ X } 10^{-15}$	1 & 2	$P_1 = 2.14 \text{ X } 10^{-19}$, $P_2 = 0$
3 (N)	2	$P_2 = 0$	3	$P_3 = 0$
4 (A2)	3	$P_3 = 0$	4 & 5	$P_4 = 0$, $P_5 = 5.21 \text{ X } 10^{-131}$
5 (A)	4	$P_4 = 4.38 \text{ X } 10^{-90}$	6	$P_6 = 2.52 \text{ X } 10^{-85}$

Table S4: Cell line archetypes compared to tumor + cell line archetypes, related to Figure 1

Original (**cell lines** only) archetype [k* = 5]	Matching archetypes from **combined dataset** [k* = 5]	p-value of Hypergeometric test on enrichments
1 (Y)	1 & 4	$P_1 = 0$, $P_4 = 0$
2 (P)	1	$P_1 = 2.5 \text{ X } 10^{-96}$
3 (N)	2 & 4	$P_2 = 1.6 \text{ X } 10^{-221}$, $P_4 = 0$
4 (A2)	3	$P_3 = 0$
5 (A)	5	$P_5 = 1.39 \text{ X } 10^{-17}$

Table S5: Table of q values for gene enrichment at archetypes for bulk RNA-seq data. Q <.1 is considered significant; Mann-Whitney test. Related to Figure 2.

Table S6: Table of q values for gene set/task enrichment at archetypes for bulk RNA-seq data using ConsensusPathDB. Related to Figure 1.

Table S7: Table of q values for Cancer Hallmark Gene Set enrichment at archetypes for bulk RNA-seq data. Q <.1 is considered significant; Mann-Whitney test. Related to Table 1.

archetype #	Feature Name	P value (Mann-Whitney)	Median Difference	Mean Difference	Significant after Benjamini-Hochberg correction?	Is first bin maximal?
SCLC-A2	Evading Immune Destruction	0.00067758	0.21293	0.17362	1	1
SCLC-A2	Tumor-Promoting Inflammation	0.0039469	0.18223	0.13398	1	1
SCLC-A2	Inducing Angiogenesis	0.015203	0.16121	0.09591	1	1
SCLC-P	Genome Instability and Mutation	0.00054072	0.19115	0.17446	1	1
SCLC-P	Reprogramming Energy Metabolism	0.0029794	0.10543	0.11382	1	1
SCLC-Y	Inducing Angiogenesis	2.93E-06	0.35277	0.35909	1	1
SCLC-Y	Resisting Cell Death	2.08E-06	0.28998	0.26927	1	1
SCLC-Y	Evading Immune Destruction	0.00042998	0.21291	0.20354	1	1
SCLC-Y	Genome Instability and Mutation	5.21E-05	0.20138	0.20953	1	1
SCLC-Y	Sustaining Proliferative Signaling	4.84E-05	0.20129	0.17892	1	1
SCLC-Y	Evading Growth Suppressors	6.05E-05	0.19485	0.18497	1	1
SCLC-Y	Tumor-Promoting Inflammation	0.00020448	0.18833	0.1922	1	1
SCLC-Y	Enabling Replicative Immortality	6.51E-05	0.18322	0.15924	1	1
SCLC-Y	Activating Invasion and Metastasis	0.00065623	0.16008	0.14933	1	1
SCLC-Y	Reprogramming Energy Metabolism	0.00037661	0.14285	0.14035	1	1

Table S8: NE stem cell and transit-amplifying genes, related to Figure 2.

	NEstem	TA	NE
Notch2	1	1	0
Hes1	1	1	0
Nrarp	1	0	0
p53	0	0	0
Rb	0	0	0
Hes6	1	0	0
Jag1	1	0	0
Piezo2	1	0	1
Ret	1	0	1
Ptprz1	1	0	1
Scnn1a	1	0	1
Trpm7	1	0	1
Chga	1	0	1

Ddc	1	0	1
Gad1	1	0	1
Pcsk1	1	0	1
Ptprn	1	0	1
Scg2	1	0	1
Snap25	1	0	1
Syt7	1	0	1
Casr	1	0	1
Kcnk2	1	0	1
Slit	0	0	1
Resp18	0	0	1
Rapgef4	0	0	1
Slitrk2	0	0	1
Fn13	0	0	1
Rn18s	0	0	1
Gm6548	0	0	1
Zrsr1	0	0	1
Ascl1	1	0	1
Calca	0	0	1
Scgb3a2	0	1	0
Cbr2	0	1	0
Lyz2	0	1	0
Foxj1	0	1	0
Myb	0	1	0
Sftpc	0	1	0
Fmo2	0	1	0
Cyp2f2	0	1	0
Sftpa1	0	1	0
Sparc	0	1	0
Serping1	0	1	0
Lyz1	0	1	0
Ppic	0	1	0
Fstl1	0	1	0
Anxa3	0	1	0
Foxq`	0	1	0
Cd74	0	1	0
Ctsh	0	1	0
Ldhb	0	1	0
Ptgr1	0	1	0
Slcbal4	0	1	0
Sftpd	0	1	0

Table S9: Archetype gene signature expression data, related to Figure 3.

Gene	SCLC-A	SCLC-A2	SCLC-N	SCLC-P	SCLC-Y
TAGLN3	8.85	6.10	5.62	1.55	-0.70
RBP1	8.16	5.32	5.73	2.88	2.54
ISL1	7.48	3.40	2.62	1.36	0.07
ELAVL3	7.35	4.22	5.78	0.18	0.50
PCP4	7.29	5.64	3.03	1.91	0.75
TCEAL2	6.82	2.07	4.43	-0.50	1.10
SOX1	6.44	3.00	1.18	1.19	-0.50
GHRH	6.01	1.57	0.35	1.89	0.23

NSG1	5.97	1.82	4.31	3.31	0.98
CD200	5.86	2.12	3.62	1.11	-0.80
MBNL3	5.57	3.53	2.06	3.10	0.51
PTN	5.43	2.52	2.25	-0.85	3.07
DCX	5.02	1.08	4.86	-0.83	-0.63
SHD	4.86	0.64	4.18	-1.04	0.43
GAD2	4.31	2.94	1.24	-0.33	-0.01
FGD3	4.30	2.77	0.85	1.67	-0.04
ILDR2	4.15	2.50	1.33	0.68	-0.02
NNAT	3.81	1.52	2.27	1.57	1.05
FLI1	3.08	2.86	0.28	1.61	-0.32
AVP	1.21	0.20	0.23	-0.22	-0.24
ASCL1	8.02	10.81	1.73	1.05	0.88
GRP	4.85	9.72	0.85	-1.38	0.90
ELF3	3.93	8.77	-0.12	4.77	1.63
SCNN1A	6.18	8.37	0.43	4.49	-0.31
CEACAM5	2.21	8.25	-0.38	0.71	0.27
WFDC2	6.20	7.63	1.45	3.38	2.61
MS4A8	3.76	7.26	0.31	0.06	-1.11
TMEM176A	3.18	7.23	0.98	0.83	0.22
FAM3B	2.66	7.17	-0.47	2.27	0.16
TMEM176B	2.84	6.91	0.98	0.27	0.74
CALCA	0.28	6.91	-0.57	1.83	1.56
TSPAN1	1.80	6.18	0.69	0.03	2.33
TSPAN8	1.17	5.95	-0.24	1.06	1.65
NPTX1	2.59	5.93	4.09	0.87	2.50
SCIN	1.53	5.89	0.51	2.84	0.00
RASSF6	5.10	5.77	0.31	4.39	-0.21
KLK11	2.41	5.39	-1.02	2.29	0.94
SKAP1	0.61	5.11	0.32	0.62	0.96
KLK12	1.39	5.06	-1.06	2.26	0.34
GJB1	0.83	4.81	-0.38	0.53	-0.64
AOC1	0.16	3.19	-0.52	0.59	0.18
NEUROD1	1.81	0.68	7.19	0.29	0.33
KCNQ2	3.58	0.15	6.83	0.02	2.16
OLFM1	3.78	0.93	6.66	-0.77	4.40
CAMKV	2.43	1.15	6.11	1.74	0.26
MFAP4	2.68	-0.03	5.81	0.26	1.40
PPP1R17	1.76	-0.13	5.57	-0.38	-0.46
CNTN1	2.06	2.49	5.30	2.68	0.98
CERKL	2.06	0.22	5.16	0.19	0.99
ADCYAP1R1	0.96	-0.57	4.64	0.73	0.17
SSTR2	1.93	0.06	4.61	-0.04	0.14
RBFOX3	1.81	-0.55	4.58	-0.73	1.88
SLC38A5	0.24	0.26	4.50	1.36	0.56
CTNND2	2.16	0.25	4.32	2.58	0.28
TSPAN18	0.86	0.47	4.22	0.36	1.83
ANGPTL2	0.84	0.25	4.07	0.22	0.75
KCNJ3	0.60	0.88	4.03	0.96	0.19
NHLH2	1.50	-0.16	3.85	-0.27	-0.14
NEUROD2	0.77	-0.33	3.77	0.07	-0.29
PLCH2	0.56	0.17	3.70	0.06	0.02

ZFPM2	0.28	-0.21	3.11	0.48	1.21
NEUROD6	0.45	-0.51	3.01	-0.39	-0.23
YBX3	2.63	4.28	3.54	8.76	8.51
AVIL	1.34	3.02	0.60	6.85	1.26
SPATS2L	3.43	3.70	2.97	6.64	6.59
ANXA4	2.21	4.39	2.59	6.39	6.31
POU2F3	0.40	0.19	-0.07	6.26	0.70
LRMP	0.44	0.28	1.14	6.08	0.55
PLCG2	0.20	1.93	0.44	5.72	1.53
PLA2G4A	-0.22	0.10	0.68	4.98	1.05
LGALS3	1.23	1.87	1.13	4.96	4.72
RGS13	-0.11	0.07	0.08	4.89	-0.03
BMX	0.31	0.20	0.47	4.81	-0.15
AZGP1	0.55	2.99	-0.13	4.78	1.28
EHF	1.54	3.04	-0.78	4.75	1.61
CRYM	1.33	3.46	0.14	4.68	1.84
GAL	0.68	1.01	2.72	4.37	2.20
SOX9	3.33	2.44	1.92	3.92	3.37
GFI1B	-0.04	0.02	0.01	3.52	0.14
TRIM58	0.14	0.40	-0.07	3.30	1.73
VSNL1	0.47	0.42	1.19	2.88	0.97
LGALS1	1.76	4.37	2.35	2.19	11.42
VIM	1.83	1.45	3.71	3.84	11.02
GSTP1	6.06	6.71	4.65	9.23	10.31
ANXA1	1.18	2.27	0.80	7.01	9.55
IFITM3	1.15	2.80	0.64	5.09	9.07
CNN2	1.15	2.48	2.18	3.58	8.33
FSTL1	3.19	0.25	3.02	1.47	8.22
TPM2	1.81	2.10	2.45	4.19	7.97
YAP1	0.69	0.18	0.31	2.80	6.62
CRIM1	0.15	1.82	1.54	3.34	6.57
MRC2	0.84	0.58	2.02	0.30	6.43
CAV1	0.20	0.69	0.70	2.19	6.42
GPX8	0.58	0.12	1.46	2.16	6.31
MAGEA4	0.70	1.83	1.99	2.12	6.18
MICA	0.19	0.74	0.24	1.70	6.17
AHNAK	0.52	1.55	0.87	3.31	6.05
TNFRSF10B	0.60	1.10	1.09	3.77	6.02
OSMR	0.07	0.88	0.23	1.86	5.80
EMP1	0.58	0.50	0.65	1.47	5.76
HOXC10	1.43	2.18	2.77	4.56	5.39
MSRB3	0.04	0.03	0.72	0.95	5.21
AXL	-0.06	0.12	0.18	0.70	5.05
MYL9	-0.29	0.37	0.36	1.68	4.50
SLPI	0.83	3.00	-0.59	0.78	3.07

Table S10: Adjusted p-values for bulk archetype signature enrichment for each RPM archetype, related to Figures 5 and S6.

	SCLC-A	SCLC-A2	SCLC-N	SCLC-P	SCLC-Y
1	1	1	1	0.0	0.0

2	1	1	1	0.184	1
3	1	1	1	1	0.0
4	0.0	1	1	1	1
5	1	0.0	0.0	1	1
6	1	1	1	1	0.0

Table S11: Somatic variants in normal tissue and days 4 and 23 of RPM time series.