

Figure S1

Figure S1. Related to Figures 1 and 2 and to Table S1.

(A) Clinical attributes of the SCLC cohort in the single-cell atlas (N = 21 tumor samples from 19 patients). Early-line and later-line treatment history is shown in the top two rows with treatment order specified in the order of full square, thin rectangle, and small square. Other annotations include smoking status, tissue, stage, disease status, and SCLC subtype of the major subclone.

(B) UMAP projection of the SCLC cohort at the global level (N=155,098 cells) colored by cell type (top) and by Shannon entropy of patients in a k-nearest neighborhood of each cell (k=30, **STAR Methods**) (bottom).

(C) Flowchart detailing the preprocessing and analytical steps taken at iterative subsets of the single-cell cohort.

(D) Oncoprint showing mutational information for all SCLC samples with matched MSK-IMPACT bulk targeted DNA sequencing available (N = 14 patients), arranged by genes (rows, sorted by most to least recurrent) and patients (columns). Point mutations are depicted by bars, copy number changes by full squares. Bar plot shows tumor mutation burden (TMB).

(E) TMB of SCLC samples, grouped by subtype. Error bars span 25th to 75th percentile.

(F) CNV burden of SCLC, LUAD, and normal lung samples. CNV burden is estimated by percent of genome altered by CNV (left) or Pearson's correlation to CNV profiles of the matched MSK-IMPACT data where available (right). Cancer cells were called based on a threshold of either percent of genome with CNV > 0.1, or bulk correlation $\rho > 0.2$ (**see STAR methods**).

(G) UMAP projection of SCLC cells colored by MAGIC-imputed expression *MYCL* (left), *MYC* (middle), and *MYCN* (right) (**STAR Methods**). Consistent with previous literature (Rudin et al., 2019), we observed that *MYCL* expression was associated with SCLC-A (though one SCLC-N sample, Ru1293, had predominant *MYCL* expression), and that *MYC* expression was relatively low in SCLC-A and highest in the SCLC-P sample (Ru1322) We observed strong *MYCN* expression particularly in one predominant SCLC-N sample (Ru1215).

(H) Heatmap showing immunohistochemistry data available for staining of *ASCL1*, *NEUROD1*, *POU2F3* and *YAP1* transcription factors in the SCLC cohort. Hatched boxes indicate marker staining that was not available due to limited tissue availability. Row annotations indicate the predominant SCLC subtype identified in our scRNA-seq dataset.

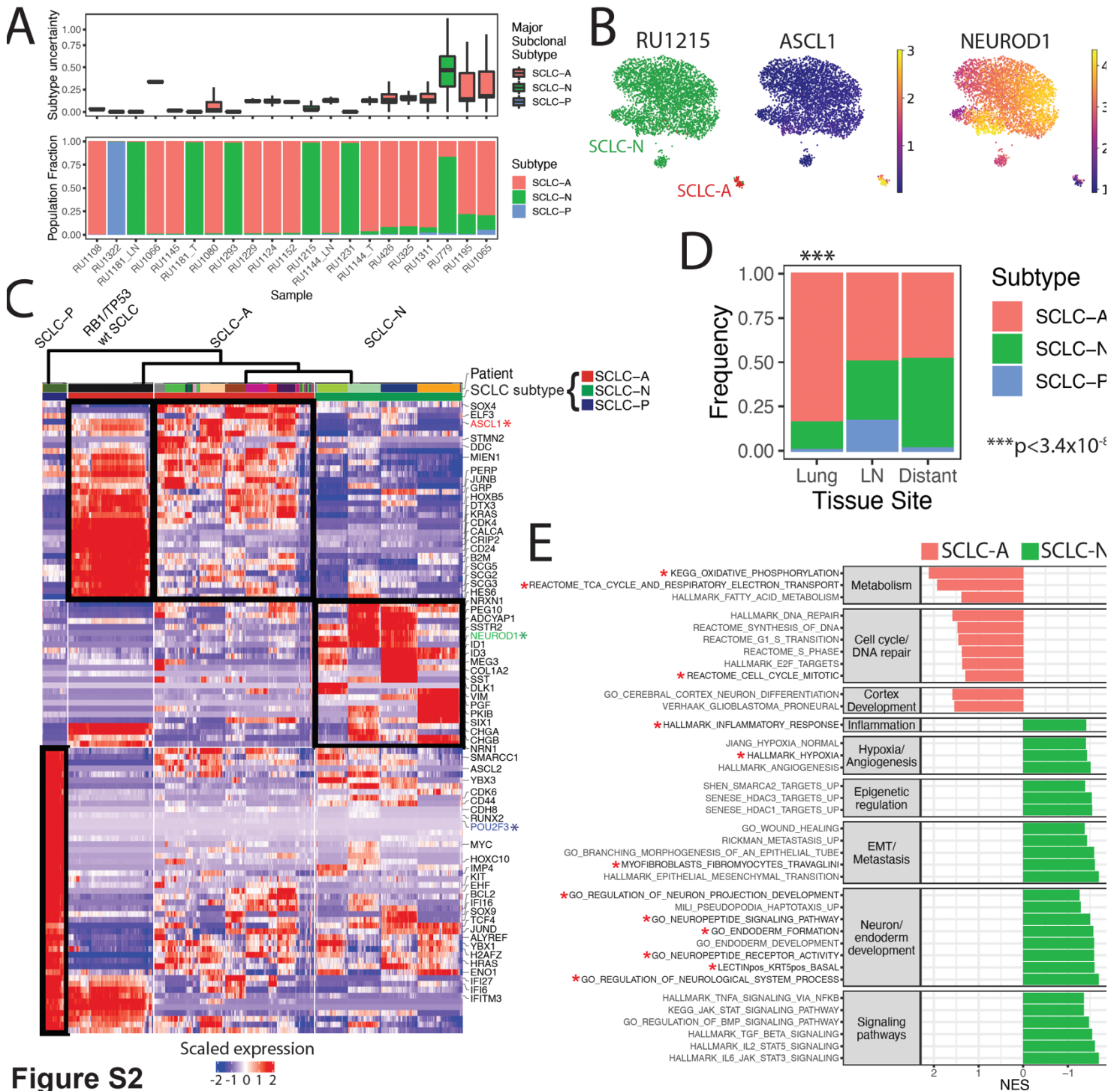


Figure S2

Figure S2. Related to Figures 1 and 2 and to Table S1.

(A) Statistics of SCLC subtype representation in samples (columns). Top: SCLC subtype uncertainty per cell as measured by Shannon entropy of SCLC subtype probabilities. Bottom: SCLC subtype fractions in each sample, as determined by maximum likelihood of Markov absorption probabilities. Error bars span 25th to 75th percentile.

(B) UMAP of sample Ru1215 showing admixed SCLC-A and SCLC-N subtyping, supported by *ASCL1* and *NEUROD1* expression. Cells are colored by maximum likelihood of Markov

absorption probability (left) and MAGIC-imputed gene expression of ASCL1 (middle) and NEUROD1 (right). (**STAR Methods**).

(C) Single-cell gene signatures of SCLC subtypes, with scaled expression (z-score) of differentially expressed genes (DEGs) per SCLC subtype (rows) for each cell (columns). Expression values are imputed using MAGIC (k=30, t=3). Representative DEGs, including canonical TFs are labeled. Top dendrogram produced by hierarchical clustering using Pearson's correlation as a distance metric with complete linkage. Top annotations include patient and SCLC subtype.

(D) Bar plot of frequency of SCLC subtype per tissue site (primary lung vs regional lymph node (LN) vs distant metastasis). The frequency of each subtype per sample was first determined and then summed by tissue site. A final frequency was normalized by the number of samples to produce the stacked bar plot. In this way, each sample was considered with equal weight. SCLC-A was significantly enriched in primary lung, whereas non-SCLC-A subtypes are enriched in nodal and distant metastases (Dirichlet regression, $p < 3.4 \times 10^{-8}$, **STAR Methods**).

(E) Gene programs significantly enriched in SCLC-A (red bars) or SCLC-N (green bars) using Gene Set Enrichment Analysis (Benjamini-Hochberg adjusted FDR < 0.05, absolute value of normalized enrichment score (NES) > 1) (**Table S3**). Enriched pathways also confirmed in an independent cohort of bulk RNA-seq are indicated by red asterisks (**STAR Methods**).

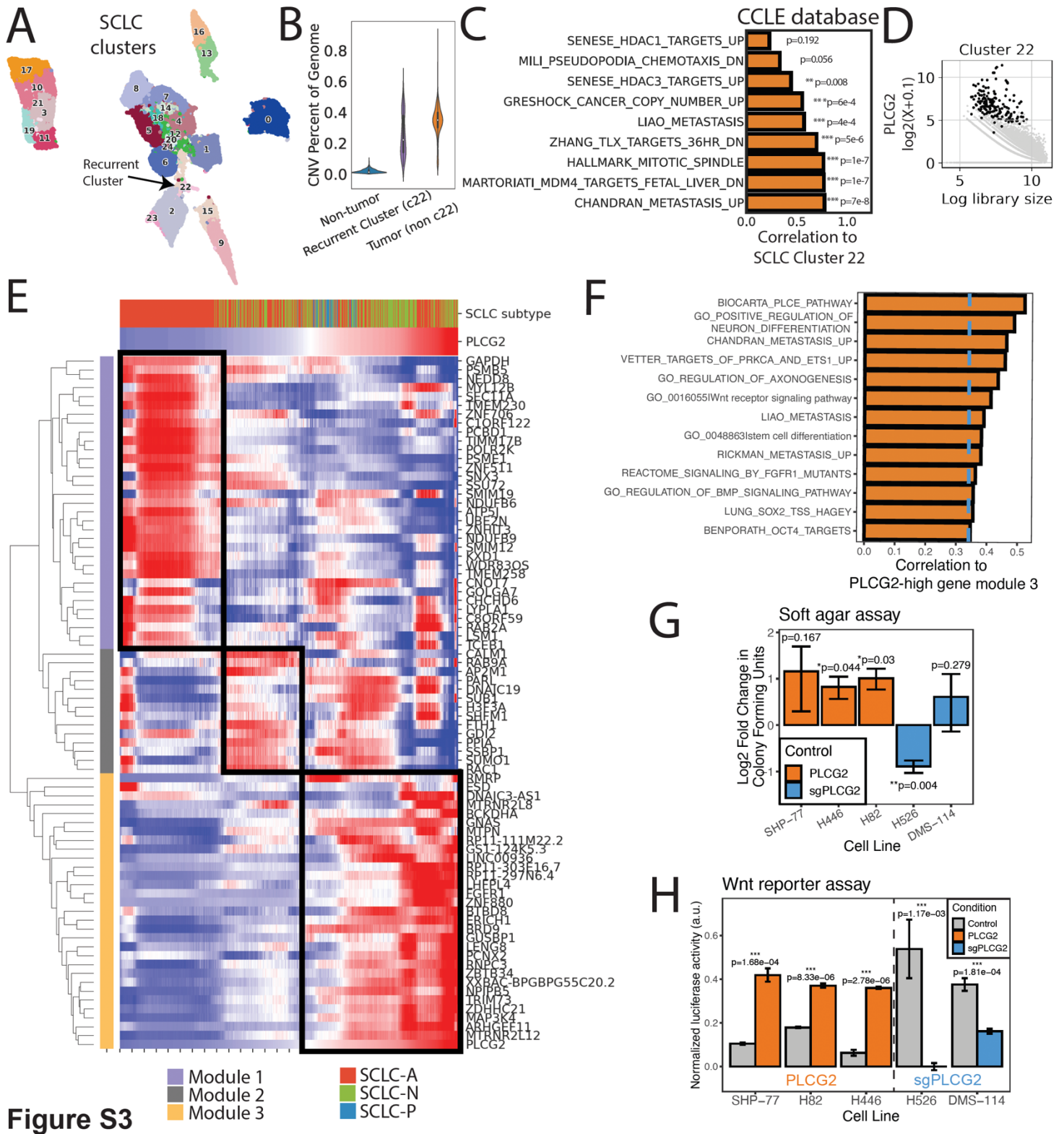


Figure S3

Figure S3. Related to Figures 3 and 4 and to Table S2.

(A) UMAP projection of Phenograph clusters in the SCLC compartment with arrow indicating the recurrent cluster 22.
 (B) CNV burden of recurrent cancer cluster 22 compared to normal lung and other SCLC cells.

CNV burden is estimated by percent of genome altered by inferred CNV (**see STAR methods**).

(C) Barplot showing Spearman's correlation of pathways enriched in the recurrent cancer cluster 22 in scRNA-seq to the recurrent cluster gene signature in the CCLE microarray database (N=54; **STAR Methods**).

(D) Unimputed *PLCG2* expression vs log library size in recurrent cluster 22. Library size is transformed by natural log with pseudocount of 1.

(E) Scaled expression (z-score, imputed by MAGIC with $k = 30$, $t = 3$) of genes with high knnDREMI conditioned on *PLCG2* > 1 (rows), with cells ordered by *PLCG2* expression (columns). For visualization, expression was smoothed over the ordered cells with a rolling window of 100 cells. Hierarchical clustering of genes on the unsmoothed, imputed expression was performed with complete linkage and Pearson correlation as a distance metric, identifying 3 gene modules that predict low, medium, and high *PLCG2* (purple, gray, and yellow respectively) (**Table S13**). Top annotations include *PLCG2* expression and SCLC subtype.

(F) Pathways with average z-scores of gene expression that are highly correlated (>95th percentile, indicated by blue dashed line) with the average z-score of gene expression in the high-*PLCG2* gene module (**STAR Methods, Table S14**).

(G) Barplot showing log₂ fold change in the number of colonies generated by *PLCG2*-overexpressing (SHP77, H446, and H82) and *PLCG2*-KO (H526 and DMS-114) polyclonal cell lines in soft agar colony formation assay, normalizing to the control condition. Error bars: standard deviation.

(H) Barplot showing luminiscence activity of a luciferase Wnt reporter in *PLCG2*-overexpressing (SHP77, H446, and H82) and *PLCG2*-KO (H526 and DMS-114) polyclonal cell lines. Error bars: standard deviation.

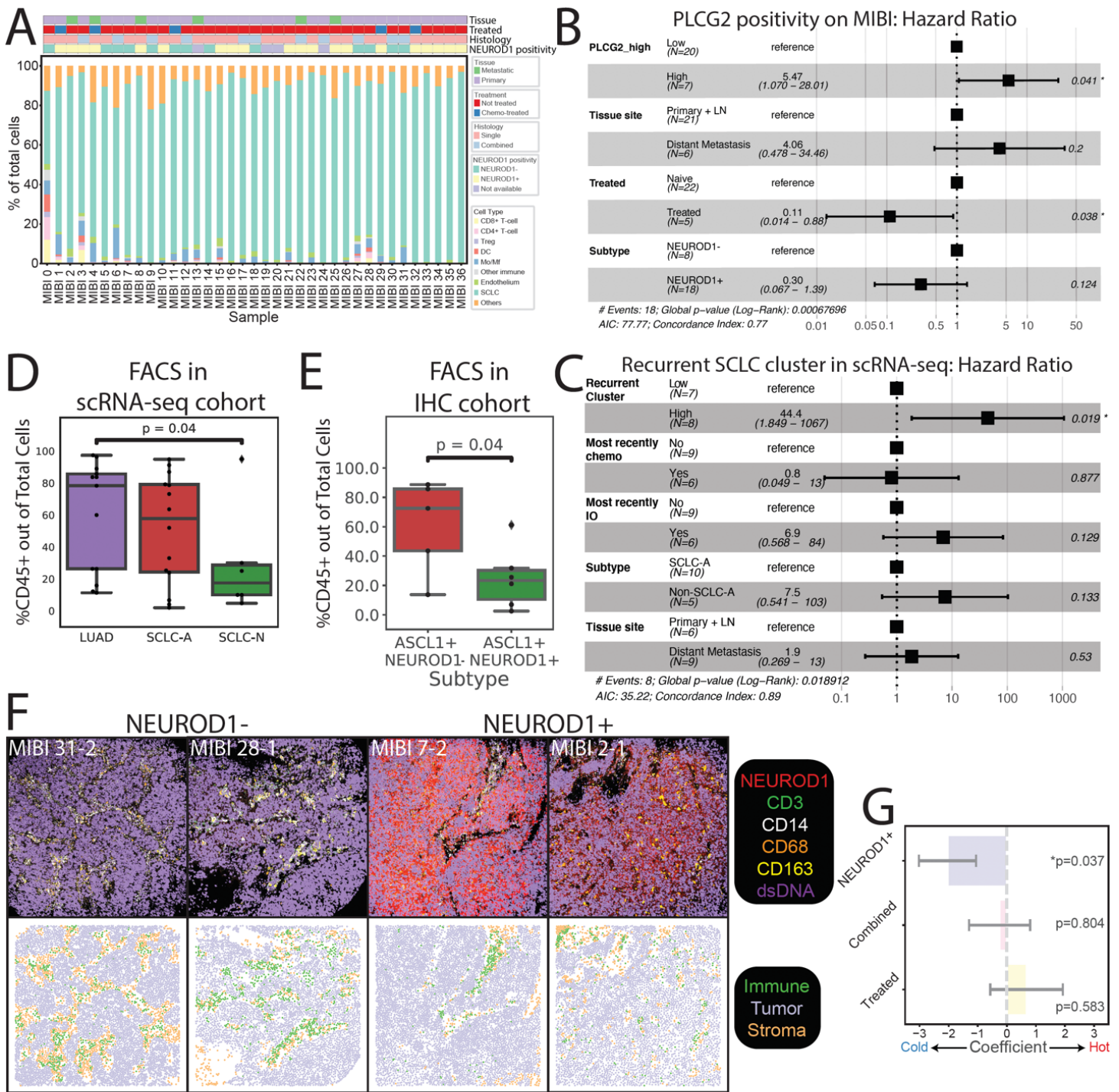


Figure S4

Figure S4. Related to Figure 4 and 5 and to Table S1.

(A) Stacked barplot showing the per-patient abundance of cell types identified by MIBI-TOF.

(B) Multivariate analyses (Cox regression) for overall survival in samples included in an independent cohort of SCLC samples assessed by MIBI-TOF, comparing in patients with high vs

low *PLCG2* positivity (> 7% of SCLC cells, comparing patients with high vs low *PLCG2* positivity (>7% vs ≤7% of SCLC cells with high *PLCG2* staining intensity). The regression model was adjusted for presence of SCLC subtype (classical vs variant), treated with chemotherapy vs naive, and distant metastasis vs primary/regional lymph node. Only samples that were ever extensive-stage (ultimately with metastatic relapse) were considered.

(C) Multivariate analyses (Cox regression) for overall survival in samples included in the single cell cohort, comparing tumors with >0.75% composition of high *PLCG2*-expressing recurrent cluster versus the rest. The regression model was adjusted for presence of SCLC subtype (classical vs variant), most recently treated with chemotherapy alone vs immunotherapy vs naive, and distant metastasis vs primary/regional lymph node. Only samples that were ever extensive-stage (ultimately with metastatic relapse) were considered.

(D) Boxplot of the percentage of CD45+ infiltrate in LUAD vs SCLC-A vs SCLC-N samples in our single-cell cohort (Mann-Whitney test), as assessed by flow cytometry. Error bars: 95% confidence interval.

(E) Boxplot of the percentage of CD45+ infiltrate in an independent cohort of SCLC tumors stratified by SCLC subtype, as assessed by flow cytometry. For SCLC subtype, we compared ASCL1+ NEUROD1- vs ASCL1+ NEUROD1+ samples. No ASCL1- NEUROD1+ samples were available. Differences between conditions were assessed using Mann-Whitney test. Error bars: 95% confidence interval.

(F) Comparison of MIBI images depicting NEUROD1- SCLC tumors (MIBI 31 FoV 2 and MIBI 28 FoV 1 on the left) and NEUROD1+ SCLC tumors (MIBI 7 FoV 2 and MIBI 2 FoV 1 on the right) (each FoV 800 μm by 800 μm), illustrating differences in immune abundance and sequestration. Top: Color overlay of NEUROD1 (red), CD3 (green), CD14 (white), CD68 (orange), CD163 (yellow), and dsDNA (violet) channels. Bottom: FoV from the top now visualized with segmented cancer cells using Mesmer, represented by dots colored by cell type (immune, tumor, and stroma).

(G) Barplot of logistic regression coefficients of 4 covariates (NEUROD1 positivity, combined vs single histology, chemotherapy-treated vs naive, primary vs metastatic) fitted to explain immune hot vs cold tumors (number of immune cells > 250 in a tumor) (Student's t-test). Error bars indicate standard error. Coefficient for primary vs metastatic covariate is not shown due to quasi-complete separation.

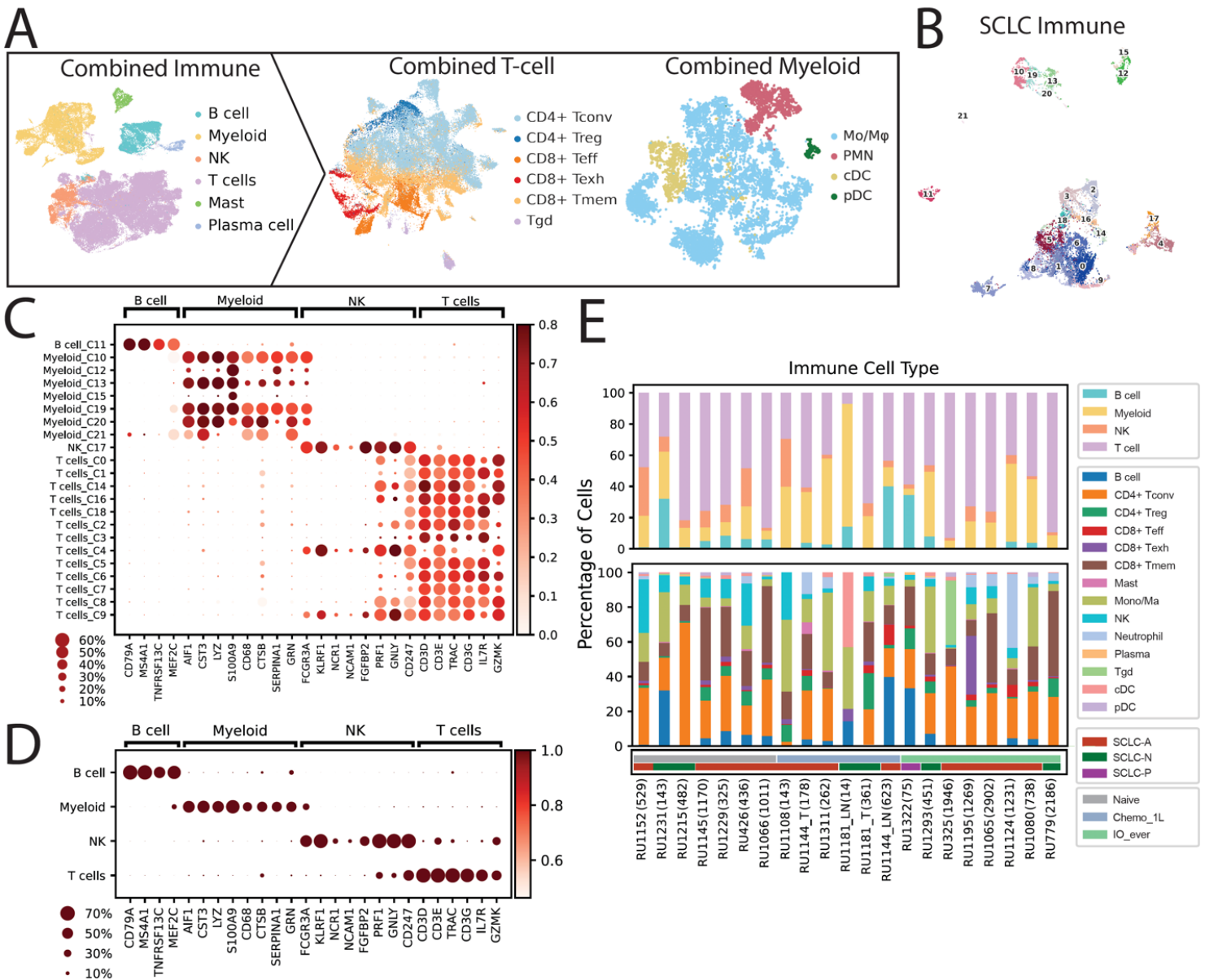


Figure S5

Figure S5. Related to Figure 5 and to Table S15, S16.

(A) UMAP projections of the immune subsets of SCLC with LUAD and adjacent normal lung reference at the global level (top, n=73,047 cells), T-cell compartment (middle, n=46,140 cells), and myeloid compartment (bottom, n=14,072 cells).

(B) UMAP projection of all immune cells from 21 SCLC samples (N = 16,475 cells), annotated by Phenograph clusters.

(C) Dot plots show relative frequency and expressing cells, and mean normalized expression of canonical immune cell type markers per Phenograph cluster.

(D) Dot plots show relative frequency and expressing cells, and mean normalized expression of canonical immune cell type markers per coarse cell type.

PLCG2 = *PLCG2* overexpression, sgPLCG2 = CRISPR knockout.

(E) Stacked barplots of the percentage of coarse immune cell type (top) and fine immune cell type (bottom) in each sample, annotated by SCLC subtype and treatment. Total number of immune cells per sample is indicated in parentheses.

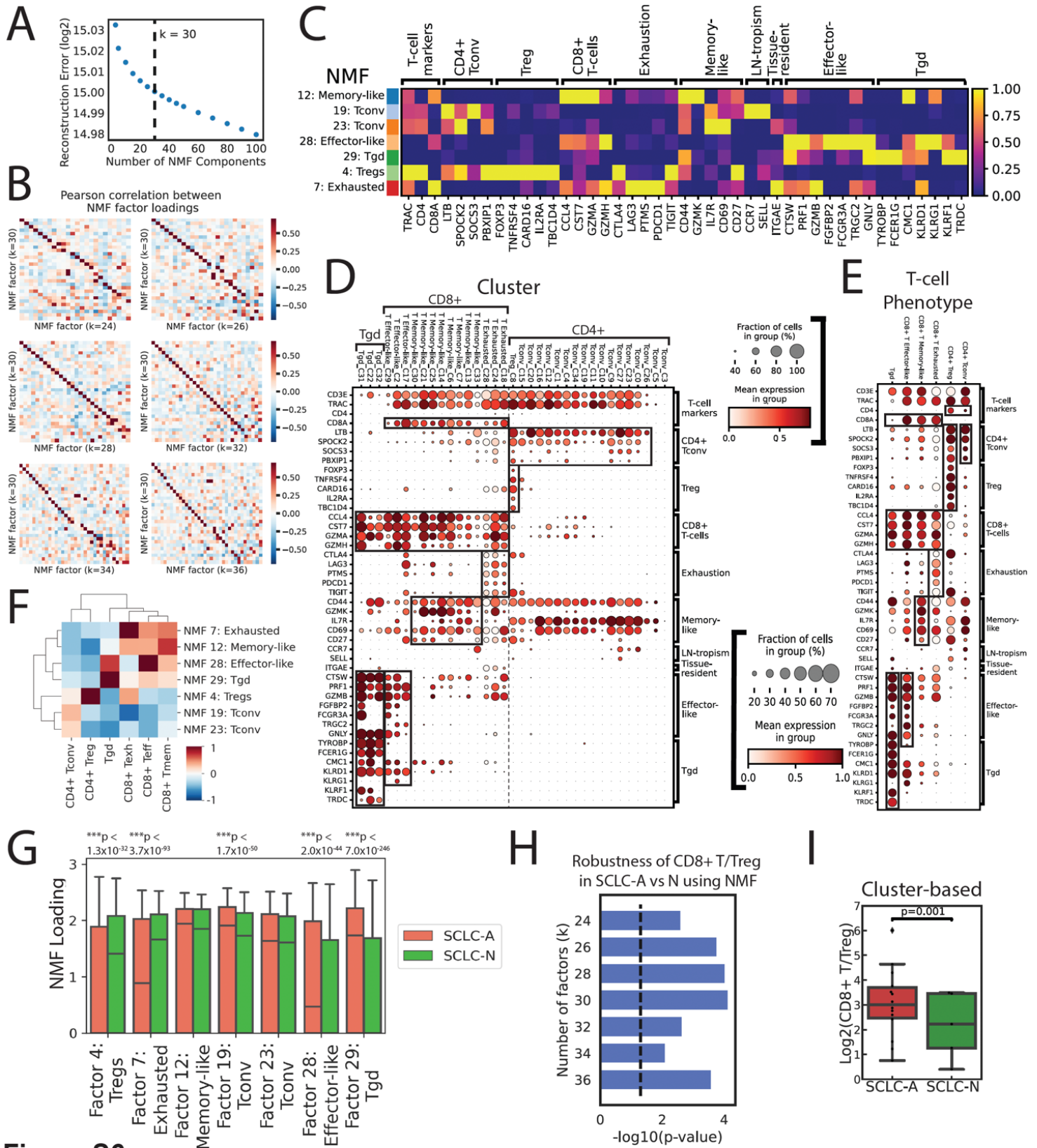


Figure S6

Figure S6. Related to Figure 5.

- (A) Scatterplot of log₂ reconstruction error vs the number of NMF components k . The dashed line at $k=30$ factors represents the kneepoint of the log₂ reconstruction error.
- (B) Robustness analysis of NMF factors. Heatmaps show correlation between NMF cell loadings identified at $k=30$ components and over a range of k (24-36).
- (C) Heatmap of NMF gene loadings for factors associated with T-cell phenotype. Genes are grouped by T-cell function. Each factor is z-scored across genes, and loadings are subsequently scaled from 0 to 1 across factors (STAR Methods). Dot plots show gene markers of each cluster (D) and T-cell phenotype (E). Dot size = % cells expressing gene; dot color = mean expression scaled from 0 to 1.
- (F) Heatmap of cell loading of select NMF factors averaged within each T-cell phenotype based on clusters. The resulting averaged loading was z-scored across factors (**STAR Methods**).
- (G) Barplot comparing scaled NMF cell loadings for factors related to T-cell phenotype between SCLC-A and SCLC-N (two-sample t-test adjusted for treatment and tissue site, error bars: 95% confidence interval; **STAR Methods**). *** $p < 0.001$.
- (H) Robustness analysis of the decreased CD8+ T/Treg ratio in SCLC-N vs SCLC-A based on NMF. Barplot shows the significance or $-\log_{10}(p\text{-value})$ of the difference in ratio between subtypes over a range of k (number of components). (I) Barplot comparing CD8+ T_{eff}/Treg log ratio for based on clusters associated with T-cell phenotype in SCLC-A versus SCLC-N in our single-cell cohort (N=19), adjusted for treatment and tissue site (weighted t-test, error bars: 95% confidence interval; **STAR Methods**).

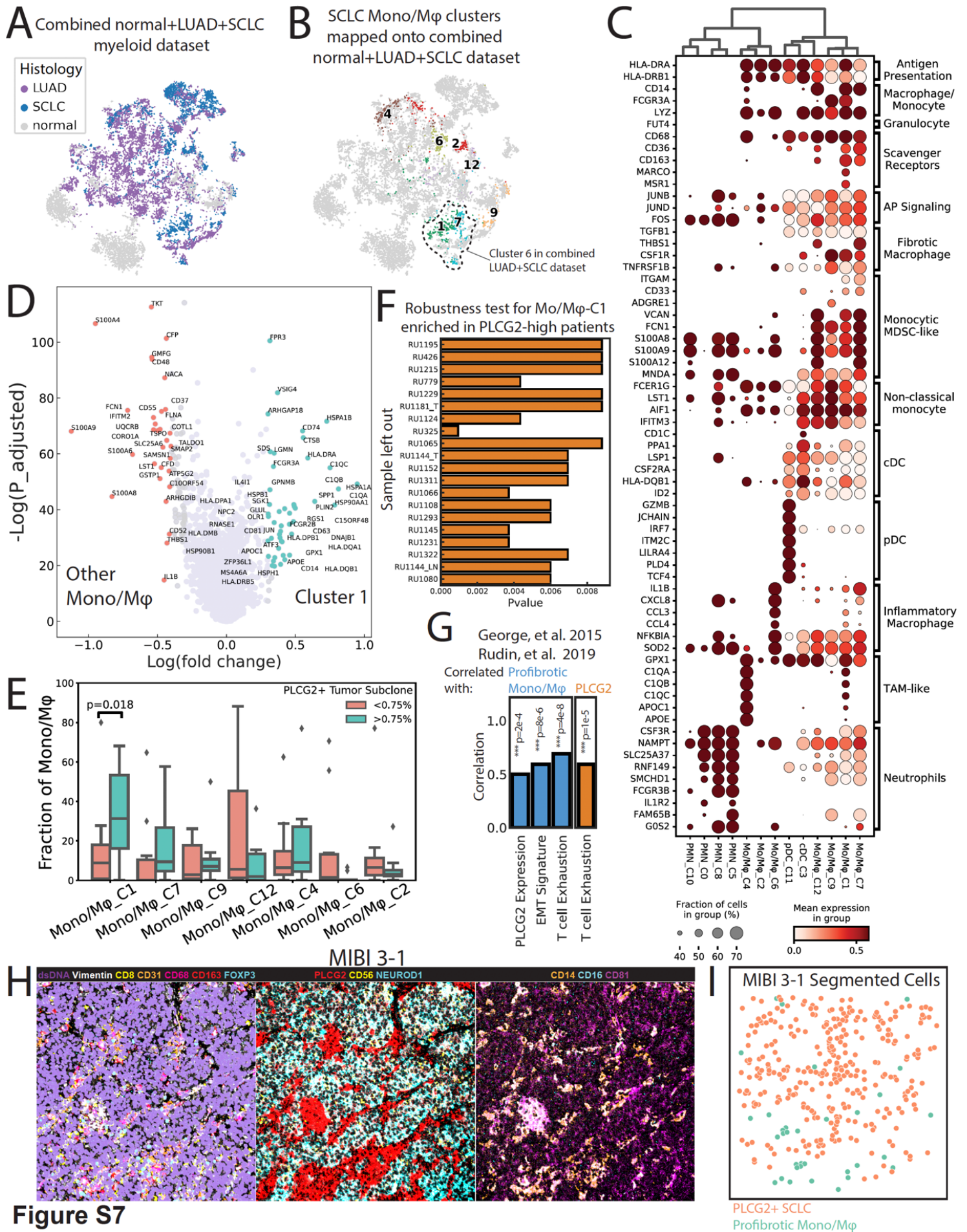


Figure S7

Figure S7. Related to Figures 6 and 7.

UMAP projections of combined myeloid compartment (SCLC, LUAD, and normal adjacent lung), annotated by (A) tumor histology and (B) Mono/M ϕ clusters identified from the SCLC-restricted myeloid compartment. In (B), the dashed outline indicates cells identified as cluster 6 in the clustering of the combined myeloid dataset of SCLC and LUAD samples.

(C) Dot plot of SCLC myeloid subsets and expression of relevant markers. Myeloid subsets are ordered based on hierarchical clustering of the first 50 principal components (**STAR Methods**). Dot size = % cells expressing gene; dot color = mean expression scaled from 0 to 1.

(D) Volcano plot of differentially expressed genes in SCLC Mono/M ϕ cluster 1 of the SCLC myeloid compartment versus other Mono/M ϕ subsets (**Table S20**).

(E) Boxplots showing the composition of Mono/M ϕ clusters stratified by PLCG2+ SCLC subpopulation comprising > 0.75% of cancer cells within each tumor. Mono/M ϕ clusters are arranged along the X-axis in order from high to low score for IPF-associated Mono/M ϕ (shown in Fig. 7C). We show significant overrepresentation of Mono/M ϕ cluster 1 in samples harboring the recurrent PLCG2-high SCLC subpopulation. Error bars: 95% confidence interval.

(F) Bar plot showing robustness of covariate-adjusted Spearman correlation p-values using leave-one-sample-out cross-validation.

(G) Barplot showing Spearman's correlation of the profibrotic Mono/M ϕ gene signature to *PLCG2* expression, EMT, and T-cell exhaustion gene signatures, as well as *PLCG2* to T-cell exhaustion, in published bulk RNA-seq datasets from George, et al. 2015 and Rudin, et al. 2019 (N=81; **STAR Methods**).

(H) Color overlay of SCLC tumor MIBI 3 at FoV 1 (size 500 μ m x 500 μ m) showing the co-occurrence of the PLCG2-positive SCLC cells and the putative profibrotic Mono/M ϕ , similar to MIBI 12 at FoV 1 in Figure 7C. Left: Channels dsDNA (violet), Vimentin (white), CD8 (yellow), CD31 (orange), CD68 (red), CD163 (red), and FOXP3 (cyan) illustrate the global tumor environment structure. Middle: Channels PLCG2 (red), CD56 (yellow), and NEUROD1 (cyan) illustrate PLCG2+ tumor. Right: Channels CD14 (orange), CD16 (cyan), and CD81 (magenta) illustrate the putative profibrotic Mono/M ϕ .

(I) FoV from the (H) now visualized with segmented cancer cells using Mesmer, represented by dots colored by PLCG2+ SCLC cells vs profibrotic Mono/M ϕ .