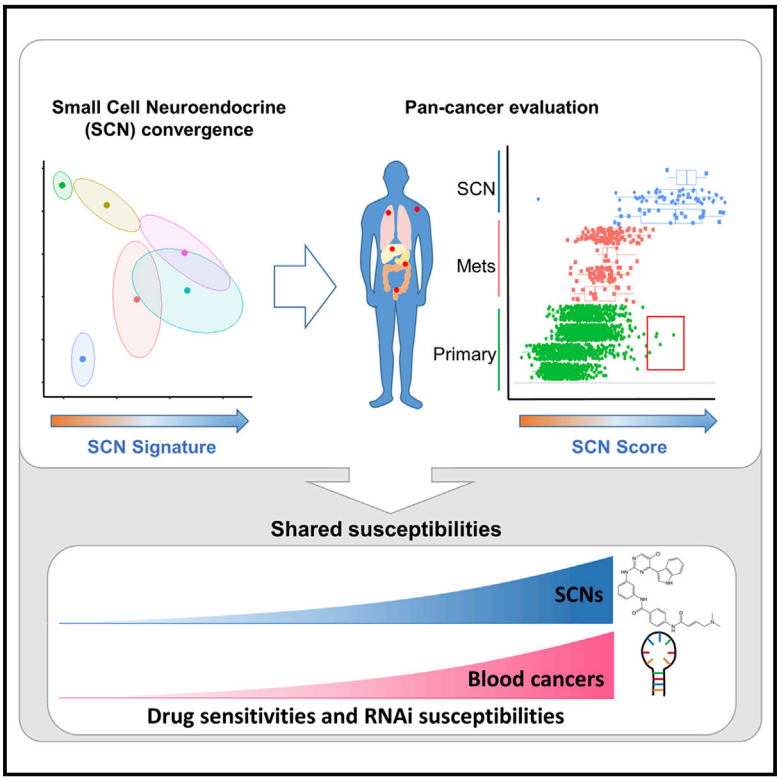
Article

Cancer Cell

# Pan-cancer Convergence to a Small-Cell Neuroendocrine Phenotype that Shares Susceptibilities with Hematological Malignancies

## Graphical Abstract




## Highlights

- Pan-tissue convergent molecular programs in small-cell neuroendocrine (SCN) cancers
- Poorer prognosis for tumors with high SCN signatures across epithelial cancers
- Drug sensitivity profiles are shared between SCN cancers and blood cancers
- SCN cancers and hematological malignancies have common functional dependencies

## Authors

Nikolas G. Balanis, Katherine M. Sheu, Favour N. Esedebe, ..., Jiaoti Huang, Owen N. Witte, Thomas G. Graeber

## Correspondence

owenwitte@mednet.ucla.edu (O.N.W.), tgraeber@mednet.ucla.edu (T.G.G.)

## In Brief

Balanis, Sheu, et al. identify a small-cell neuroendocrine (SCN) state across various epithelial cancer types that shares genome-wide expression, methylation, and copy-number alteration patterns and associates with poor prognosis. SCN cancers have common vulnerabilities that, unexpectedly, are shared with blood cancers.

Balanis et al., 2019, Cancer Cell *36*, 17–34
July 8, 2019 
https://doi.org/10.1016/j.ccell.2019.06.005

CellPress

Cancer Cell
Article

# Pan-cancer Convergence to a Small-Cell Neuroendocrine Phenotype that Shares Susceptibilities with Hematological Malignancies

Nikolas G. Balanis,[1,9] Katherine M. Sheu,[1,9] Favour N. Esedebe,[1] Saahil J. Patel,[1] Bryan A. Smith,[2] Jung Wook Park,[2] Salwan Alhani,[1] Brigitte N. Gomperts,[3,5,8] Jiaoti Huang,[6] Owen N. Witte,[2,3,8,*] and Thomas G. Graeber[1,3,4,7,8,10,*]
[1]Department of Molecular and Medical Pharmacology, UCLA, Los Angeles, CA 90095, USA
[2]Department of Microbiology, Immunology, and Molecular Genetics, UCLA, Los Angeles, CA 90095, USA
[3]Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, CA 90095, USA
[4]Crump Institute for Molecular Imaging, UCLA, Los Angeles, CA 90095, USA
[5]UCLA Children's Discovery and Innovation Institute, Mattel Children's Hospital UCLA, Department of Pediatrics, David Geffen School of Medicine, UCLA, Los Angeles, CA 90095, USA
[6]Department of Pathology, Duke University School of Medicine, Durham, NC 27708, USA
[7]California NanoSystems Institute, UCLA, Los Angeles, CA 90095, USA
[8]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, UCLA, Los Angeles, CA 90095, USA
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: owenwitte@mednet.ucla.edu (O.N.W.), tgraeber@mednet.ucla.edu (T.G.G.)
https://doi.org/10.1016/j.ccell.2019.06.005

## SUMMARY

**Small-cell neuroendocrine cancers (SCNCs) are an aggressive cancer subtype. Transdifferentiation toward an SCN phenotype has been reported as a resistance route in response to targeted therapies. Here, we identified a convergence to an SCN state that is widespread across epithelial cancers and is associated with poor prognosis. More broadly, non-SCN metastases have higher expression of SCN-associated transcription factors than non-SCN primary tumors. Drug sensitivity and gene dependency screens demonstrate that these convergent SCNCs have shared vulnerabilities. These common vulnerabilities are found across unannotated SCN-like epithelial cases, small-round-blue cell tumors, and unexpectedly in hematological malignancies. The SCN convergent phenotype and common sensitivity profiles with hematological cancers can guide treatment options beyond tissue-specific targeted therapies.**

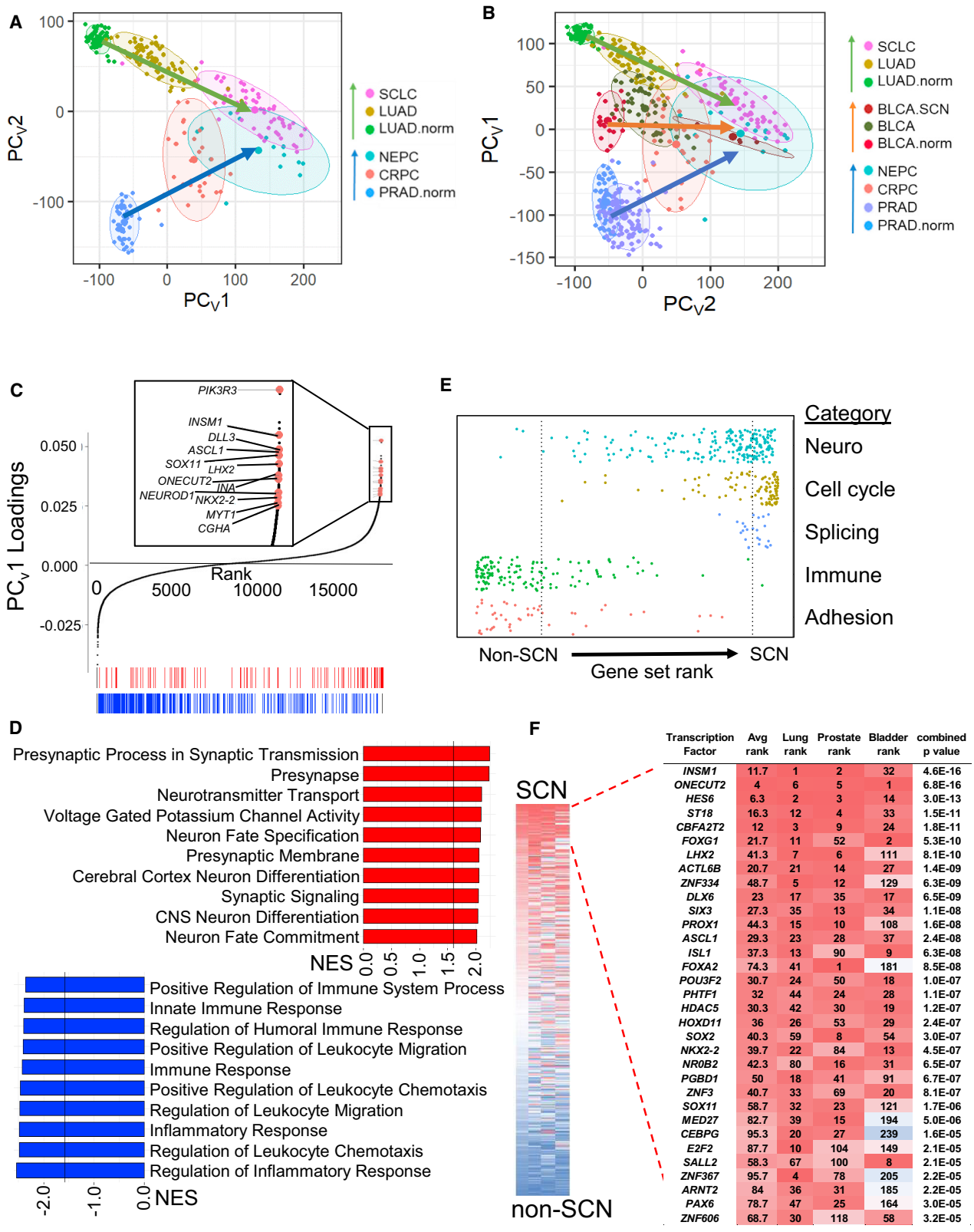
## INTRODUCTION

Small-cell neuroendocrine cancers (SCNCs) are highly aggressive and arise in multiple tissues, commonly reported in lung and therapy-refractive prostate cancers. SCNCs across tissues share morphology and marker-based histology such as high nuclear to cytoplasm ratios, frequent mitotic figures, and granular chromatin (Klimstra et al., 2015). At the molecular level, *TP53* and *RB1* loss and/or inactivating mutations are essentially obligatory for SCNCs of the lung and highly enriched in SCNCs of the prostate (Beltran et al., 2016; George et al., 2015). In addition, SCNCs express common neuroendocrine markers such as chromogranin A (*CHGA*) and synaptophysin (*SYP*) (Oberg et al., 2015).

The cell of origin of SCNCs across tissues is unclear. One proposed mechanism of origin is through transdifferentiation from a non-neuroendocrine cell lineage, which has been observed in patient tumors and in experimental models (Mu et al., 2017; Niederst et al., 2015; Yang et al., 2018). Transdifferentiation into small-cell lung carcinoma (SCLC) can be a mechanism for treatment resistance to epidermal growth factor receptor (EGFR) kinase inhibitors for lung adenocarcinomas

### Significance

SCNCs are aggressive and histologically similar across tissue types, and no effective treatment modalities are available. Lung and prostate adenocarcinomas can transdifferentiate to SCNCs in response to targeted therapy. This has important consequences in that SCNCs, once considered rare, may become increasingly common with the emergence of resistance cases from targeted therapies. Here we define molecular signatures for SCNCs, and we find that SCNCs share similar drug and RNAi vulnerabilities with blood cancers. Our results guide the detection of SCN-like cases in the clinic, and support the exploration of treatments for SCNCs that mimic treatments for blood cancers.




| Transcription Factor | Avg rank | Lung rank | Prostate rank | Bladder rank | combined p value |
|---|---|---|---|---|---|
| *INSM1* | 11.7 | 1 | 2 | 32 | 4.6E-16 |
| *ONECUT2* | 4 | 6 | 5 | 1 | 6.8E-16 |
| *HES6* | 6.3 | 2 | 3 | 14 | 3.0E-13 |
| *ST18* | 16.3 | 12 | 4 | 33 | 1.5E-11 |
| *CBFA2T2* | 12 | 3 | 9 | 24 | 1.8E-11 |
| *FOXG1* | 21.7 | 11 | 52 | 2 | 5.3E-10 |
| *LHX2* | 41.3 | 7 | 6 | 111 | 8.1E-10 |
| *ACTL6B* | 20.7 | 21 | 14 | 27 | 1.4E-09 |
| *ZNF334* | 48.7 | 5 | 12 | 129 | 6.3E-09 |
| *DLX6* | 23 | 17 | 35 | 17 | 6.5E-09 |
| *SIX3* | 27.3 | 35 | 13 | 34 | 1.1E-08 |
| *PROX1* | 44.3 | 15 | 10 | 108 | 1.6E-08 |
| *ASCL1* | 29.3 | 23 | 28 | 37 | 2.4E-08 |
| *ISL1* | 37.3 | 13 | 90 | 9 | 6.3E-08 |
| *FOXA2* | 74.3 | 41 | 1 | 181 | 8.5E-08 |
| *POU3F2* | 30.7 | 24 | 50 | 18 | 1.0E-07 |
| *PHTF1* | 32 | 44 | 24 | 28 | 1.1E-07 |
| *HDAC5* | 30.3 | 42 | 30 | 19 | 1.2E-07 |
| *HOXD11* | 36 | 26 | 53 | 29 | 2.4E-07 |
| *SOX2* | 40.3 | 59 | 8 | 54 | 3.0E-07 |
| *NKX2-2* | 39.7 | 22 | 84 | 13 | 4.5E-07 |
| *NR0B2* | 42.3 | 80 | 16 | 31 | 6.5E-07 |
| *PGBD1* | 50 | 18 | 41 | 91 | 6.7E-07 |
| *ZNF3* | 40.7 | 33 | 69 | 20 | 8.1E-07 |
| *SOX11* | 58.7 | 32 | 23 | 121 | 1.7E-06 |
| *MED27* | 82.7 | 39 | 15 | 194 | 5.0E-06 |
| *CEBPG* | 95.3 | 20 | 27 | 239 | 1.6E-05 |
| *E2F2* | 87.7 | 10 | 104 | 149 | 2.1E-05 |
| *SALL2* | 58.3 | 67 | 100 | 8 | 2.1E-05 |
| *ZNF367* | 95.7 | 4 | 78 | 205 | 2.2E-05 |
| *ARNT2* | 84 | 36 | 31 | 185 | 2.2E-05 |
| *PAX6* | 78.7 | 47 | 25 | 164 | 3.0E-05 |
| *ZNF606* | 68.7 | 30 | 118 | 58 | 3.2E-05 |

**Figure 1. Pan-cancer Convergence of SCN Carcinoma**

(A) Varimax-rotated PCA ($PCA_V$) of adjacent normal (norm), adenocarcinoma, and SCNC for lung and prostate patient tumors. Ellipses represent 80% confidence regions.

*(legend continued on next page)*

(LUAD) (Lee et al., 2017; Marcoux et al., 2018; Niederst et al., 2015). *De novo* neuroendocrine prostate cancers (NEPCs) are rare (approximately 1% of cases). As in lung, transdifferentiation from a prostate adenocarcinoma to an SCN state, in response to androgen suppression therapy, has been reported (Beltran et al., 2016), with 25%–40% of resulting resistant tumors expressing neuroendocrine markers (Alanee et al., 2015; Watson et al., 2015). Thus, across multiple tissues types, adenocarcinomas have been observed to escape targeted therapy through evolution to an SCN-like phenotype, highlighting the need to characterize and develop therapies for this aggressive cancer outcome.

There are currently no effective therapies for SCNCs, and SCNC typically leads to early and widespread metastases (Wong et al., 2009). Etoposide- and platinum-based chemotherapies are the primary first-line treatment modalities (Klimstra et al., 2015; Nadal et al., 2014), but they are only transiently effective, 5-year survival rates of patients with SCN lung or prostate cancers are less than 20% (Alanee et al., 2015). Here we sought to provide a molecular and functional underpinning for the observed pathology-based similarities among SCNCs arising in multiple tissue types.

## RESULTS

### SCNCs across Tissues Converge on a Common Molecular Phenotype

The advent of global molecular profiling of SCNCs (Beltran et al., 2016; George et al., 2015; Robertson et al., 2018) allows a molecular understanding of pan-tissue SCNC similarities. An unsupervised principal-component analysis (PCA) performed on RNA sequencing data of tumor biopsies of LUAD, castration-resistant prostate adenocarcinoma, SCLC, and castration-resistant NEPC, along with normal lung and prostate tissues, revealed a strongly convergent expression signature (Figure 1A; Table S1), as previously reported in comparisons of normal tissue and SCN tumors (Park et al., 2018). We further find that, as tumor cell states progress along a transdifferentiation trajectory from adenocarcinoma to SCNC (arrows), the tumors become increasingly independent of tissue of origin, with SCNCs of different tissues more similar to each other than adenocarcinomas of different tissues.

When included in the PCA analysis, four samples of SCN bladder cancer in the TCGA (Robertson et al., 2018) also support convergence of SCN cases, with the result reflecting a developmental landscape for the three tissues (Figure 1B). Hierarchical clustering supported that SCNCs of these three tissues were more similar to each other than to their non-SCNC counterparts (Figure S1A). Genes strongly contributing to the SCN signature include SCN-associated genes, such as *CHGA* and *INSM1*, and genes related to neural transcriptional programs, such as *ASCL1*, *NEUROD1*, *SEZ6*, *INA*, and *NKX2-2* (Figure 1C; Table S1). It is important to note that any one cancer incidence may contain only a subset of these markers (Figure S1B) and hence be missed by traditional classification schemes based on only a few markers (Oberg et al., 2015).

Enrichment analysis revealed the functional categories of SCN signature-associated genes. Enriched on the SCN side were gene sets related to neuron development and function, to splicing, and to cell cycle, whereas de-enriched gene sets included those related to adhesion and to immune response and inflammation (hereafter referred to as immune gene sets) (Figures 1D and 1E). To determine the contribution of proliferation genes on convergence, we removed proliferation and proliferation-associated genes, and found that the convergence of SCNCs across tissues was maintained (Figure S1C). Enrichment analysis on ranked genes from the proliferation gene-removed PCA confirmed the loss of enrichment of cell-cycle gene sets, and the maintenance of enrichment of neuronal gene sets (Figure S1D). Gene expression-based PCA defined on LUAD and SCLC cell lines in the Cancer Cell Line Encyclopedia (CCLE) faithfully predicted SCN cases of lung tumors (Figure S1E), but not other transitions such as melanoma dedifferentiation (Tsoi et al., 2018) (Figure S1F), supporting cell lines as an informative model for the interrogation of the SCN and non-SCN dichotomy. Interestingly, immune gene sets were also de-enriched in SCLC cell lines, indicating that the reduction of canonical immune and inflammation mediators in the SCN state is in part cancer cell-intrinsic (Figure S1G).

Inference of transcription factor activity from a data-guided transcription network across lung, prostate, and bladder SCN and non-SCN tumor datasets revealed a strong similarity across SCNCs in these tissues (Figure 1F). The transcription factor network identified included multiple factors central to neural development and brain patterning, such as LHX2, HES6, PROX1, PAX6, MYT1, and NKX2-2 (Table S1).

### The Shared SCN Gene Expression Signature Has an Epigenomic Basis

Signature overlap analysis using rank-rank hypergeometric overlap (RRHO) on partial least-squares regression (PLSR) loadings-based signatures, created individually from either lung or bladder cancers, showed that the DNA methylation signatures of the SCN transition were highly similar (Figure 2A) (Plaisier et al., 2010). To characterize the methylation sites distinguishing SCNCs from their adenocarcinoma counterparts,

(B) $PCA_v$ of samples in (A) and bladder patient tumor data. TCGA bladder cancer (BLCA) includes four SCN samples that were labeled separately (BLCA.SCN). PC1 and PC2 are reversed to show the SCN signature on the x axis as in (A).
(C) Gene loadings and selected top SCN-related genes of the varimax-rotated first principal component of the PCA ($PC_V1$) of (A). Gene set enrichment analysis (GSEA) was run on this ranked gene list. Shown along the bottom are the top gene sets in the "neuro" (red) and "immune" (blue) categories, also shown in (D).
(D) Top 10 gene sets from neuro and immune gene set categories. Lines mark false discovery rate (FDR) q value of 0.05.
(E) Distribution of gene sets from the C5 MSigDB collection ranked by normalized enrichment score. All listed categories enrichments are nominally significant ($p < 0.001$) by Kolmogorov-Smirnov test. Dashed lines mark FDR q value < 0.05 for individual gene sets in each direction.
(F) Average rank ordering of VIPER activities across the three tissue types (left) and zoom in to top of rank-based inferred activity in each cancer separately (right). Combined p value across the three tissues by Stouffer's method.
See also Figure S1 and Table S1.

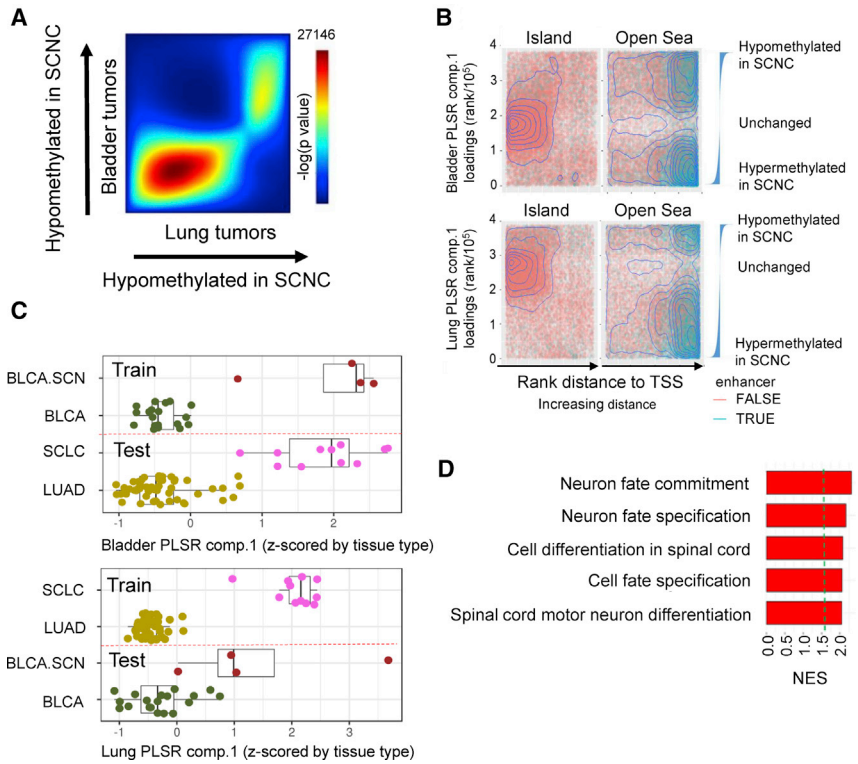



**Figure 2. An Epigenetic Basis for the Shared SCN Gene Expression Signature**

(A) RRHO heatmap of lung and bladder SCN signatures defined by PLSR loadings. Number shown is the maximum $-\log_{10}$(p value) of the RRHO heatmap.

(B) Enrichment analysis of methylation sites that define the lung or bladder SCN versus non-SCN dichotomy (random, representative sample of 100k sites shown for visibility). The x axis is the increasing (arrows) rank distance of the CG site from the nearest TSS. The y axis is the rank value of individually run lung or bladder PLSR component 1 loadings, thus extreme values represent sites with differential methylation between SCN and non-SCN tumors. Waterfall plots show the relative values of the ranked loadings.

(C) PLSR analysis of DNA methylation data from patient bladder SCN and non-SCN tumor samples, and projection of lung LUAD and SCLC tumor samples onto this framework, or vice versa. PLSR component 1 is *Z* scored by tissue type. Dashed red lines separate training data (above line) from testing data (below line). Lines inside boxplots represent the 25th, 50th, and 75th quantiles. Whiskers extend to 1.5 times the interquartile range.

(D) Top 5 enrichment analysis (GSEA) terms for genes hypomethylated in SCN (averaged lung, prostate, and bladder PLSR component 1 loadings). Dashed green line is at FDR q value = 0.05.

See also Figure S2.

we plotted methylation sites by their rank in the PLSR loadings. This analysis revealed that sites within open-sea regions (transcriptional start site [TSS] distal), rather than CG island regions (TSS proximal), are important in contribution to the SCN to non-SCN distinction (one-sided Kolmogorov-Smirnov [KS] test $p < 2.2 \times 10^{-16}$) (Figure 2B). Distinct from the canonical role of DNA methylation in regulating transcription at gene promoters, these distal methylation changes that delineate the epigenetic differences between SCNCs and non-SCNCs are consistent with enhancer-based regulation of gene expression programs (Sur and Taipale, 2016).

Methylation sites distinguishing SCN from non-SCN tumors in one tissue could separate the same groupings in another tissue. When we projected lung cancer samples to a PLSR methylation signature based on the bladder dichotomy, we observed that these sites on average distinguished SCLC from LUAD (Figure 2C). Training on lung samples likewise predicted bladder samples, further highlighting the concordance of non-SCN versus SCN methylation profiles across tissues (Figure 2C). Projection of gene-summarized methylation values of lung tumor samples onto a PLSR analysis of lung cell lines confirmed concordance of methylation patterns between tumors and cell lines (Figure S2A). Pairwise RRHO analyses supported the methylation-based concordance of SCN tumors from lung, bladder, and prostate tissues (Figure S2B). Gene-based summarization of methylation sites across the three tissues revealed enrichment of neuronal development gene sets (Figure 2D), supporting that the pan-tissue convergent similarity in SCNCs is functionally maintained across their methylomes.

## Selection for SCN-Associated DNA Copy-Number Alteration Signatures across Tissue Types

Recurrent genomic scars in cancer are reflective of selective forces that confer fitness advantages through copy-number alterations (CNA) in distinct genomic regions (Graham et al., 2017). PCA of CNA patterns in SCLC and LUAD cell lines showed SCN-specific differences in amplification and deletion patterns along the second principal component (Figure 3A; Table S2), with the first principal component describing the overall degree of aneuploidy (Figure 3B). Projection of tumor samples onto this PCA framework demonstrated that CNA patterns in SCLC cell lines were reflected in both SCLC (Figure 3C) and NEPC (Figure 3D) patient tumors. Consistently amplified or deleted regions in SCN cases of the two tissue types, analyzed together (Figures 3E and 3F) or independently (Figure 3G; Table S2), support that common selective forces act on SCN variants in both tissues. For example, SCNCs shared 1p amplification and 3p deletion (Figure 3F). As a positive control, we noted that *RB1* was included in the consistent deletion region on chromosome 13 (Figure 3G). Furthermore, these specific CNA changes are seen in both *de novo* (lung) and treatment-induced transdifferentiation (prostate) SCN cases.

## A Pan-cancer Predictor of SCNCs Reveals Unannotated Cases

Clinical classification of SCNCs relies on histology and pathology defined by a few sets of morphological and molecular features. To uncover cases missed by these traditional approaches we analyzed purported non-SCN tumors using the gene expression-defined PCA SCN convergence framework (Figure 1A).

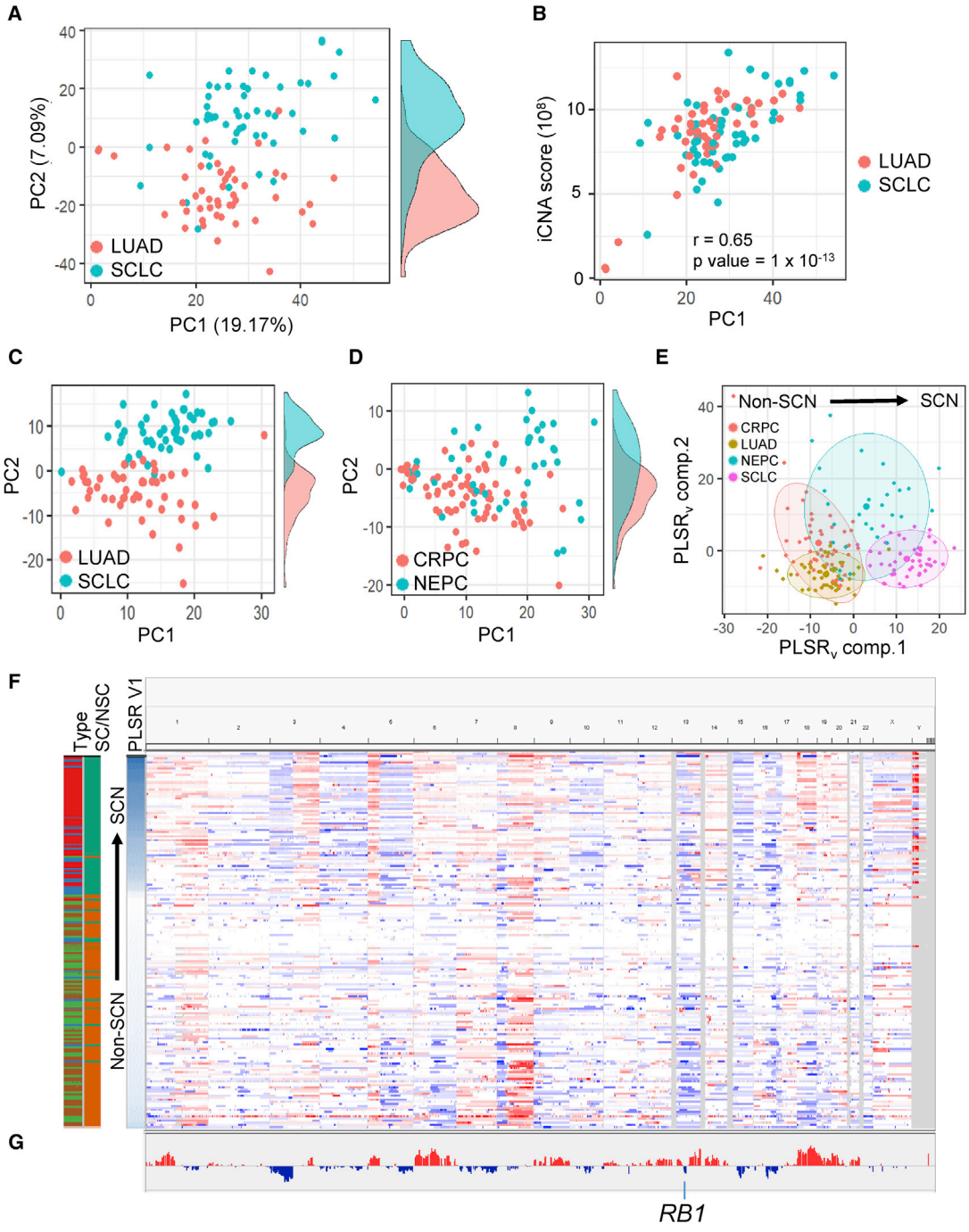



**Figure 3. SCNCs of Lung and Prostate Origin Share DNA Copy-Number Alteration Patterns**

(A) PCA on lung CCLE cell line CNA profiles. Side tracks are density plots of points along PC2.

(B) Cell line PC1 score reflects degree of aneuploidy (integrated CNA [iCNA] score).

(C) Projection of lung tumors onto the cell line PCA of (A). LUAD category randomly down-sampled for clarity. Cell lines are from the same batch.

(D) Projection of prostate tumors onto PCA of (A).

(E) PLSR of lung and prostate tumor CNA profiles, regressed on SCN or non-SCN status. LUAD category randomly down-sampled to match numbers in other categories.

(F) Genome-wide view of CNA patterns. Each row is a tumor sample: SCLC, red; LUAD, light green; NEPC, blue; castration-resistant prostate cancer (CRPC), brown; small cell (SC), green; non-SC (NSC), orange.

(G) Visualization of copy-number changes that are observed in both lung and prostate SCNC signatures. Each cancer type was analyzed by PLSR independently and then combined. The y axis represents the mean of the concordant PLSR loadings.

See also Table S2.

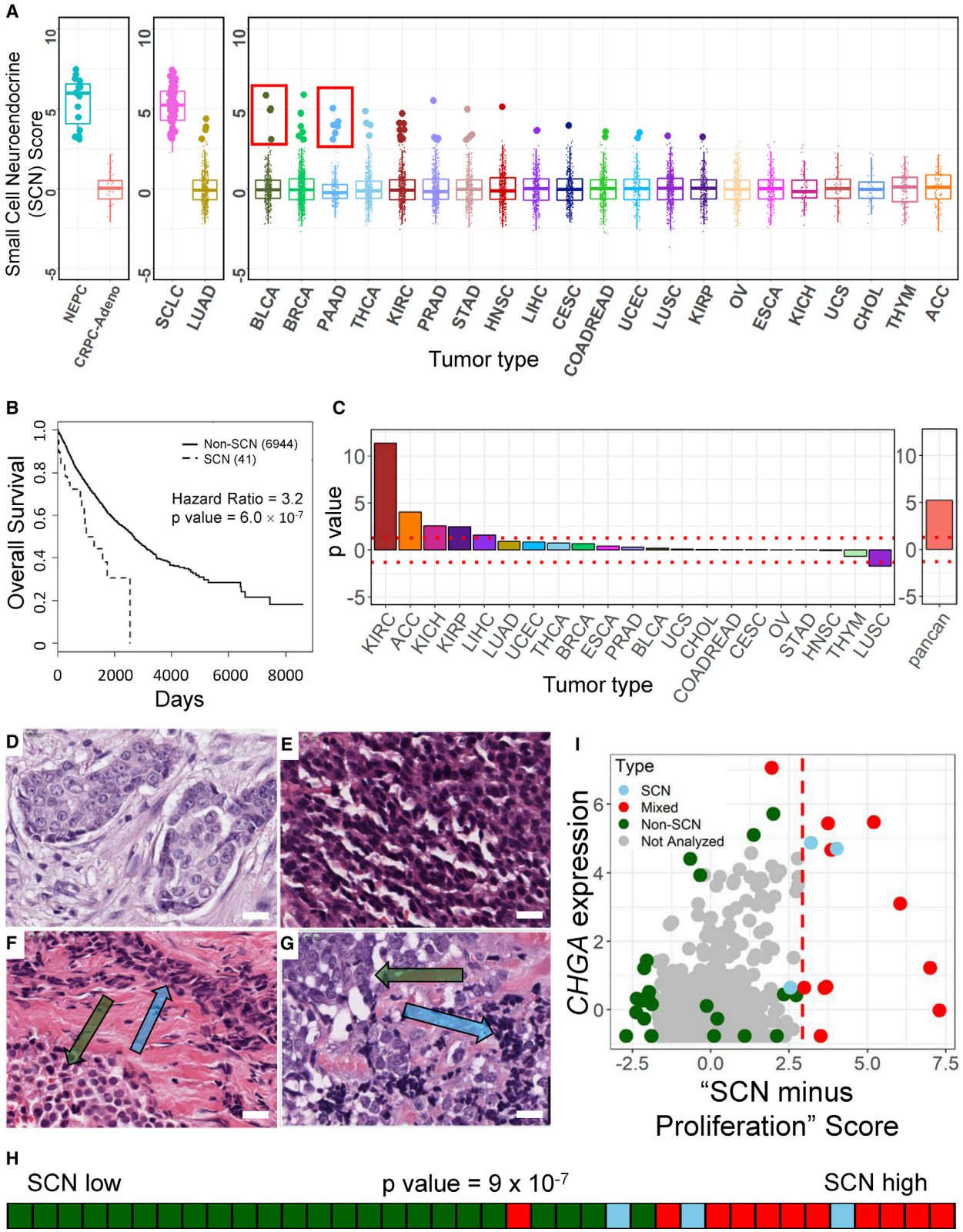



**Figure 4. Pan-cancer Identification of Primary Tumors with an SCN Signature**

(A) Gene expression-based prediction of SCN phenotype in epithelial TCGA patient tumors. Predictions made by projection onto $PC_v1$ from Figure 1A. For each cancer type, samples >3 SD from the mean are highlighted as enlarged data points. Red inset boxes indicate known cases of non-SCN PNETS in the TCGA PAAD

*(legend continued on next page)*

Through a projection approach, the signature weightings for each gene (Table S3) were used to predict SCN phenotypes across approximately 7,000 primary TCGA tumors from 21 different epithelial tumor types (Figure 4A). Boxplots of the first component score reveal a low frequency of primary tumor cases that contain an SCN component (~1%). Except for a subset of low-grade non-SCN pancreatic neuroendocrine tumors (PNETs), these samples were high-grade tumors (Table S3), consistent with reports in the literature for pathology-defined SCN cases (Oronsky et al., 2017; Watson et al., 2018). Of note, many other high-grade samples had low SCN scores. Thus, SCN versus non-SCN high-grade tumors may need to be addressed differently in the clinic.

We next compared the predicted SCN (referred to as SCN-like) with non-SCN cases, controlling for tumor type and excluding the indolent non-SCN PNET cases, and found a significant decrease in patient overall survival (Figure 4B; Table S3). Pan-cancer association of the SCN phenotype with poor overall survival was robust to changes in the SCN score threshold, and further supported in the majority of epithelial cancers by survival analysis based on continuous SCN score (Figure 4C; Table S3). Unlike the traditional classification scheme that creates a dichotomy (Oronsky et al., 2017), our model supports evaluating each cancer sample along a spectrum. Notably, even cancers with SCN scores in the range that are not overtly annotated as SCN by pathology analysis have poorer outcomes.

We next investigated the histological features of the tumors most strongly predicted to be SCN-like. A number of these cases received SCN scores >3 SD above the mean although almost none was diagnosed as SCNCs based on the original pathology reports. To confirm the predictions, we chose samples based on the proliferation-removed SCN score across multiple tissue types for pathologic examination (Gutman et al., 2013). Fourteen of 16 cases were confirmed by the analysis to have SCN or neuroendocrine features (Table S3; web resource).

Because SCN breast cancers are considered rare, with a reported incidence of 0.1% (Wang et al., 2014), we next investigated the presence of SCN histology features in a panel of predicted SCN-like BRCA cases. A pathologist thoroughly examined slides of TCGA breast tumor cases covering a full range of SCN scores (Figures 4D–4G and S3A–S3H; Table S3; web resource). The gene expression-based SCN score was indeed statistically predictive of BRCA samples with pathology-based SCN features (Figure 4H). Thus, the pan-tissue pathology analysis validated the transcriptome-based SCN signature score as a predictor of tumors with SCN morphology features, and supports an under-diagnosis of SCN cases.

In the pathology analysis, high SCN signature-score BRCA samples were typically called either SCN-positive, or more often as having mixed histology (Table S3). Three of four breast cancer subtypes (basal, luminal B, and luminal A) displayed regions with SCN pathology, usually in cases with accompanying genetic dysregulation of *TP53* and *RB1*, suggesting a subset of cases for which closer pathologic interrogation will be beneficial to uncover the often focal regions of SCN morphology (Figure S3I; Table S3). The SCN score-high BRCA samples did not uniformly express the traditional SCN markers of *CHGA*, *SYP*, and *NCAM1* (Figure 4I; Table S3). This finding reinforces the appreciation that heterogeneity in expression limits the use of only a small set of markers in the clinical identification of aggressive SCN signature-positive tumor cases.

REST/NRSF is a transcriptional repressor that negatively regulates neural gene expression. Loss of REST transcriptional repression has been linked to promoting the SCN phenotype in NEPC, and REST activation inhibits neuroendocrine differentiation in SCLC (Lim et al., 2017; Zhang et al., 2015). Our analysis supports that REST activity regulates the SCN phenotype in a pan-cancer manner (Figures S3J–S3L). The expression of REST repressional target genes is positively correlated with SCN score across many epithelial and other tumor types, consistent with loss of REST-mediated repression (Figure S3K). The expression of *REST* itself is somewhat positively correlated with SCN score in the majority of cancers types, downweighting REST's potential as a direct marker of non-NE tissues in samples with mixed histology. Overall the SCN score and pathology analyses above validate that tumors with SCN-like molecular signatures can be missed by standard pathology examination, and thus may be present more commonly than currently appreciated in cancer types in which SCNCs are considered rare.

## Mutations Associated with the Pan-tissue SCN Phenotype

Mutations and loss of *TP53* and *RB1* are associated with lung and prostate SCNCs (Beltran et al., 2016; George et al., 2015), and contribute to transdifferentiation and SCN-like histology *in vivo* (Ku et al., 2017). We next sought to identify additional

cohort (6 of 8; 2 missed are just subthreshold) and SCNCs in BLCA (3 of 4, 1 missed is the next sample subthreshold). Cancer types in rightmost box are left-right sorted based on the average of top 3 scores per cancer type. Lines inside boxplots represent the 25th, 50th, and 75th quantiles. Whiskers extend to 1.5 times the interquartile range.

(B) Kaplan-Meier overall survival analysis for predicted SCN versus non-SCN cases from TCGA epithelial cancers (samples in A right box plus LUAD; PAAD, SCLC, NEPC, and castration-resistant prostate adenocarcinoma [CRPC-Adeno] not included). The p value is calculated controlling for tumor type.

(C) The p values from a Cox regression survival analysis in individual cancer types using the continuous SCN score (left), and pan-cancer Cox regression using the continuous score, accounting for cancer type (right).

(D–G) TCGA BRCA H&E-stained diagnostic slides of invasive ductal carcinoma (D; TCGA-D8-A1XD), small-cell neuroendocrine carcinoma (E) (TCGA-BH-A0HL), mixed tumor with components of invasive ductal carcinoma (lower left, green arrow) and small-cell neuroendocrine carcinoma (upper right, blue arrow) (F) (TCGA-E9-A245), mixed tumor with components of large cell neuroendocrine carcinoma (upper left, green arrow) and small-cell neuroendocrine carcinoma (lower right, blue arrow) (G) (TCGA-A1-A0SK). Scale bars, 20 μm.

(H) Rug plot and Kolmogorov-Smirnov enrichment p value of breast cases scored by pathologist for SCN features ordered by their proliferation-removed SCN score.

(I) Scatterplot of all samples in TCGA BRCA cohort (x axis, proliferation-removed SCN score; y axis: *Z* scored chromogranin A expression. Cases to the right of the dashed red line, x = 3, were computationally predicted as SCN-like).

See also Figures S3 and S4 and Table S3.

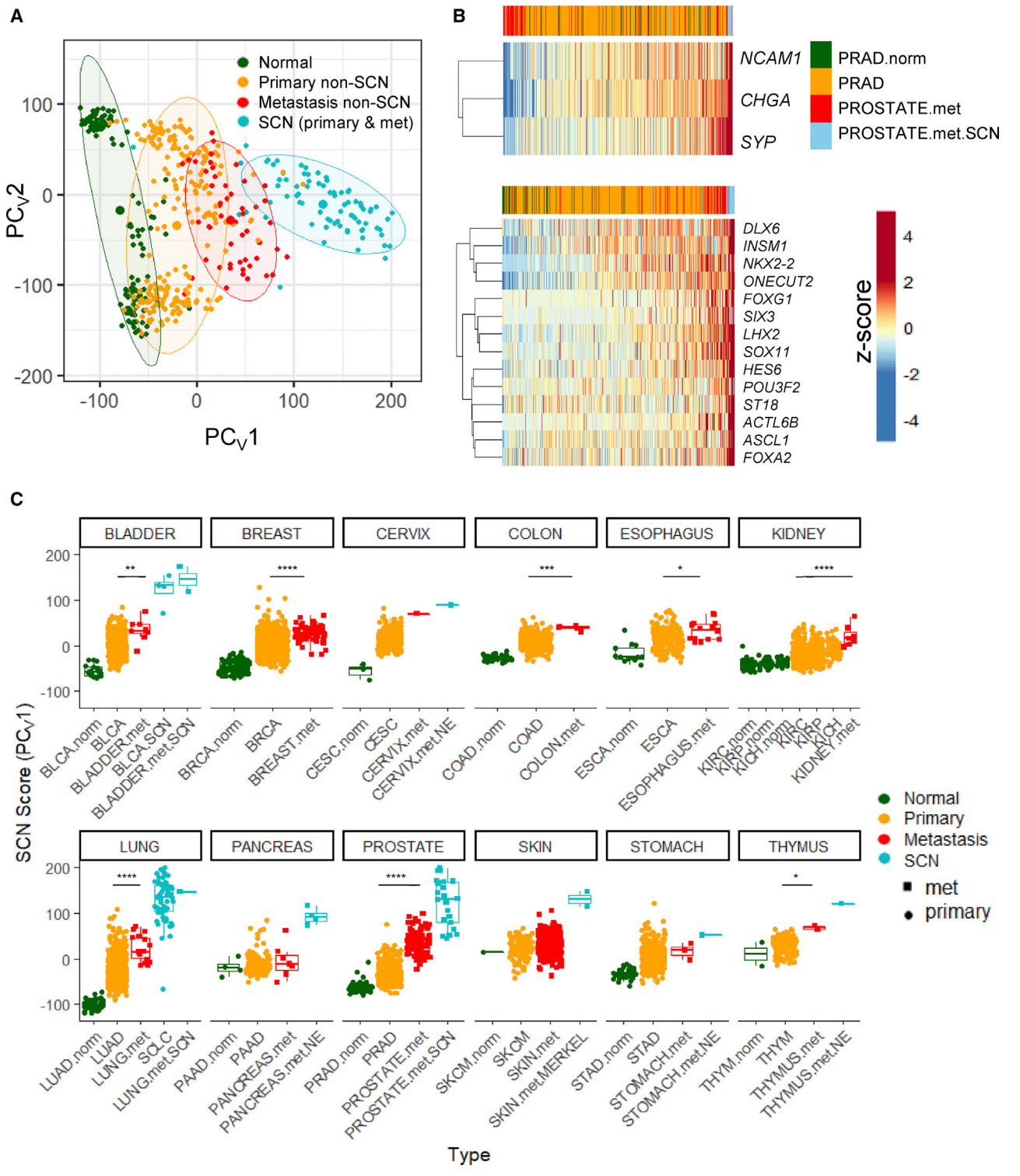



**Figure 5. Metastases across Multiple Tissue Types Have Increased Expression of SCN Features**

(A) Projection of lung, prostate, bladder normal tissue, primary, and metastatic non-SCN tumor samples, and SCN tumor samples onto the PCA framework of Figure 1A. Centroids and 80% confidence regions for the indicated groups are displayed. Here, the pancreatic, cervix, stomach, and thymus blue samples are those annotated with a neuroendocrine (NE)-related term in Robinson et al. (2017).

(B) Z-score-based heatmap of the expression of canonical SCN marker genes (top) and key SCN transcription factor genes (bottom) for prostate normal, primary adeno, metastatic adeno, and metastatic SCN. Samples ordered from left to right by the sum *Z* score of the genes displayed, as indicated at the top of the heatmaps with sample type as indicated by the upper key.

*(legend continued on next page)*

mutations associated with SCN features in a pan-tissue context by directly leveraging the spectrum of SCN signature scores across primary tumors (Figure 4A). First using known cases of SCNC in lung and prostate tissues, we confirmed significant associations of *TP53* and *RB1* mutations with high SCN score, as well as a significant association of *FOXA1* mutation, whose wild-type expression inhibits transition to NEPC (Kim et al., 2017) (Figure S4A; Table S3). In other epithelial cancer types, in addition to *TP53* and *RB1* mutations, we uncovered SCN score-associated mutations in *NRAS* (neuroblastoma-RAS) and genes such as *OBSCN* and *BCLAF1*, which have been previously associated with tissue-specific SCNCs (Figure S4B; Table S3) (Cho et al., 2016; Rudin et al., 2012). In contrast, *KRAS* mutations were enriched in cases on the non-SCN side of the spectrum. Even in cancer types that are already highly neuronal, such as glioblastoma, *TP53* mutations were associated with higher SCN score (Figure S4C; Table S3). These results point to mutations beyond *TP53* and *RB1* that are common to SCN-like cancers across tissue types, and supports their mutational contribution to the development of SCN phenotypes.

### Metastatic Non-SCN Tumors Express the SCN Signature Profile More Strongly Than Primary Non-SCN Tumors

Given that SCNCs can arise from epithelial tissues, we investigated the extent of the SCN signature in metastatic adenocarcinomas (MAd) in comparison with primary adenocarcinomas (PAd). For lung, prostate, and bladder tissues, we found that the expression profiles of MAd samples were more similar to SCNCs than were their respective PAds (Figure 5A). MAd samples typically do not express canonical SCN markers, but express increasing levels of other SCN-associated genes and transcription factors (Figure 5B). To determine the SCN signature score of MAd and SCNC samples on a pan-tissue scale, cases from the MET500 dataset (Robinson et al., 2017) were projected onto the SCN framework of Figure 1A (Figure S5A). Plotting the expression levels of the top 50 genes of the SCN signature in prostate cancers visually demonstrated the increased similarity of prostate MAds to prostate SCNCs (Figure S5B). Metastatic non-SCN samples tended to have SCN score distributions significantly shifted upward on the SCN spectrum in relation to their respective primary non-SCN samples, in multiple different tissues (Figure 5C).

Because the SCN score contains both neuronal and proliferation components (Figure 1E), we sought to deconvolute these components in the primary, metastatic, and SCN samples. Although a proliferation signal is a contributing component, metastatic cases have a significant increase in the expression of the SCN program even when the proliferation component is removed (Figure S5C). Furthermore, this signature deconvolution separates low-grade PNET cases, which have a low proliferation score (red box, Figure S5C), from pancreatic SCN cases, consistent with reports that these cancers are distinct (Yachida et al., 2012). Additional cases of low-grade PNETs, as well as neuroendocrine tumors of the rectum and small intestine (Alvarez et al., 2018), both fell in the same pattern of having low proliferation but high neuroendocrine features (Figure S5D). Taken together, these results support that MAds derive elements of their aggressive phenotypes from both a proliferation program and from a program associated with the neuronal programs of SCNCs.

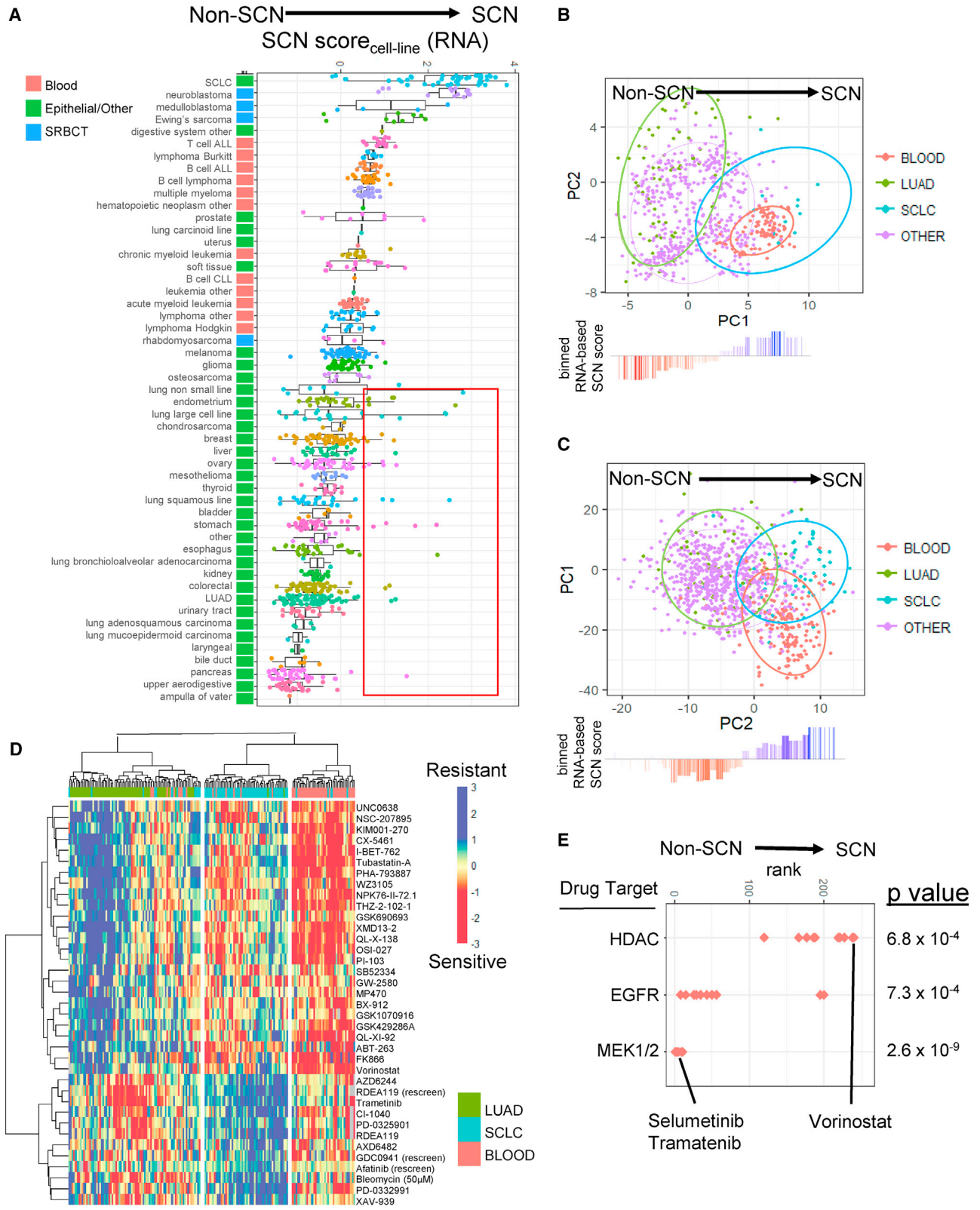
### Hematopoietic Cancers Share Expression Profiles and Drug Sensitivities with SCNCs

We next leveraged the concordance of SCN tumors and cell lines (Figures S1, S2, and 3) and the availability of drug screening data to gain insight into potential therapeutic vulnerabilities of SCNCs. Training on LUAD and SCLC cell lines, we scored all remaining CCLE cell lines using a gene expression- and PLSR-based SCN score (Figure 6A; Table S4). Consistent with the tumor findings, (1) known SCN cell lines had higher SCN scores than most epithelial lines, and (2) the epithelial cases included a few cell lines that were not annotated as small cell but nonetheless had a strong SCN score (red box in Figure 6A). Unexpectedly, hematopoietic cancer cell lines had higher SCN expression signatures in comparison with the non-blood epithelial cancers (Figure 6A), which was likewise observed when the same analysis was performed on tumor data (Figure S6A; Table S4).

We next analyzed protein expression signatures using reverse-phase protein array measurements across cell lines from various tissues (Li et al., 2017). PCA well segregated SCLC and LUAD cell lines (Figure 6B). Further supporting the unanticipated similarity in SCNCs and blood cancers, blood cancers as a group had projection-based protein profiles highly similar to SCLC (Figure 6B; Table S4), with increased expression of proteins such as BCL2, and regulators of cell cycle such as ATM, CHK1, and E2F1, which are candidate therapeutic targets in SCLC (Doerr et al., 2017) (Figure S6B; Table S4).

The similarity between SCNCs and blood cancers was seen again in large-scale drug sensitivity data (Iorio et al., 2016). Analogous to the expression-based analysis, PCA was performed on the matrix of all drug sensitivity (half maximal inhibitory concentration [$IC_{50}$]) values across LUAD and SCLC cell lines. This unsupervised approach revealed a sensitivity/resistance profile for SCLC cases (Figure 6C). Parallel to the protein-based analysis, blood cancers projected into the SCLC drug sensitivity space, while epithelial tissue cancer types projected into the LUAD space (Figure 6C). Hierarchical clustering based on the drugs with top differential sensitivity between LUAD and SCLC lines also supported the drug sensitivity similarities between SCLC and blood cancers (Figure 6D). Compared with LUAD lines, SCLC lines were more sensitive to drugs that inhibit histone deacetylation such as Vorinostat (Figure 6E), and more resistant to drugs that target EGFR or components of extracellular signal-regulated kinase (ERK)/mitogen-activated protein kinase (MAPK) signaling pathways, such as Trametinib and Selumatinib (Figure 6E). Taken together, these findings support that hematopoietic cancers have similarities to SCNCs that range from expression profiles to drug sensitivity-based phenotype profiles.

(C) Pan-cancer projections onto framework of Figure 1A. Data are from TCGA normal samples, TCGA primary tumors, MET500 metastatic tumors, SCLC tumors (George et al., 2015), and CRPC and NEPC tumors (Beltran et al., 2016). Plotted are the $PC_V1$ values, representing SCN score. The SKIN.met cohort has two sources, the TCGA and MET500 databases. Wilcoxon-Mann-Whitney p values are shown comparing primary (orange) and metastatic (red) cases. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. Lines inside boxplots represent the 25th, 50th, and 75th quantiles. Whiskers extend to 1.5 times the interquartile range. See also Figure S5.




*(legend on next page)*

### Small-Round-Blue Cell Tumors Share Drug Sensitivities with Lung SCN and Blood Cancers

In addition to blood cancers, our gene signature analysis revealed that small-round-blue cell tumor (SRBCT) cell lines have high SCN gene expression signature scores (Figure 6A). These similarities between SRBCTs and SCNCs were also observed in protein profile (Figure S7A) and drug sensitivity profile analysis (Figure S7B).

Drugs with high differential sensitivity both in SCLC versus LUAD and in blood versus non-blood comparisons potentially act through a shared SCN-blood mechanism. The top drugs with such a pan-SCN-blood sensitivity profile included FK866 and THZ-2-102-1, to which SRBCTs also displayed sensitivity (Figure S7C; Table S4). FK866 is one of several nicotinamide phosphoribosyltransferase (NAMPT) inhibitors shown to have indications for efficacy in both SCLC and neuronal cancers (Cole et al., 2017; Watson et al., 2009). THZ-2-102-1, a cyclin-dependent kinase 7 (CDK7) inhibitor, targets components of transcriptional regulation and has been shown to be highly effective in *MYCN*-amplified neuroblastoma (Chipumuro et al., 2014).

Of note, blood cancers and the majority of, but not all, SCNCs grow in suspension *in vitro*. Nevertheless, SCLC suspension and SCLC adherent lines are intermingled in their PCA-based drug sensitivity profiles, and both are distinct from LUAD (Figure S7D). The parallel drug sensitivities of SCLC and blood cancers (R = 0.68) was maintained when the analysis was restricted to the 31% of SCLC lines in the drug sensitivity database that grow adherent in culture (R = 0.69; n = 20 adherent, n = 6 semi-adherent, n = 38 suspension) (Figure S7D). Furthermore, SRBCTs largely grow adherent in culture (n = 50 adherent, n = 6 semi-adherent, n = 6 suspension). Taken together, these results support that the commonalities in drug sensitivities among blood, SCN, and SRBCT cancers are not solely due to suspension culture characteristics (Figure S7D).

### An Expression-Based SCN Classifier Is Predictive of Sensitivity to SCN-Targeting Drugs in Non-SCN Epithelial Cancers across Tissues

Comparing gene expression and drug profiles in cell lines we found that increased SCN expression score generally correlated with increased sensitivity to SCN-targeting drugs (Figure S7E). This correlation had an overall combined p value of $1.4 \times 10^{-5}$, and reached statistical significance in four individual epithelial tissue types. In addition, tissues with a greater mean SCN expression signature also had drug sensitivity profiles that were closer to that of SCLC (Figure S7F). Thus, as epithelial cancers from various tissues develop SCN gene expression programs, they become increasingly susceptible to a similar panel of SCN-targeting drugs. These results support that the small-cell phenotype exists along an expression signature-defined spectrum that has influence on therapeutic vulnerabilities in individual cancers.

### Validation of Shared SCN and Blood Susceptibilities Based on Gene Dependencies

To validate the finding that SCN and blood cancers shared drug susceptibility, we analyzed a genome-scale RNAi (short hairpin RNA-based) functional screen across a large panel of cell lines (Tsherniak et al., 2017). We first trained a prediction model on the gene dependency data between LUAD and SCLC cell lines (Figure 7A). A PLSR-based dependency prediction approach had 92% cross-validation accuracy for predicting the identity of lung cell lines that have concordant annotation and gene expression signatures (see the STAR Methods). Applying the prediction model to the remaining cell types showed that blood and SRBCT lines share gene dependencies with the SCLC lines (Figure 7B). In reverse, using a blood versus non-blood RNAi sensitivity framework as the predictor confirmed that SCNCs and SRBCTs have more blood-like dependencies (Figures S8A and S8B). Thus, the RNAi data validate our drug-based findings that SCNCs, hematopoietic cancers, and SRBCTs have shared susceptibilities.

Investigating biological pathways revealed that SCN and blood cancer cell lines shared susceptibility to disruption of immune pathways and lipid and sterol metabolism (Figures 7C, 7D, and S8C). Notably, this susceptibility enrichment was not exclusively tied to elevated (or depressed) gene expression. In fact, SCN cell lines have lower expression of immune genes but remain sensitive to their disruption. In contrast, SCN and blood lines share increased expression of cell-cycle genes, but both had decreased sensitivity to the knockdown of this gene category (Figure 7E). Although SCNCs had high expression of neural genes (Figure 1), and a few distinct neural gene sets did show genetic sensitivity to knockdown (Table S5; Figure 7F, upper graph), there was no substantial enrichment of susceptibility to knockdown of neural genes—supporting that many components of the upregulated neuronal gene programs promote phenotypes distinct from the regulation of cell survival (Figure 7F). These expression versus sensitivity results highlight that cells are not particularly dependent on genes with elevated expression. Furthermore, even the knockdown of genes with reduced expression can have functional consequences.

While our transcriptome, proteome, drug sensitivity, and gene susceptibility analyses above uncovered many similarities between SCNCs and blood cancers (Figures S8C and S8D),

**Figure 6. Blood Cancers Have SCN-like Gene and Protein Expression Profiles and Drug Sensitivities**

(A) Cell line gene expression SCN score (component 1 score of projection on PLSR of LUAD and SCLC lines). *Z* score adjusted across all cell line samples together. Cancer types top-bottom sorted based on average SCN score per type. Red box, non-SCN annotated epithelial lines with strong SCN score. Lines inside boxplots represent the 25th, 50th, and 75th quantiles. Whiskers extend to 1.5 times the interquartile range.

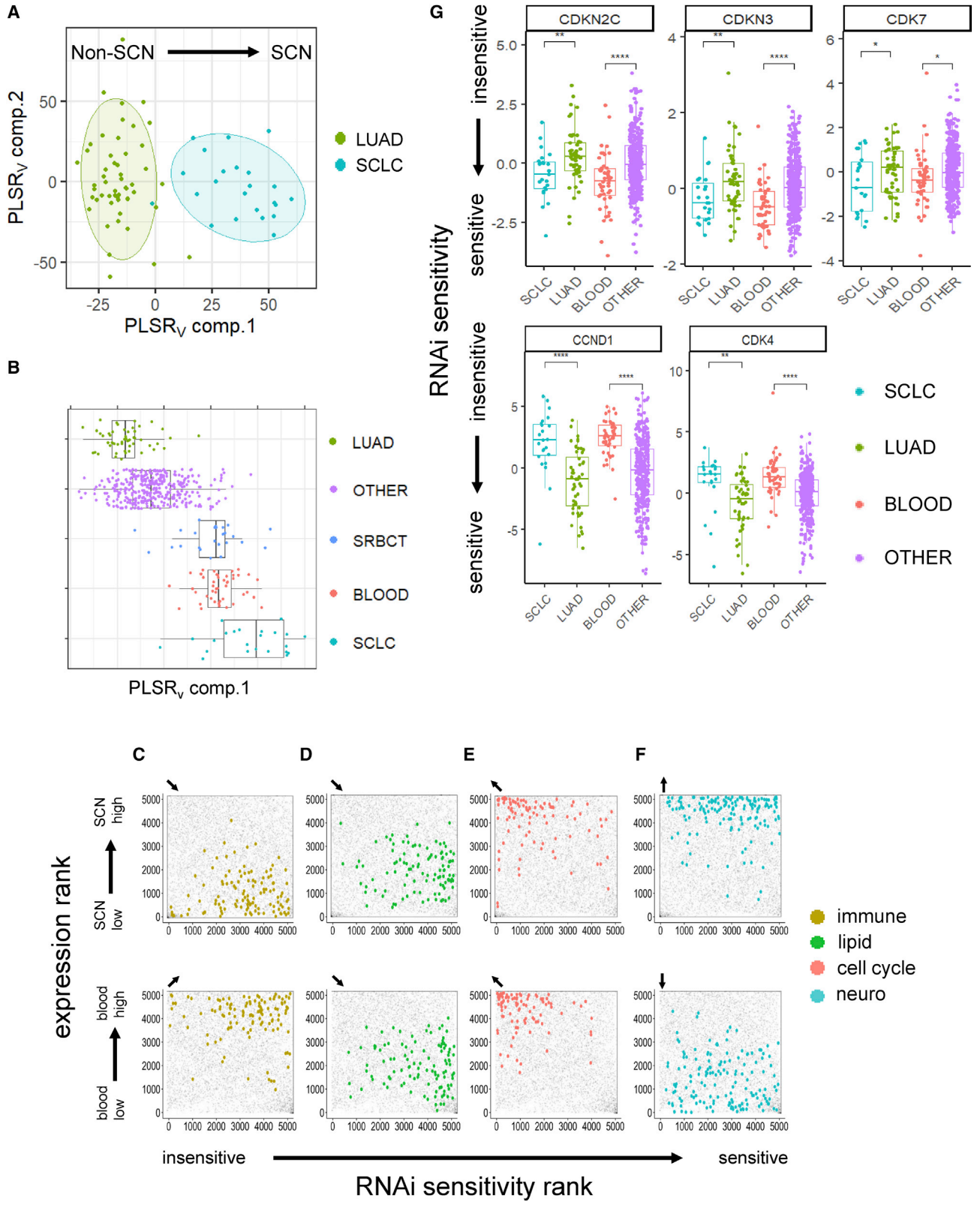
(B) Protein profile of blood cell lines, from projection onto PCA of LUAD and SCLC lines.

(C) Drug sensitivity profile of blood cell lines, from projection onto PCA of $IC_{50}$ values for LUAD and SCLC lines. Ellipses represent 80% confidence regions. In (B) and (C), waterfall plots show binned gene expression SCN score of projected cell lines.

(D) Heatmap of drugs with differential sensitivity between LUAD and SCLC cell lines (t test $p < 0.0001$). Hierarchical clustering is done on $IC_{50}$ measurements for LUAD, SCLC, and blood lines.

(E) Enrichment of drug targets for drugs more effective in LUAD or SCLC lines. Kolmogorov-Smirnov test p values.

See also Figures S6 and S7 and Table S4.




*(legend on next page)*

naturally there were also many differences, especially at the gene expression level, with blood cancers expressing more immune-annotated genes, and SCNCs expressing more neuro-annotated genes (Figure S8E).

We further investigated dependencies on protein targets of interest currently in clinical trials for various cancer types. We focused on CDKs and CDK antagonists, which had strong differential sensitivities in the LUAD versus SCLC, and blood versus non-blood comparisons. CCND1 and CDK4 knockdown were more effective in LUAD compared with either SCLC or blood cancers, while knockdown of CDKN2C, a CDK4 inhibitor, was more effective in SCLC and blood cancers (Figure 7G). The CCND1-CDK4 complex has been shown to inhibit phospho-RB to promote cell-cycle progression, and loss of RB function is a hallmark in the development of the SCN phenotype (Figures 3G, S4A, and S4B). These results predict that epithelial cancers could use de-differentiation to an SCN phenotype as a resistance mechanism to CDK4 inhibition. This finding is consistent with RB loss as a resistance mechanism to CDK4 antagonism observed in preclinical models of liver cancer and glioblastoma, and in patients with metastatic breast cancer (Bollard et al., 2017; Condorelli et al., 2018). Notably, CDK7 knockdown was more effective in SCLC and blood cancers, paralleling the drug sensitivity finding for THZ-2-102-1, a CDK7 inhibitor (Figures 7G and S7C). Thus, CDK7 sensitivity provides a specific example of a previously documented sensitivity in both SCNCs and blood cancers (Cayrol et al., 2017; Christensen et al., 2014), which is a part of the wider panel of shared sensitivities revealed by our pan-cancer and functional screen analysis.

### Concordant Gene Expression-Based Drug Sensitivity Profiles in SCN and Blood Tumors

We next investigated if the gene expression profiles associated with drug susceptibilities are present in primary tumors. We built an elastic net (ENET)-based predictor of drug sensitivities in tumors trained on the LUAD and SCLC cell line gene expression profiles. In general, drugs that were differentially potent in LUAD versus SCLC cell lines were likewise predicted to be differentially potent in LUAD versus SCLC tumors (Figures 8A and 8B). This result included a drug that targets BCL2 (ABT-263), which is often overexpressed in SCLC as a resistance mechanism to etoposide treatment (van Meerbeeck et al., 2011); and the PLK inhibitor NPK76-II-72-1 (Figure 8C, red dots Figure 7A). To determine the global sensitivity status of individual tumors, ENET-predicted tumor sensitivities were projected onto the PCA drug sensitivities of LUAD and SCLC cell lines (as shown in Figure 8C). We found that epithelial tumors with previously unannotated SCN features (based on gene expression profiles, Figure 4A) typically had drug sensitivity profiles more similar to annotated SCLC tumors than to LUAD tumors (Figure 8D). Ten of 11 tumor types had a significant association of expression-based predicted SCN status to predicted SCN-like sensitivity (KS test-based; Figure 8D; Table S6). These results support that the expression profiles associated with drug sensitivities *in vitro* are reflective of the *in vivo* tumor setting.

As in the cell line empirical-sensitivity data case, patient blood cancers shared predicted-drug sensitivity profiles with SCN tumors. These shared profiles included the BCL2 inhibitor ABT-263, which is in a class of drugs approved to treat chronic lymphoblastic leukemia, and is in development for other hematological malignancies (Hantel et al., 2018; Seymour et al., 2018) (Figures 8E and 8F). Thus, we observe a similarity between SCNCs and blood cancers that spans gene and protein expression, and drug sensitivity and gene vulnerability profiles.

## DISCUSSION

The highly convergent molecular profiles of SCN tumors reported here match well-documented pathology findings of related morphologies and biomarkers across different tissue types (Klimstra et al., 2015; Rickman et al., 2017). The strength of these shared profiles allowed us to define a molecular classifier that identified cases with SCN or neuroendocrine features across numerous purportedly non-SCN primary tumors. These cases may not have been reported in the original pathology because their SCN component is typically focal and largely unexpected in epithelial tumors. Survival analysis supported the aggressiveness of these SCN-like epithelial cancers (Chen et al., 2018). As such, screening for and targeting the SCN-like phenotype early in individual patient cases could have clinical benefit.

The implications of identifying this spectrum of SCNCs is further apparent in functional data on drug response and gene dependency. We found three categories of tumors to generally share expression and sensitivity features with pathology-defined SCN cases: SCN-like epithelial cancers, SRBCTs, and hematopoietic malignancies. Epithelial tumor types in which SCN tumors rarely occur, but have poor prognosis (Inno et al., 2016), present statistical challenges in advancing care through clinical trials, and thus cross-tissue learning supported by shared molecular profiles will likely be required. Our findings thus support and guide efforts to identify rare SCNC cases and develop therapies that will target SCNCs from multiple tissue sites of origin.

**Figure 7. Validation of Shared Vulnerabilities Based on Genome-Scale Functional RNAi Screens**

(A) Varimax-rotated PLSR model trained on the genome-scale RNAi sensitivity values for LUAD and SCLC cell lines. Ellipses represent 80% confidence regions.

(B) Prediction of RNAi sensitivity profile for blood, SRBCTs, and all other cell lines in the dataset. SCN sensitivity score based on projection to the varimax PLSR component 1 from (A) (which included LUAD and SCLC).

(C–F) Comparison of gene set expression rank with gene set sensitivity rank for cell line SCLC versus LUAD (top) and blood versus non-blood (bottom) for gene sets containing selected keywords. Gene set RRHO scatterplots are subcategorized and colored by immune (C), lipid (D), cell cycle (E), and neuro (F) gene sets, with all other gene sets colored gray. Arrows in top left corner of individual panels indicate direction of significance ($q < 0.01$) by Kolmogorov-Smirnov test (diagonal arrows indicate significance in both expression and sensitivity directions; Benjamani-Hochberg correction).

(G) Select targets with differential SCLC versus LUAD and blood versus non-blood RNAi sensitivity. The y axis (RNAi sensitivity) is the published Demeter score. Student's t test p values. *$p < 0.1$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. Lines inside boxplots represent the 25th, 50th, and 75th quantiles. Whiskers extend to 1.5 times the interquartile range.

See also Figure S8 and Table S5.

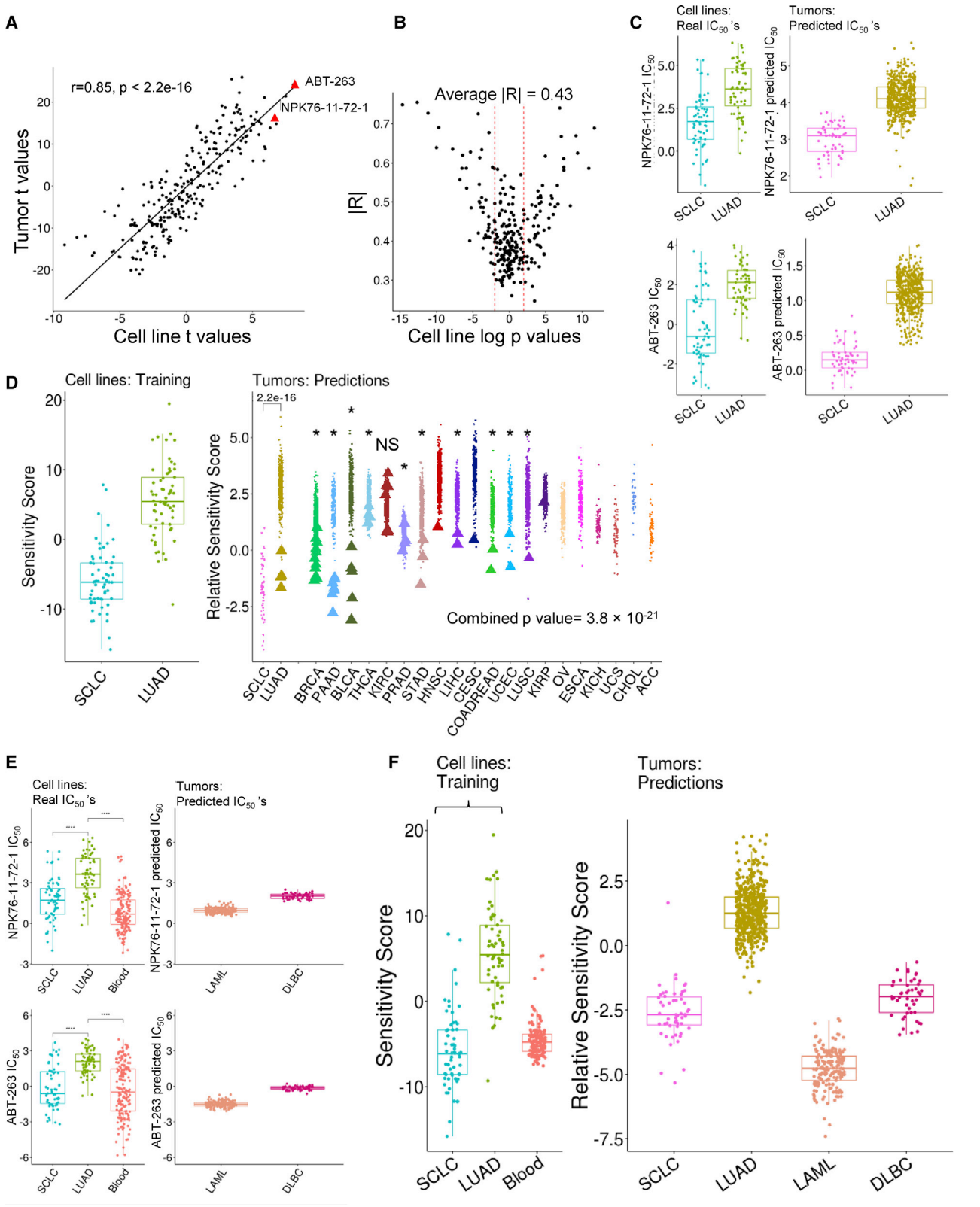



**Figure 8. Tumors Recapitulate SCN and Blood Cell Line Sensitivity Signatures**

(A) Scatterplot of t values from t test of true $IC_{50}$ values of SCLC versus LUAD in cell lines (x axis) and predicted $IC_{50}$ values in tumors (y axis).

(B) Comparison of cross-validation model $R^2$ values (cell line-based) with t test signed log p values of cell line SCLC versus LUAD true drug sensitivities ($IC_{50}$). $R^2$ values are from true cell line drug sensitivities versus cross-validation-based predicted cell line sensitivities. Dashed lines are p value 0.05 (log p value = 1.3).

*(legend continued on next page)*

The finding that small molecules preferentially effective in SCNs are also effective in hematopoietic malignancies was initially unexpected. However, there are several lines of phenotypic and molecular evidence in support of the sensitivity profile similarity. SCNCs such as those of the lung and prostate, as well as blood cancers, are among the most highly undifferentiated cancers, and are characterized by upregulation of stem-like signaling pathways (Malta et al., 2018; Smith et al., 2018). In culture, SCNCs display some phenotypic characteristics of hematopoietic cancer cell lines, such as a propensity to grow in suspension. In patients, SCN tumors downregulate genes involved in adhesion function compared with non-SCN tumors, corresponding with the widespread nature of SCN metastases (Wong et al., 2009). However, both adherent SCN and suspension SCN cell lines display similar drug sensitivity profiles to blood cancers, supporting that their drug sensitivity profiles are not defined solely by their suspension characteristic. SCNCs and lymphomas also share histological similarities. Namely, the small-round-blue cell phenotype seen in both tumor types can make distinguishing them a challenge despite their notably different origin, often requiring staining for more specific molecular markers (De Las Casas et al., 2004). Importantly, case reports have warned of the misdiagnosis of hematopoietic cancers as SRBCTs owing to both morphological and histological similarities. For example, CD99, a marker used in the diagnosis of Ewing sarcoma, is frequently highly expressed in hematopoietic malignancies (Lucas et al., 2001; Ozdemirli et al., 1998).

The commonalities between SCNCs and blood cancers in functionally defined susceptibilities provide guidance and avenues for hematopoietic malignancy therapies to inform SCN therapies, and vice versa. Importantly, the sensitivity profiles demonstrate predicted concordance in tumor data. Our analysis of the functional screens and previous studies point to the disruption of epigenetic regulators through use of histone deacetylase (HDAC) inhibitors (Platta et al., 2007) or to the targeting of antiapoptotic factors such as BCL2 (Lochmann et al., 2018) as approaches to target both blood and SCNCs. CDK7 inhibitors have been shown to be effective in SCLCs, MYCN-driven SRBCTs, and blood cancers (Cayrol et al., 2017; Chipumuro et al., 2014; Christensen et al., 2014)—matching the cancer types that we co-classified in our sensitivity profiling analysis. Future work thus should investigate whether other therapeutic approaches successful in one cancer type can be applied to the cancer types reported here to have parallel sensitivity profiles. For example, CDK7 or HDAC inhibitors in combination with BCL2 inhibition are reported leads for treating lymphomas and leukemias, respectively (Cyrenne et al., 2017), and thus warrant testing in SCNCs.

In analysis of gene dependencies, we found that cell-cycle-related CDKs were more effective in non-SCN and non-blood cancers. In contrast, knockdown or inhibition of CDK7, part of the CDK family sub-group involved in regulating transcription (Lapenna and Giordano, 2009), was more effective in SCNCs and blood cancers. Disruption of CDK7 has been shown to interfere with RNA polymerase II pausing, allowing concomitant decrease of super-enhancer-driven strong transcription of sets of oncogenes (Nilson et al., 2015). This concept of de-amplifying the effect of multiple oncogenic transcription factors offers an approach to target cancers of distinct cell types (e.g., SCNCs, SRBCTs, or blood cancers) that may upregulate different transcription factors, but through the same molecular mechanism, as previously postulated (Nilson et al., 2015). Our analysis also uncovered similarities for further therapeutic investigation. For example, both SCN and blood cancers share elevated sensitivities to targeting lipid pathways, which are generally known to regulate cancer cell signaling and metastasis (Pascual et al., 2017).

The cell of origin for SCNCs is not universally defined. SCNCs have been alternatively reported to arise from primary normal neuroendocrine cell precursors in some cases and transdifferentiation from adenocarcinoma in others (Watson et al., 2015). The cell of origin may vary depending on factors such as whether the SCNCs arise *de novo* or as a resistance mechanism to therapeutics (Feng et al., 2017; Rickman et al., 2017). In either case, the transdifferentiation-linked path to an SCN cancer converges, and thus leads to molecular profiles and phenotypes increasingly independent of the tissue of origin. This model of convergence is supported by our demonstration that normal prostate and lung cells can be reprogrammed into prostate and lung SCNCs, respectively, using a common set of small-cell cancer-associated genetic perturbations (Park et al., 2018). Despite using distinct cell types as the starting material, the resulting tumor models had highly similar molecular profiles that matched the overlapping profiles of patient prostate and lung SCNCs. Combination therapies that both treat non-SCNCs and prevent escape toward SCNCs may help curtail the incidence of the disease. The pan-tissue scope of shared SCN susceptibilities supports the potential for pan-cancer therapy development to impact a large number of otherwise aggressive, metastatic, and treatment-resistant cancers.

## Web Resource: SCN Signatures

The web resource at https://systems.crump.ucla.edu/scn/ can be used to (1) predict samples that have SCN features based on RNA sequencing data by projection to the varimax-rotated PCA in Figure 1A; (2) predict relative drug sensitivities of samples that may have SCN features using the ENET model trained on SCLC and LUAD cell line data (Figure 8); and (3) view high-power

(C) Real and predicted sensitivities for NPK76-II-72-1 and ABT-263. The y axis is the log $IC_{50}$.

(D) Projection of RNA sequencing, ENET-based predicted sensitivities of epithelial tumors onto PCA framework of SCLC and LUAD cell lines. More negative "Relative Sensitivity Scores" correspond to more SCN-like drug sensitivity profiles. Asterisks (*) denote individual tumor type significance by Kolmogorov-Smirnov test. NS, not significant. SCLC and LUAD tumor predicted sensitivity compared via a Wilcoxon-Mann-Whitney test ($p < 2.2 \times 10^{-16}$). Combined p value calculated with Stouffer's test.

(E) Real and predicted sensitivities of NPK76-II-72-1 and ABT-263 for blood tumors. The y axis is the log $IC_{50}$ values.

(F) Projection of ENET-based sensitivities of epithelial tumors and blood tumors onto PCA framework of SCLC and LUAD cell lines (open bracket); combined p value from Stouffer's test. More negative relative sensitivity scores correspond to more SCN-like drug sensitivity profiles. Lines inside boxplots represent the 25th, 50th, and 75th quantiles. Whiskers extend to 1.5 times the interquartile range.

See also Table S6.

pathology images of SCN regions of the tumors investigated in Figure 4 and Table S3.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
  - Abbreviations
- METHOD DETAILS
  - Data Acquisition and Processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - PCA/PLSR
  - Small Cell Neuroendocrine (SCN) and Proliferation Scores
  - Transcription Factor Analysis
  - GSEA-Squared
  - REST Analysis
  - Rank Rank Hypergeometric Overlap (RRHO)
  - Methylation Analysis
  - CNA Analysis
  - Pathology Analysis
  - Survival Analysis
  - Mutation Analysis
  - RPPA Analysis
  - Drug Sensitivity Analysis
  - Gene Dependency
  - Tumor Drug Sensitivity Prediction
- DATA AND SOFTWARE AVAILABILITY
  - Public Data Resources

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.ccell.2019.06.005.

## ACKNOWLEDGMENTS

We thank members of T.G.G. and O.N.W. labs for helpful discussion. We thank Weihong Yan for assistance in creating the web resource and pathology image webpages. T.G.G. and O.N.W. are supported by the NIH NCI R01 CA222877, the UCLA SPORE in Prostate Cancer (NIH NCI P50 CA092131), the W.M. Keck Foundation, and the UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research Hal Gaba Director's Fund for Cancer Stem Cell Research. B.N.G. and T.G.G. are supported by NIH NCI R01 CA208303, UC Tobacco-Related Disease Research Program (TRDRP; 26IP-0036). J.H. is supported by NIH NCI R01 CA205001. K.M.S. is supported by the UCLA Medical Scientist Training Program (NIH NIGMS T32 GM008042) and the Systems and Integrative Biology Training Grant at UCLA (NIH T32 GM008185). B.A.S. is supported by a Prostate Cancer Foundation Young Investigator Award. Biostatistics supported by the National Center for Advancing Translational Sciences UCLA CTSI Grant UL1TR000124. J.W.P. is supported by UCLA Broad Stem Cell Research Center postdoctoral fellowship and an NIH NCI K99/R00 Pathway to Independence Award K99 CA218731.

## AUTHOR CONTRIBUTIONS

Conceptualization, N.G.B., K.M.S., B.A.S., J.W.P., B.N.G., O.N.W., and T.G.G.; Methodology, N.G.B., K.M.S., and T.G.G.; Formal Analysis, N.G.B., K.M.S., F.N.E., and S.J.P.; Investigation, N.G.B., K.M.S., and J.H. Resources/Data Curation, N.G.B., K.M.S., and S.A.; Writing – Original Draft, N.G.B., K.M.S., and T.G.G.; Writing – Review & Editing, N.G.B., K.M.S., O.N.W., and T.G.G.; Funding Acquisition, B.N.G., O.N.W., and T.G.G. Supervision and Project Administration, T.G.G.

## DECLARATION OF INTERESTS

N.G.B., K.M.S., F.N.E, J.W.P., O.N.W., and T.G.G. are inventors on a patent application related to this work, filed by UCLA. The lab of T.G.G. has completed a research agreement with ImmunoActiva. O.N.W. currently has consulting, equity, and/or board relationships with Trethera Corporation, Kronos Biosciences, Sofie Biosciences, and Allogene Therapeutics. None of these companies contributed to or directed any of the research reported in this article.

Received: November 5, 2018
Revised: March 4, 2019
Accepted: June 6, 2019
Published: July 8, 2019

## REFERENCES

Alanee, S., Moore, A., Nutt, M., Holland, B., Dynda, D., El-Zawahry, A., and McVary, K.T. (2015). Contemporary incidence and mortality rates of neuroendocrine prostate cancer. Anticancer Res. *35*, 4145–4150.

Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat. Genet. *48*, 838–847.

Alvarez, M.J., Subramaniam, P.S., Tang, L.H., Grunn, A., Aburi, M., Rieckhof, G., Komissarova, E.V., Hagan, E.A., Bodei, L., Clemons, P.A., et al. (2018). A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. Nat. Genet. *50*, 979–989.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

Beltran, H., Prandi, D., Mosquera, J.M., Benelli, M., Puca, L., Cyrta, J., Marotz, C., Giannopoulou, E., Chakravarthi, B.V.S.K., Varambally, S., et al. (2016). Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat. Med. *22*, 298–305.

Bollard, J., Miguela, V., Ruiz de Galarreta, M., Venkatesh, A., Bian, C.B., Roberto, M.P., Tovar, V., Sia, D., Molina-Sánchez, P., Nguyen, C.B., et al. (2017). Palbociclib (PD-0332991), a selective CDK4/6 inhibitor, restricts tumour growth in preclinical models of hepatocellular carcinoma. Gut *66*, 1286–1296.

Carlson, M. (2019). org.Hs.eg.db: genome wide annotation for human. R Package Version 3.8.2.

Cayrol, F., Praditsuktavorn, P., Fernando, T.M., Kwiatkowski, N., Marullo, R., Calvo-Vidal, M.N., Phillip, J., Pera, B., Yang, S.N., Takpradit, K., et al. (2017). THZ1 targeting CDK7 suppresses STAT transcriptional activity and sensitizes T-cell lymphomas to BCL2 inhibitors. Nat. Commun. *8*, 14290.

Chen, G.-L., and Miller, G.M. (2018). Alternative REST splicing underappreciated. ENeuro *5*, https://doi.org/10.1523/ENEURO.0034-18.2018.

Chen, F., Zhang, Y., Gibbons, D.L., Deneen, B., Kwiatkowski, D.J., Ittmann, M., and Creighton, C.J. (2018). Pan-cancer molecular classes transcending tumor lineage across 32 cancer types, multiple data platforms, and over 10,000 cases. Clin. Cancer Res. *24*, 2182–2193.

Chipumuro, E., Marco, E., Christensen, C.L., Kwiatkowski, N., Zhang, T., Hatheway, C.M., Abraham, B.J., Sharma, B., Yeung, C., Altabef, A., et al. (2014). CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. Cell *159*, 1126–1139.

Cho, S.Y., Choi, M., Ban, H.-J., Lee, C.H., Park, S., Kim, H., Kim, Y.-S., Lee, Y.S., and Lee, J.-Y. (2016). Cervical small cell neuroendocrine tumor mutation profiles via whole exome sequencing. Oncotarget *8*, 8095–8104.

Christensen, C.L., Kwiatkowski, N., Abraham, B.J., Carretero, J., Al-Shahrour, F., Zhang, T., Chipumuro, E., Herter-Sprie, G.S., Akbay, E.A., Altabef, A., et al. (2014). Targeting transcriptional addictions in small cell lung cancer with a covalent CDK7 inhibitor. Cancer Cell *26*, 909–922.

Clinical Lung Cancer Genome Project (CLCGP), and Network Genomic Medicine (NGM). (2013). A genomics-based classification of human lung tumors. Sci. Transl. Med. *5*, 209ra153.

Cole, J., Guiot, M.-C., Gravel, M., Bernier, C., Shore, G.C., and Roulston, A. (2017). Novel NAPRT specific antibody identifies small cell lung cancer and neuronal cancers as promising clinical indications for a NAMPT inhibitor/niacin co-administration strategy. Oncotarget *8*, 77846–77859.

Condorelli, R., Spring, L., O'Shaughnessy, J., Lacroix, L., Bailleux, C., Scott, V., Dubois, J., Nagy, R.J., Lanman, R.B., Iafrate, A.J., et al. (2018). Polyclonal RB1 mutations and acquired resistance to CDK 4/6 inhibitors in patients with metastatic breast cancer. Ann. Oncol. *29*, 640–645.

Cyrenne, B.M., Lewis, J.M., Weed, J.G., Carlson, K.R., Mirza, F.N., Foss, F.M., and Girardi, M. (2017). Synergy of BCL2 and histone deacetylase inhibition against leukemic cells from cutaneous T-cell lymphoma patients. Blood *130*, 2073–2083.

Doerr, F., George, J., Schmitt, A., Beleggia, F., Rehkämper, T., Hermann, S., Walter, V., Weber, J.-P., Thomas, R.K., Wittersheim, M., et al. (2017). Targeting a non-oncogene addiction to the ATR/CHK1 axis for the treatment of small cell lung cancer. Sci. Rep. *7*, 15511.

Feng, N., Yin, Y., He, Y., and Huang, J. (2017). Alternative splicing provides a novel molecular mechanism for prostatic small-cell neuroendocrine carcinoma. Eur. Urol. *71*, 79–80.

George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., et al. (2015). Comprehensive genomic profiles of small cell lung cancer. Nature *524*, 47–53.

Graham, N.A., Minasyan, A., Lomova, A., Cass, A., Balanis, N.G., Friedman, M., Chan, S., Zhao, S., Delgado, A., Go, J., et al. (2017). Recurrent patterns of DNA copy number alterations in tumors reflect metabolic selection pressures. Mol. Syst. Biol. *13*, 914.

Gutman, D.A., Cobb, J., Somanna, D., Park, Y., Wang, F., Kurc, T., Saltz, J.H., Brat, D.J., Cooper, L.A.D., and Kong, J. (2013). Cancer Digital Slide Archive: an informatics resource to support integrated in silico analysis of TCGA pathology data. J. Am. Med. Inform. Assoc. *20*, 1091–1098.

Hantel, A., Wynne, J., Lacayo, N., Khaw, S.L., Rubnitz, J., Mullighan, C., Schmidt, M., Zhou, Y., Ross, J.A., Rosenwinkel, L., et al. (2018). Safety and efficacy of the BCL inhibitors venetoclax and navitoclax in combination with chemotherapy in patients with relapsed/refractory acute lymphoblastic leukemia and lymphoblastic lymphoma. Clin. Lymphoma Myeloma Leuk. *18*, S184–S185.

Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell *173*, 291–304.e6.

Inno, A., Bogina, G., Turazza, M., Bortesi, L., Duranti, S., Massocco, A., Zamboni, G., Carbognin, G., Alongi, F., Salgarello, M., et al. (2016). Neuroendocrine carcinoma of the breast: current evidence and future perspectives. Oncologist *21*, 28–32.

Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). A landscape of pharmacogenomic interactions in cancer. Cell *166*, 740–754.

Kim, J., Jin, H., Zhao, J.C., Yang, Y.A., Li, Y., Yang, X., Dong, X., and Yu, J. (2017). FOXA1 inhibits prostate cancer neuroendocrine differentiation. Oncogene *36*, 4072–4080.

Klimstra, D.S., Beltran, H., Lilenbaum, R., and Bergsland, E. (2015). The spectrum of neuroendocrine tumors: histologic classification, unique features and areas of overlap. Am. Soc. Clin. Oncol. Educ. Book *35*, 92–103.

Ku, S.Y., Rosario, S., Wang, Y., Mu, P., Seshadri, M., Goodrich, Z.W., Goodrich, M.M., Labbé, D.P., Gomez, E.C., Wang, J., et al. (2017). Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. Science *355*, 78–83.

Kuhn, M., Wing, J, Weston S, Williams A, Keefer, C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, The R Core Team, et al. (2018). caret: Classification and Regression Training. R package version 6.0-81. https://CRAN.R-project.org/package=caret.

Lachmann, A., Giorgi, F.M., Lopez, G., and Califano, A. (2016). ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. Bioinformatics *32*, 2233–2235.

Lapenna, S., and Giordano, A. (2009). Cell cycle kinases as therapeutic targets for cancer. Nat. Rev. Drug Discov. *8*, 547–566.

De Las Casas, L.E., Gokden, M., Mukunyadzi, P., White, P., Baker, S.J., Hermonat, P.L., You, H., Korourian, S., Malak, S.F., and Miranda, R.N. (2004). A morphologic and statistical comparative study of small-cell carcinoma and non-Hodgkin's lymphoma in fine-needle aspiration biopsy material from lymph nodes. Diagn. Cytopathol. *31*, 229–234.

Lee, J.-K., Lee, J., Kim, S., Kim, S., Youk, J., Park, S., An, Y., Keam, B., Kim, D.-W., Heo, D.S., et al. (2017). Clonal history and genetic predictors of transformation into small-cell carcinomas from lung adenocarcinomas. J. Clin. Oncol. *35*, 3065–3074.

Li, J., Zhao, W., Akbani, R., Liu, W., Ju, Z., Ling, S., Vellano, C.P., Roebuck, P., Yu, Q., Eterovic, A.K., et al. (2017). Characterization of human cancer cell lines by reverse-phase protein arrays. Cancer Cell *31*, 225–239.

Lim, J.S., Ibaseta, A., Fischer, M.M., Cancilla, B., O'Young, G., Cristea, S., Luca, V.C., Yang, D., Jahchan, N.S., Hamard, C., et al. (2017). Intratumoural heterogeneity generated by Notch signalling promotes small-cell lung cancer. Nature *545*, 360–364.

Lochmann, T.L., Floros, K.V., Naseri, M., Powell, K.M., Cook, W., March, R.J., Stein, G.T., Greninger, P., Maves, Y.K., Saunders, L.R., et al. (2018). Venetoclax is effective in small-cell lung cancers with high BCL-2 expression. Clin. Cancer Res. *24*, 360–369.

Lucas, D.R., Bentley, G., Dan, M.E., Tabaczka, P., Poulik, J.M., and Mott, M.P. (2001). Ewing sarcoma vs lymphoblastic lymphoma. A comparative immunohistochemical study. Am. J. Clin. Pathol. *115*, 11–17.

Mahamdallie, S.S., Hanks, S., Karlin, K.L., Zachariou, A., Perdeaux, E.R., Ruark, E., Shaw, C.A., Renwick, A., Ramsay, E., Yost, S., et al. (2015). Mutations in the transcriptional repressor REST predispose to Wilms tumor. Nat. Genet. *47*, 1471–1474.

Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell *173*, 338–354.e15.

Marcoux, N., Gettinger, S.N., O'Kane, G., Arbour, K.C., Neal, J.W., Husain, H., Evans, T.L., Brahmer, J.R., Muzikansky, A., Bonomi, P.D., et al. (2018). *EGFR*-mutant adenocarcinomas that transform to small-cell lung cancer and other neuroendocrine carcinomas: clinical outcomes. J. Clin. Oncol. *37*, 278–285.

Mayrhofer, M., Viklund, B., and Isaksson, A. (2016). Rawcopy: improved copy number analysis with Affymetrix arrays. Sci. Rep. *6*, 36158.

van Meerbeeck, J.P., Fennell, D.A., and De Ruysscher, D.K.M. (2011). Small-cell lung cancer. Lancet *378*, 1741–1755.

Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41.

Mohammad, H.P., Smitheman, K.N., Kamat, C.D., Soong, D., Federowicz, K.E., Van Aller, G.S., Schneck, J.L., Carson, J.D., Liu, Y., Butticello, M., et al. (2015). A DNA hypomethylation signature predicts antitumor activity of LSD1 inhibitors in SCLC. Cancer Cell *28*, 57–69.

Mu, P., Zhang, Z., Benelli, M., Karthaus, W.R., Hoover, E., Chen, C.-C., Wongvipat, J., Ku, S.-Y., Gao, D., Cao, Z., et al. (2017). *SOX2* promotes lineage plasticity and antiandrogen resistance in *TP53*- and *RB1*-deficient prostate cancer. Science *355*, 84–88.

Nadal, R., Schweizer, M., Kryvenko, O.N., Epstein, J.I., and Eisenberger, M.A. (2014). Small cell carcinoma of the prostate. Nat. Rev. Urol. *11*, 213–219.

Niederst, M.J., Sequist, L.V., Poirier, J.T., Mermel, C.H., Lockerman, E.L., Garcia, A.R., Katayama, R., Costa, C., Ross, K.N., Moran, T., et al. (2015). RB loss in resistant EGFR mutant lung adenocarcinomas that transform to small-cell lung cancer. Nat. Commun. *6*, 6377.

Nilson, K.A., Guo, J., Turek, M.E., Brogie, J.E., Delaney, E., Luse, D.S., and Price, D.H. (2015). THZ1 reveals roles for Cdk7 in co-transcriptional capping and pausing. Mol. Cell *59*, 576–587.

Oberg, K., Modlin, I.M., De Herder, W., Pavel, M., Klimstra, D., Frilling, A., Metz, D.C., Heaney, A., Kwekkeboom, D., Strosberg, J., et al. (2015). Consensus on biomarkers for neuroendocrine tumour disease. Lancet Oncol. *16*, e435–e446.

Oronsky, B., Ma, P.C., Morgensztern, D., and Carter, C.A. (2017). Nothing but NET: a review of neuroendocrine tumors and carcinomas. Neoplasia *19*, 991–1002.

Ozdemirli, M., Fanburg-Smith, J.C., Hartmann, D.P., Shad, A.T., Lage, J.M., Magrath, I.T., Azumi, N., Harris, N.L., Cossman, J., and Jaffe, E.S. (1998). Precursor B-lymphoblastic lymphoma presenting as a solitary bone tumor and mimicking Ewing's sarcoma: a report of four cases and review of the literature. Am. J. Surg. Pathol. *22*, 795–804.

Park, J.W., Lee, J.K., Sheu, K.M., Wang, L., Balanis, N.G., Nguyen, K., Smith, B.A., Cheng, C., Tsai, B.L., Cheng, D., et al. (2018). Reprogramming normal human epithelial tissues to a common, lethal neuroendocrine cancer lineage. Science *362*, 91–95.

Pascual, G., Avgustinova, A., Mejetta, S., Martín, M., Castellanos, A., Attolini, C.S.-O., Berenguer, A., Prats, N., Toll, A., Hueto, J.A., et al. (2017). Targeting metastasis-initiating cells through the fatty acid receptor CD36. Nature *541*, 41–45.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Pedersen, B.S., Eyring, K., De, S., Yang, I.V., and Schwartz, D.A. (2014). Fast and accurate alignment of long bisulfite-seq reads. arXiv, 1401.1129.

Plaisier, S.B., Taschereau, R., Wong, J.A., and Graeber, T.G. (2010). Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. Nucleic Acids Res. *38*, e169.

Platta, C.S., Greenblatt, D.Y., Kunnimalaiyaan, M., and Chen, H. (2007). The HDAC inhibitor trichostatin a inhibits growth of small cell lung cancer cells. J. Surg. Res. *142*, 219–226.

Rickman, D.S., Beltran, H., Demichelis, F., and Rubin, M.A. (2017). Biology and evolution of poorly differentiated neuroendocrine tumors. Nat. Med. *23*, 1–10.

Robertson, A.G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A.D., Hinoue, T., Laird, P.W., Hoadley, K.A., Akbani, R., et al. (2018). Comprehensive molecular characterization of muscle-invasive bladder cancer. Cell *174*, 1033.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics *12*, 77.

Robinson, D.R., Wu, Y.-M., Lonigro, R.J., Vats, P., Cobain, E., Everett, J., Cao, X., Rabban, E., Kumar-Sinha, C., Raymond, V., et al. (2017). Integrative clinical genomics of metastatic cancer. Nature *548*, 297–303.

Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixOmics: an R package for 'omics feature selection and multiple data integration. PLoS Comput. Biol. *13*, e1005752.

Rudin, C.M., Durinck, S., Stawiski, E.W., Poirier, J.T., Modrusan, Z., Shames, D.S., Bergbower, E.A., Guan, Y., Shin, J., Guillory, J., et al. (2012). Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nat. Genet. *44*, 1111–1116.

Seymour, J.F., Kipps, T.J., Eichhorst, B., Hillmen, P., D'Rozario, J., Assouline, S., Owen, C., Gerecitano, J., Robak, T., De la Serna, J., et al. (2018). Venetoclax-rituximab in relapsed or refractory chronic lymphocytic leukemia. N. Engl. J. Med. *378*, 1107–1120.

Smith, B.A., Balanis, N.G., Nanjundiah, A., Sheu, K.M., Tsai, B.L., Zhang, Q., Park, J.W., Thompson, M., Huang, J., Witte, O.N., et al. (2018). A human adult stem cell signature marks aggressive variants across epithelial cancers. Cell Rep. *24*, 3353–3366.e5.

Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J. (2007). pcaMethods–a bioconductor package providing PCA methods for incomplete data. Bioinformatics *23*, 1164–1167.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., and Lander, E.S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U S A *102*, 15545–15550.

Sur, I., and Taipale, J. (2016). The role of enhancers in cancer. Nat. Rev. Cancer *16*, 483–493.

Therneau, T.M., and Grambsch, P.M. (2000). Modeling Survival Data: Extending the Cox Model (Springer Science & Business Media).

Triche T, Jr. (2014). IlluminaHumanMethylation450k.db: Illumina Human Methylation 450k annotation data. R package version 2.0.9.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a cancer dependency map. Cell *170*, 564–576.e16.

Tsoi, J., Robert, L., Paraiso, K., Galvan, C., Sheu, K.M., Lay, J., Wong, D.J.L., Atefi, M., Shirazi, R., Wang, X., et al. (2018). Multi-stage differentiation defines melanoma subtypes with differential vulnerability to drug-induced iron-dependent oxidative stress. Cancer Cell *33*, 890–904.e5.

Vivian, J., Rao, A.A., Nothaft, F.A., Ketchum, C., Armstrong, J., Novak, A., Pfeil, J., Narkizian, J., Deran, A.D., Musselman-Brown, A., et al. (2017). Toil enables reproducible, open source, big biomedical data analyses. Nat. Biotechnol. *35*, 314–316.

Wang, J., Wei, B., Albarracin, C.T., Hu, J., Abraham, S.C., and Wu, Y. (2014). Invasive neuroendocrine carcinoma of the breast: a population-based study from the surveillance, epidemiology and end results (SEER) database. BMC Cancer *14*, 147.

Watson, M., Roulston, A., Belec, L., Billot, X., Marcellus, R., Bedard, D., Bernier, C., Branchaud, S., Chan, H., Dairi, K., et al. (2009). The small molecule GMX1778 is a potent inhibitor of NAD+ biosynthesis: strategy for enhanced therapy in nicotinic acid phosphoribosyltransferase 1-deficient tumors. Mol. Cell. Biol. *29*, 5872–5888.

Watson, P.A., Arora, V.K., and Sawyers, C.L. (2015). Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. Nat. Rev. Cancer *15*, 701–711.

Watson, G.A., Ahmed, Y., Picardo, S., Chew, S., Cobbe, S., Mahony, C., Crotty, J., Wallis, F., Shelly, M.J., Kiely, P., et al. (2018). Unusual sites of high-grade neuroendocrine carcinomas: a case series and review of the literature. Am. J. Case Rep. *19*, 710–723.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis, 2016 (Springer-Verlag New York).

Wong, Y.N.S., Jack, R.H., Mak, V., Henrik, M., and Davies, E.A. (2009). The epidemiology and survival of extrapulmonary small cell carcinoma in South East England, 1970–2004. BMC Cancer *9*, 209.

Yachida, S., Vakiani, E., White, C.M., Zhong, Y., Saunders, T., Morgan, R., de Wilde, R.F., Maitra, A., Hicks, J., DeMarzo, A.M., et al. (2012). Small cell and large cell neuroendocrine carcinomas of the pancreas are genetically similar and distinct from well-differentiated pancreatic neuroendocrine tumors. Am. J. Surg. Pathol. *36*, 173–184.

Yang, D., Denny, S.K., Greenside, P.G., Chaikovsky, A.C., Brady, J.J., Ouadah, Y., Granja, J.M., Jahchan, N.S., Lim, J.S., Kwok, S., et al. (2018). Intertumoral heterogeneity in SCLC is influenced by the cell type of origin. Cancer Discov. *8*, 1316–1331.

Zaykin, D.V. (2011). Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. J. Evol. Biol. *24*, 1836–1841.

Zhang, X., Coleman, I.M., Brown, L.G., True, L.D., Kollath, L., Lucas, J.M., Lam, H.-M., Dumpit, R., Corey, E., Chery, L., et al. (2015). SRRM4 expression and the loss of REST activity may promote the emergence of the neuroendocrine phenotype in castration-resistant prostate cancer. Clin. Cancer Res. *21*, 4698–4708.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Deposited Data | | |
| gene expression data: TCGA pan-cancer | (Hoadley et al., 2018) | https://xenabrowser.net/datapages/?hub=https://pancanatlas.xenahubs.net:443 |
| Methylation beta values: TCGA pan-cancer | Genomic Data Commons | https://portal.gdc.cancer.gov/repository |
| Gene expression data: GEP-NETS | (Alvarez et al., 2018) | GEO: GSE98894 |
| RNAseq and Bisulfite FASTQs: NEPC/CRPC tumors | (Beltran et al., 2016) | dbGAP: phs000909.v1.p1 |
| Copy Number data: NEPC/CRPC | (Beltran et al., 2016) | Table S6 |
| SNP 6.0 chip and RNAseq FASTQs: SCLC Tumors | (George et al., 2015) | EGA: EGAS00001000925 |
| RNAseq FASTQs: MET500 | (Robinson et al., 2017) | dbGAP: phs000673.v2.p1 |
| gene expression data: pan-cancer CCLE | Office of Cancer Genomics: NIH | https://ocg.cancer.gov/ctd2-data-project/translational-genomics-research-institute-quantified-cancer-cell-line-encyclopedia |
| RNAseq FASTQs: lung CCLE | Broad Institute | https://portals.broadinstitute.org/ccle |
| Methylation beta values: CCLE | (Barretina et al., 2012) | GEO: GSE68379 |
| Methylation beta values: SCLC Tumors | (Mohammad et al., 2015) | GEO: GSE66298 |
| Drug $IC_{50}$:CCLE | (Iorio et al., 2016) | https://www.cancerrxgene.org/downloads |
| RNAi Sensitivity : CCLE | (Tsherniak et al., 2017) | https://depmap.org/portal/download/Achilles 2.20.2dataset |
| Software and Algorithms | | |
| Gene Set Enrichment Analysis (GSEA) | (Subramanian et al., 2005) | http://software.broadinstitute.org/gsea/index.jsp |
| R (v3.2) | N/A | https://www.r-project.org/ |
| Rank-rank Hypergeometric Overlap (RRHO) | (Plaisier et al., 2010) | http://systems.crump.ucla.edu/rankrank/rankranksimple.php |
| TOIL | (Vivian et al., 2017) | https://github.com/BD2KGenomics/toil-rnaseq |
| ARACNe | (Lachmann et al., 2016) | https://sourceforge.net/projects/aracne-ap/ |
| VIPER | (Alvarez et al., 2016) | https://www.bioconductor.org/packages/release/bioc/html/viper.html |
| pcaMethods | (Stacklies et al., 2007) | |
| mixOmics | (Rohart et al., 2017) | https://cran.r-project.org/web/packages/mixOmics/index.html |
| bwameth | (Pedersen et al., 2014) | https://github.com/brentp/bwa-meth |
| MethylDackel | N/A | https://github.com/dpryan79/MethylDackel |
| Caret | (Kuhn, 2016) | http://caret.r-forge.r-project.org/ |
| ggplot2 | (Wickham, 2016) | https://ggplot2.tidyverse.org/ |
| org.Hs.eg.db | (Carlson, 2019) | https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html |
| Rawcopy | (Mayrhofer et al., 2016) | http://rawcopy.org/ |
| GISTIC2.0 | (Mermel et al., 2011) | http://portals.broadinstitute.org/cgi-bin/cancer/publications/pub_paper.cgi?mode=view&paper_id=216&p=t |
| pROC | (Robin et al., 2011) | https://cran.r-project.org/web/packages/pROC/index.html |
| Survival | (Therneau and Grambsch, 2000) | https://cran.r-project.org/web/packages/survival/index.html |

(*Continued on next page*)

| Continued | | |
|---|---|---|
| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
| metap | N/A | https://cran.r-project.org/web/packages/metap/index.html |
| KSTEST2 | N/A | https://github.com/franapoli/signed-ks-test |
| Other | | |
| Resource website for the SCN publication | This paper | https://systems.crump.ucla.edu/scn/ |

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Thomas G. Graeber (TGraeber@mednet.ucla.edu).

### Abbreviations

TCGA cancer type abbreviations are found at https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations.

## METHOD DETAILS

### Data Acquisition and Processing

Gene expression data of primary tumors was obtained from the TCGA (https://xenabrowser.net/hub/), of SCLC from (George et al., 2015), of human NEPC and CRPC from (Beltran et al., 2016), of the MET500 data which contains non SCN and SCN metastatic samples across multiple cancers from (Robinson et al., 2017), of all lung lines in the CCLE from (Barretina et al., 2012). To minimize batch effects, we uniformly processed the raw FASTQs through a single analysis pipeline (TOIL) (Vivian et al., 2017). CCLE lung cell lines were processed through TOIL when comparing to TCGA samples which were also processed through TOIL. Another version of the CCLE expression data on all cell lines was obtained in a format pre-processed through Salmon (Patro et al., 2017) (https://ocg.cancer.gov/ctd2-data-project/translational-genomics-research-institute-quantified-cancer-cell-line-encyclopedia), and all analyses done on CCLE cell lines only were performed using this data from this processing pipeline (Figure 6A). The CLCGP lung cancer microarray gene expression dataset used to build the ARACNe network was downloaded from www.uni-koeln.de/med-fak/clcgp. For RNAseq, upper quartile normalized expression values were transformed to $log_2(x + 1)$.

The 450 K methylation array data for cell lines was obtained from Iorio et al. (2016), for human SCLC tumors from Mohammad et al. (2015), and for cancers in the TCGA from Hoadley et al. (2018) in processed form. The reduced representation bisulfite sequencing (RRBS) data for human NPEC and CRPC tumors was obtained from Beltran et al. (2016) in FASTQ format (dbGAP:phs000909.v1.p1) and aligned to hg38 using bwameth (Pedersen et al., 2014). Methylation metrics were called using MethylDackel, which groups cytosines into one of three sequence contexts: CpG, CHG, or CHH. Only cytosines in CpG context were used for downstream analysis.

Drug Sensitivity data was obtained processed as $IC_{50}$ values from the Genomics of Drug Sensitivity in Cancer website (https://www.cancerrxgene.org). RNAi sensitivity data (Achilles version 2.20.2) was obtained processed (Demeter1 scores) from the Depmap website (https://depmap.org/portal/)

Cell line annotation was performed by harmonizing the annotation of the CCLE, GDSC, and Demeter datasets. Cell line annotations were cross checked with ATCC (https://www.atcc.org/), Cellosaurus (https://web.expasy.org/cellosaurus/), and DSMZ (https://www.dsmz.de/) when possible. When discrepancies arose between these 3 online sources and the primary annotation, we defaulted to the online sources, which have stringent analysis pipelines. Lines with problematic annotation as defined by Cellosaurus were left out of the analyses. Tumor and cell line annotations used are in Table S7 Cell line culture growth characteristics (e.g. adherent vs suspension) were obtained from the Dependency Map database (https://depmap.org/portal/) and (Iorio et al., 2016).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### PCA/PLSR

Log2 transformed upper quartile normalized expression of coding genes was used to perform unsupervised principal component analysis (PCA). This method uncovers latent components which are a linear combinations of the features that most strongly vary across the datasets. PCA was performed centered and unscaled using the *prcomp* function in R. Partial Least Squares Regression (PLSR) is a supervised version of PCA that seeks to find the latent vectors that maximize the covariance of the input variables (e.g. gene expression) and the response (e.g. phenotypes). Varimax rotation of the PCA ($PCA_v$) or PLSR loadings ($PLSR_v$) was performed on 2 components, without Kaiser normalization, using the R *varimax* package. Projections onto varimax-rotated PCA/PLSR frameworks were done by multiplication of the original projected sample scores by the varimax rotation matrix. When applicable

(e.g. Figure 1A) TCGA cancers (LUAD, PRAD, and BLCA) were randomly down sampled to more closely match the number of samples in data sets from George et al. and Beltran et al.

## Small Cell Neuroendocrine (SCN) and Proliferation Scores

### *Tumors*

For RNAseq data using patient tumors, ''SCN score'' is the PC1 score after projection onto the Figure 1A varimax PCA framework. Because of the nature of PCA, this score is determined as a linear combination of weights that includes every coding gene, and hence is not strictly dependent on solely one subset of genes. After projection onto this framework, the score was either z-scored in each individual tumor type as in Figure 4A (to highlight outlier samples), or left un-zscored to place cancers on a common scale as in Figures 5A, 5C, and S6A. For the two left boxes in Figure 4A, SCN scores are z-normalized w.r.t CRPC (leftmost box) and LUAD (middle box). For samples in the right box, SCN score is the value of $PC_v1$ from projection onto Figure 1A, z-score normalized by cancer type. Samples greater than 3 standard deviations from the mean in the z-scored analysis were deemed ''SCN-like'' (e.g. Figures 4A and 4B).

As SCNCs are highly proliferative and display neuroendocrine features we sought to de-convolve these two influences on SCN score. To that end we created a ''SCN minus proliferation score''. A list of proliferation genes was generated from the union of three lists of proliferation genes published by Benporath et al, Cyclebase (https://cyclebase.org/), and KEGG cell cycle genes. PCA was performed using only these proliferation genes, and the absolute value of each gene's Pearson correlation to PC1 was calculated. A ROC curve was created using two classes with the *pROC* package to choose a threshold cutoff (Youden's J statistic) for genes highly correlated with proliferation. Correlated genes above threshold and all annotated proliferation genes from the original list were removed. PCA and GSEA-squared analysis (see below) was then redone. ''SCN minus proliferation score'' is thus the PC1 score after sample projection onto the varimax PCA of the samples used in Figure 1A with this new gene list. A ''Proliferation Only Score'' was also created using the union of genes in the three lists and those removed using the ROC curve method. These proliferation-related genes were used to create another varimax PCA of the samples in Figure 1A. Proliferation Score is the varimax PC1 score after projection onto this framework. The proliferation-removed score was used in the analysis of the breast cancer slides to highlight samples that had a high probability of having neuroendocrine features. This method additionally has utility in distinguishing between SCN samples which have both neuroendocrine and proliferative features, from the indolent primitive neuroectodermal tumors which have neuroendocrine features but lack a proliferative signature. The proliferation-removed score was used in Figures 4I, S3I, S5C, and S5D. For Figures 4I and S3J, the proliferation-removed score was z-scored, since only BRCA tumors were involved in these panels. For Figure S5D, both the x and y-axes were left not z-score normalized to place all samples on a common scale.

### *Cell Lines*

For cell lines, RNA gene expression and RNAi sensitivity based SCN scores were calculated by projection onto the varimax PLSR framework of the SCLC/LUAD dichotomy. For both these data types, SCN Score is the ''component 1'' score after projection onto this framework. Samples with annotation not concordant with their expression-based predictions, from a linear-discriminant Leave-one-out Cross Validation analysis using the first three components of an expression PCA, were removed in the RNAi susceptibility-based PLSR. For protein data, the ''SCN score'' is the PC1 score of a sample upon projection onto the protein data-based framework of the SCLC/LUAD dichotomy (e.g. Figure 6B). For drug sensitivity, as the distinction between non-SCN (LUAD) and SCN cases (SCLC) is on PC2 , the drug sensitivity ''SCN score'' is the PC2 score of a sample after projection onto that framework (e.g. Figure 6C).

## Transcription Factor Analysis

ARACNe (Lachmann et al., 2016) network connections were created using all genes, and then the network nodes were restricted to 1675 transcription factors (TFs) by combining all TF gene sets in the GO gene ontology. One network was built for each of lung (Clinical Lung Cancer Genome Project (CLCGP) and Network Genomic Medicine (NGM), 2013), prostate (Beltran et al., 2016), and bladder (TCGA), using a balanced set of samples from the SCN and adenocarcinoma groups when possible, with default settings. VIPER analysis (Alvarez et al., 2016) was performed using the *msviper* function from R package *viper*, with a minimum network size of 10. The combined p value across the three tissues was calculated using Stouffer's method by converting two-way p values from *msviper* into one-way p values using the *two2one* and *sumz* functions from the *metap* package in R.

## GSEA-Squared

Gene Set Enrichment Analysis (Subramanian et al., 2005) was done on pre-ranked lists of genes using the MSigDB C5 gene sets and Kolmogorov-Smirnov (KS)-based statistics. In order to identify categories of genesets that could be enriched or de-enriched, all individual words in the genesets were collected and their frequencies were tabulated. Words with frequencies <5 or >500 were excluded. All genesets were then ranked by their NES value. Using genesets containing each particular word, all the individual words were then ranked by their KS test p value using *ks.test.2* at (https://github.com/franapoli/signed-ks-test). Top words with small p values were considered categories of interest, such as 'immune' or 'neuron'. A manual curation of other top words related to

the categories, and inspection of gene sets containing these words, was done to group related gene sets. The keywords used for each category are listed below:

| Category | Keywords |
|---|---|
| Neuro | NEURO, SYNAP, VOLTAGE, AXON, CEREBRAL, CORTEX |
| Cell cycle | CELL_CYCLE, MITOTIC, DNA_REPLICATION, CHROMOSOME_SEGREGATION, SPINDLE, CELL_DIVISION |
| Splicing | SPLICING, SPLICESOME, SPLICESOMAL |
| Immune | INFLAM, IMMUNE, IMMUNITY, INTERLUEKIN, LEUKOCYTE |
| Adhesion | ADHESION, ADHERENS |

All gene sets were then ranked by their Normalized Enrichment Score (NES), and KS tests were performed to assess the distribution of gene set categories using *ks.test.2*. This second application of KS test-based enrichment analysis led to the coining of 'GSEA-squared,' enrichment analysis first on genes, and then on geneset categories. For multiple-hypothesis testing of KS tests of enrichment in Figures 7C–7F, each category's one-way p values were corrected using the Benjamani-Hochberg method placing them within the full list of keywords.

### REST Analysis

REST/NRSF is transcriptional repressor and restricts neural gene expression. We used an aggregate measure of REST activity by calculating an inferred "REST Activity Score" for each sample from RNAseq data. "REST Activity Score" for individual samples was determined using the gene set NRSF_01 from MSIGDB (http://software.broadinstitute.org/gsea/msigdb/geneset_page.jsp?geneSetName=NRSF_01) which contains genes that have a 3' UTR motif that matches annotation for the REST transcription factor. In this model a higher "REST Activity Score" corresponds to de-repression of REST target genes, and hence conditions that support more neural gene expression, For each sample, a "single sample gene set enrichment score" (ssGSEA) was determined using the V$NRSF_01 gene set. *REST* gene expression and "REST Activity Score" were correlated to "SCN Score" in individual cancer types using the Pearson's correlation (Barbie et al., 2009). Possibilities for the discrepancy between REST activity and *REST* expression is that REST function can be lost in multiple ways such as by mutation, (Mahamdallie et al., 2015), or by truncation/alternative splicing that abrogates its function (Chen and Miller, 2018) – both of which would not require concordant changes in transcription levels.

### Rank Rank Hypergeometric Overlap (RRHO)

Rank Rank Hypergeometric Overlap was performed for 'signature overlap analysis' using the online tool and the R package RRHO, with step size 100 for expression data and gene-based methylation data, and 2000 for probe-based methylation data such as that used in the 'signature overlap analysis' (Plaisier et al., 2010).

### Methylation Analysis

Methylation levels were expressed as β-values, indicating the overall proportion of methylation at each particular site [methylated / (methylated+umethylated)]. PCAs were performed centered and unscaled on the entire data matrix. The IlluminaHumanMethylation450k.db package was used to provide annotation information on the location of the probe in relation to regulatory elements. Tissue-agnostic enhancer locations were provided by the IlluminaHumanMethylation450k.db package (Triche, 2017), which informatically determines enhancer probes using ENCODE data. For the probe-based methylation data, PLSR was run individually on lung and bladder tissues, regressing against the phenotype of SCN or non-SCN. Thus, extreme values in the PLSR loadings represent sites with differential methylation between SCN and non-SCN tumors. Non-SCN samples for lung and bladder were down-sampled to more closely match the number of SCN samples. Signature overlap analysis was then performed using RRHO on probe-level data. Prostate, which was sequenced by reduced-representation bisulfite sequencing (RRBS) was not included for site level analysis due to much fewer sites measured by the microarray platform used for bladder and lung. To test for the importance of open-sea regions, which are typically distant from the transcriptional start site (TSS) rather than CG island regions (TSS proximal), in contributing to the SCN versus non-SCN distinction, the absolute value of the loadings was ranked, and a 1-sided KS test was performed against a background of all sites. To incorporate prostate data with the lung and bladder 450K array data, the number of sites was reduced to sites represented across both platforms. Sites were then further collapsed to genes by matching probes to genes using IlluminaHumanMethylation450k annotation. Probes that matched to multiple genes based on the Illumina annotation were removed. Averaged methylation values for each gene were then ranked by PLSR loadings on each tissue type. Genes were ranked by averaged lung, prostate, bladder PLSR component 1 loadings, and GSEA was performed on this ranked list. The gene-based summarization of methylation data was also performed on the lung cell lines and PLSR was performed. The equivalent methylation data format from tumors was projected to the cell line framework.

### CNA Analysis

TCGA SNP6.0 Affymetrix derived seg files were downloaded from the GDC repository. Cell line seg files were created using RAWCOPY from the .cel files with default settings (Mayrhofer et al., 2016). Seg files were inputted into GISTIC2.0 to obtain both

thresholded calls and continuous log2 CNA values mapped to genes. PCA was performed uncentered and unscaled on the continuous log2 CNA data. IGV was used for visualization. For lung, a random sample of LUAD samples and all SCLC samples were used. For prostate, the samples were subset to include only one sample from each patient. One region containing highly focal CNAs on Chromosome 1 was removed by inspection of prostate PCA loadings, because it vastly dominated the top components' loadings of the PCA analysis and was determined to be a likely technical artifact. To determine consistently amplified or deleted regions in SCN cases of lung and prostate SCN samples, we performed PLSR on lung and prostate samples together or independently, regressing on phenotype in each case (1 and 0 representing SCN and non-SCN samples respectively). To determine consistent regions of amplifications and deletions, we evaluated the loadings of the individual PLSR analyses (consistent regions were defined by being commonly positive or commonly negative in the CNA PLSR loading values). The loadings for consistent regions were averaged, non-consistent regions were set to 0.

Integrated CNA (iCNA) score (Graham et al., 2017) for each sample was defined as:

$$iCNA = \frac{\sum_{segments} |segment\ end - segment\ start| \times |segment\ mean|}{\#base\ pairs\ in\ sample}$$

## Pathology Analysis

To evaluate the molecular profile (gene expression)-based SCN predictions, a pathologist evaluated the available TCGA hematoxylin and eosin (H&E) stained histology slide images located at the Cancer Digital Slide Archive (http://cancer.digitalslidearchive.net/) (Gutman et al., 2013). Typically, each TCGA case has sections from formalin-fixed paraffin-embedded (FFPE) tissue and from frozen tissue. To avoid artifacts associated with frozen sections, only FFPE sections were used for histology classification. For the pan-cancer histology analysis, 16 tumor samples with high proliferation-removed SCN signature scores from multiple tissue types were analyzed by a pathologist. Representative images of neuroendocrine and non-neuroendocrine regions were obtained by taking screen shots. Cases with no SCN region identified may be explained by tumors with focal SCN, and pathology calls being performed on tissue not adjacent to the tissue used for sequencing.

From the breast cancer (BRCA) cohort we ranked samples by their proliferation-removed SCN score, and sampled to cover a full range of proliferation-removed SCN scores. Thirty-eight samples were selected to include 1) the majority of SCN high cases, 2) a random sampling of tumors with middling SCN scores, 3) tumors with low SCN score, and 4) samples with high expression levels of at least one of four traditional SCN markers, *CHGA*, *SYP*, *NCAM1*, and *TTF1*. Each digital slide was divided into 4 roughly equal quadrants to ensure that the slide was examined evenly across different regions. In each quadrant, 20 non-contiguous areas were examined at low, medium and high powers to determine the histologic features of the tumor. Representative images were obtained at the indicated magnification by taking screen shots. Cases were classified as mixed tumors when two different histologic types (most commonly, invasive ductal carcinoma and small cell neuroendocrine carcinoma) co-existed. For a case to be classified as a mixed tumor, the minor component (usually, small cell neuroendocrine carcinoma) needed to occupy a substantial area of the tumor. Specifically, we required that the tumor cells of the minor histology should coalesce in a region that is equal to or larger than 2 high power fields. To determine if pathology based-SCN positive cases were enriched in samples with high SCN score, samples were rank ordered by SCN score and a Kolmogorov-Smirnov enrichment test was performed on pathology-based SCN status (Table S3). The Web Resource contains a pathology page with detailed histology images (http://systems.crump.ucla.edu/scn/). For Table S3, grade and stage information were obtained from the TCGA Pan-Cancer Clinical Data Resource (https://gdc.cancer.gov/about-data/publications/PanCan-Clinical-2018).

## Survival Analysis

For the pan-cancer analysis, Cox regressions were performed based on SCN-like/non-SCN status (SCN-like status = SCN z-score of 3 and above), controlling for cancer type using the *coxph* function in R. Various thresholds of the SCN score were used for the pan-cancer survival analysis to make Table S3. Survival analysis based on a continuous scale was performed using the $PC_v1$-based SCN score determined by projection onto the Figure 1A PCA plot for epithelial cancers, and controlling for cancer type for the pan-cancer analysis, using the *coxph* function in R. Wald-test p values were reported.

## Mutation Analysis

Mutations were assembled from the MutsigCV2 calls for each tumor type from firebrowse.org. The gene list used was the union of recurrent mutations (mutsigcv2 q-value < 0.1) from all cancers in the TCGA. To determine the association of mutations with the neuroendocrine phenotype a generalized linear model (GLM) was used. In the analysis of cancers with known SCN cases (Figure S4A) a logistic regression was conditioned on the SCN score and the tissue type on the combined data from "prostate" (consisting of NEPC, CRPC, PRAD) and "lung" (SCLC, LUAD). For the pan-epithelial cancer analysis (Figure S4B) the logistic regression was conditioned on the $PC_v1$-based SCN score (projection onto Figure 1A) and cancer type using all epithelial cancers in the TCGA. In the individual cancer cases (Figure S4C) a Wilcoxon rank sum test on neuroendocrine score of mutant vs non-mutant was used. Multiple hypothesis correction was performed with the Benjamani-Hochberg method.

### RPPA Analysis

RPPA data was obtained from (Li et al., 2017). As this data has missing values, imputation was performed using probabilistic PCA, using the *ppca* and *completeObs* functions in the *pcaMethods* package (Stacklies et al., 2007). LUAD and SCLC cell lines were processed together in one batch without other cell lines so as not to intermix test and training sets. Missing value imputation for all other cell lines were performed together in one batch. Prior to each imputation we removed all proteins with greater than 25% missing values in that batch. Samples were projected onto the PCA of the imputed values SCLC and LUAD (Figure 6B). The SRBCT group in the protein data consisted of neuroblastoma, medulloblastoma, Ewing's sarcoma, and rhabdomyosarcoma.

### Drug Sensitivity Analysis

Drug sensitivity data (log $IC_{50}$ values) of 255 small molecules across a wide panel of cancer cell lines from multiple tissue types was obtained from the GDSC (Iorio et al., 2016). For LUAD and SCLC cell lines, differential sensitivity to each drug was calculated using the Student's t-test. These drugs were then ranked by the t-test statistic. Annotation on drug target and target pathway was obtained from the pharmacogenomics screen published by Iorio et al. From the annotation on drug targets, a list of genes or biological targets was obtained. For each target, a KS test was performed (bootstrap n = 1000) on the t-test ranked list of drugs that contained that target in its annotation, resulting in a list of targets significantly enriched or de-enriched for small cell cancer sensitivity. Missing values in the drug sensitivity data were imputed using the weighted average of k-nearest neighbors (n = 3). For SCLC and LUAD samples that were used as training data, k-Nearest Neighbors (kNN) imputation was performed on the lung samples alone. kNN imputation was then then performed on all other cell lines together, excluding the SCLC and LUAD samples. PCA was performed on the imputed matrix of drug sensitivity log $IC_{50}$ values across SCLC and LUAD cell lines using all drugs in the screen. We then projected the drug data for all lines onto the PCA defined on lung SCLC and LUAD, including cancer of hematological and neuroectodermal origin. Projected points (all lines excluding lung SCLC and LUAD) in the drug sensitivity plot were annotated by their expression projection values. These values were binned along the x-axis and mean expression projection values were summarized in the corresponding waterfall plot (eg. Figures 6B and 6C). Combined p values were obtained by the weighted-Stouffer test with the one way null hypothesis that the Pearson correlation is less than zero; using the lower tail of the t-distribution for values outputted by *cor.test* in R (lung small cell not included). The square root of individual group sample sizes were used as weights to stabilize the variance of the mean in this calculation (Zaykin, 2011). The SRBCT group in the drug sensitivity data consisted of neuroblastoma, medulloblastoma, Ewing's sarcoma, and rhabdomyosarcoma.

### Gene Dependency

shRNA data was taken from the Achilles Project (Tsherniak et al., 2017). Demeter gene dependency scores were used. A lower Demeter score indicates sensitivity to downregulation of that gene. PLSR was performed on 1) LUAD and SCLC lines, and 2) blood and non-blood lines (all cell lines except SRBCT, LUAD, or SCLC lines). The SRBCT group in the gene dependency data consisted of neuroblastoma, medulloblastoma, Ewing's sarcoma, rhabdomyosarcoma, and merkel cell carcinoma. Varimax rotation was performed on 2 components. Other cell lines were projected onto the lung or blood PLSR frameworks using the varimax-rotated loadings. Student's t-test was performed on LUAD versus SCLC lines, and blood versus non-blood lines, producing two gene lists, each ranked by p value representing differential RNAi sensitivity in the two above comparisons. RRHO was performed on these two ranked lists of genes, and on their corresponding GSEA-analyzed genesets ranked by NES score. For each gene, the sum of the LUAD versus SCLC rank and blood versus non-blood rank for each gene was calculated, and the list was co-ranked using this sum. Enrichment analysis (GSEA) was performed on the co-ranked list of genes using the MSigDB C2: KEGG, C2: Reactome, and C5: all GO gene sets. GSEA-squared (see above) was performed by ranking individual words by their signed KS test p value using *ks.test.2*. The keywords used for each category are listed below:

| Category | Keywords |
|---|---|
| Telomere | TELOMERE |
| Immune | INFLAM, IMMUNE, IMMUNITY, INTERLUEKIN, LEUKOCYTE |
| Cell cycle | CELL_CYCLE, MITOTIC, DNA_REPLICATION, CHROMOSOME_SEGREGATION, SPINDLE, CELL_DIVISION |
| Lipids | COA, LIPID, STEROL |
| Neuro | NEURO, SYNAP, VOLTAGE, AXON, CEREBRAL, CORTEX |

### Tumor Drug Sensitivity Prediction

Tumor drug sensitivity prediction were performed using elastic nets (ENET) with the caret package in R, using cell line RNAseq and $IC_{50}$ drug sensitivity data. For each drug two distinct models were created. 1) Using the 1000 most variable genes at the RNA level to predict drug sensitivity in the SCLC and LUAD cell lines (used in Figures 8A–8D), and 2) using 1000 of the most variable genes across all cell lines (used in Figures 8E and 8F). This resulted in a total of 510 models (255 drugs each). For the models used in Figures 8A–8D,

model cross-validation performance improves for drugs with significant true differential sensitivities, showing that the models are built on relevant gene expression features. Using the predicted values, tumors were projected onto the PCA of real drug sensitivity values shown in Figure 6C. The relative sensitivity scores are the PC2 values from this projection.

## DATA AND SOFTWARE AVAILABILITY

### Public Data Resources

Data from the CCLE (expression) and GDSC (drug sensitivity and methylation) databases were downloaded from the Broad Institute and Cancer X Gene resources (http://www.broadinstitute.org/ccle; http://www.cancerrxgene.org/downloads), respectively.

**Cancer Cell, Volume *36***

## Supplemental Information

# Pan-cancer Convergence to a Small-Cell Neuroendocrine Phenotype that Shares Susceptibilities with Hematological Malignancies

**Nikolas G. Balanis, Katherine M. Sheu, Favour N. Esedebe, Saahil J. Patel, Bryan A. Smith, Jung Wook Park, Salwan Alhani, Brigitte N. Gomperts, Jiaoti Huang, Owen N. Witte, and Thomas G. Graeber**

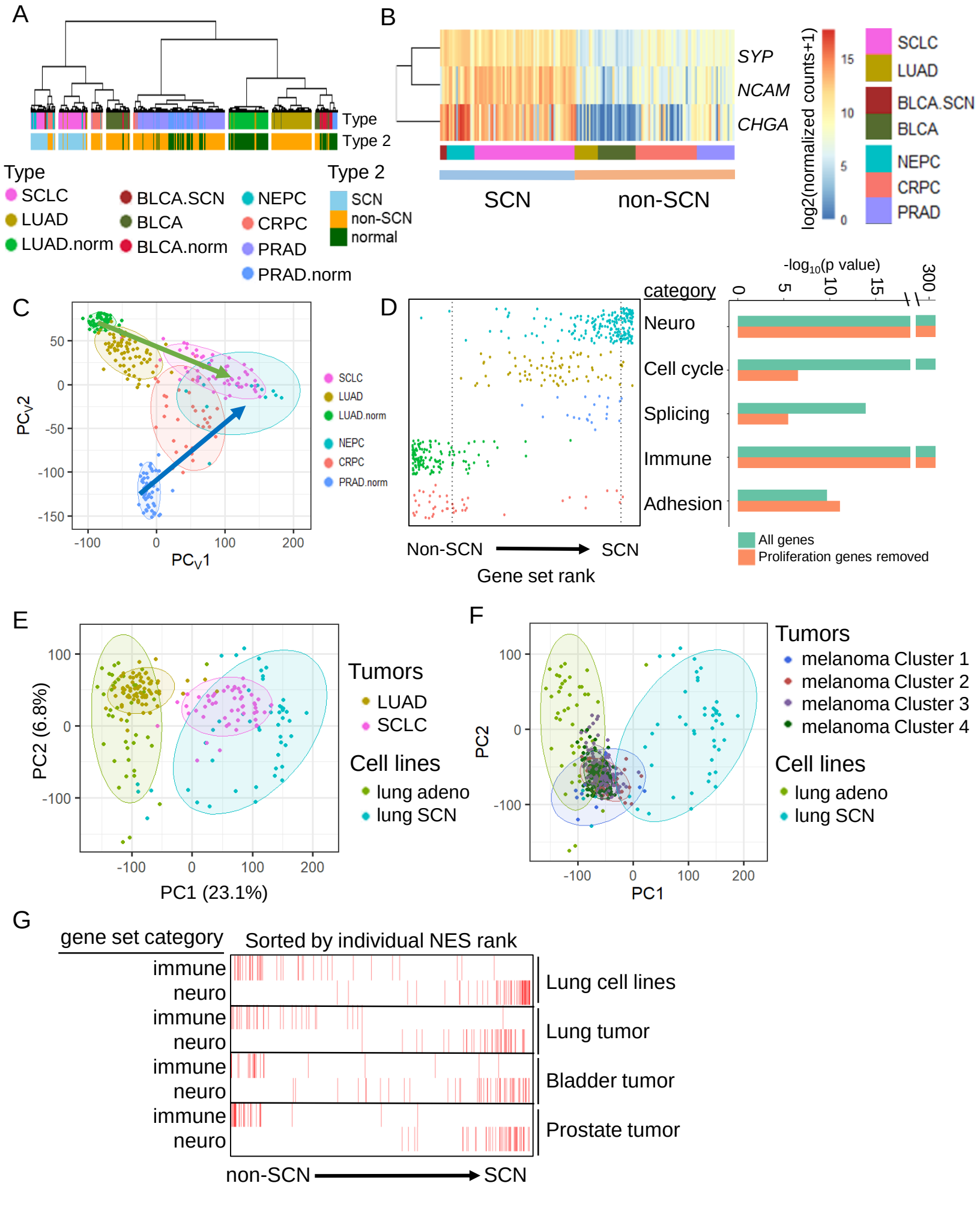

A
Type
Type 2
Type
SCLC
LUAD
LUAD.norm
BLCA.SCN
BLCA
BLCA.norm
NEPC
CRPC
PRAD
PRAD.norm
Type 2
SCN
non-SCN
normal
B
SYP
NCAM
CHGA
SCN
non-SCN
log2(normalized counts+1)
15
10
5
0
SCLC
LUAD
BLCA.SCN
BLCA
NEPC
CRPC
PRAD
C
PCV2
PCV1
50
0
-50
-100
-150
-100
0
100
200
SCLC
LUAD
LUAD.norm
NEPC
CRPC
PRAD.norm
D
-log10(p value)
category
0
5
10
15
300
Neuro
Cell cycle
Splicing
Immune
Adhesion
Non-SCN
SCN
Gene set rank
All genes
Proliferation genes removed
E
PC2 (6.8%)
PC1 (23.1%)
100
0
-100
-100
0
100
200
Tumors
LUAD
SCLC
Cell lines
lung adeno
lung SCN
F
PC2
PC1
100
0
-100
-100
0
100
200
Tumors
melanoma Cluster 1
melanoma Cluster 2
melanoma Cluster 3
melanoma Cluster 4
Cell lines
lung adeno
lung SCN
G
gene set category
Sorted by individual NES rank
immune
neuro
immune
neuro
immune
neuro
immune
neuro
Lung cell lines
Lung tumor
Bladder tumor
Prostate tumor
non-SCN
SCN

**Figure S1. Gene expression signatures of three tissue types supports convergent expression profiles of SCN cancers. Related to Figure 1.** (A) Hierarchical clustering on Pearson's correlation using all genes of patient tumors. (B) Heterogeneity of canonical neuroendocrine markers in SCNC tumors, with a representative set of non-SCNC tumors for comparison. (C) Varimax-rotated PCA with proliferation and proliferation-related genes removed on patient tumors. Ellipses represent 80% confidence regions. (D) Left: GSEA on proliferation-removed $PC_V1$ loadings of (C). Dashed lines mark FDR q-value <0.05 in each direction. Right: Bar chart of individual categories Kolmogorov-Smirnov test $-\log_{10}$(p values) for enrichment in SCN (for Neuro, Cell cycle and Splicing) or in Non-SCN (for Immune and Adhesion). (E) Projection of patient tumor samples onto lung cell lines. Lung cell lines are from the same batch, with these results thus mitigating concerns of a batch effect between patient tumor datasets. Ellipses represent 80% confidence regions. (F) Melanoma (SKCM) tumors projected onto lung cell line-defined PC components. The neuroendocrine dedifferentiation signature is distinct from other dedifferentiation trajectories, such as melanoma dedifferentiation (clusters 1-4 from Tsoi et al.). (G) GSEA of a t-test defined SCNC versus adenocarcinoma signature. Gene sets are ranked by NES score. Kolmogorov-Smirnov p values $< 2\times10^{-16}$ for all.

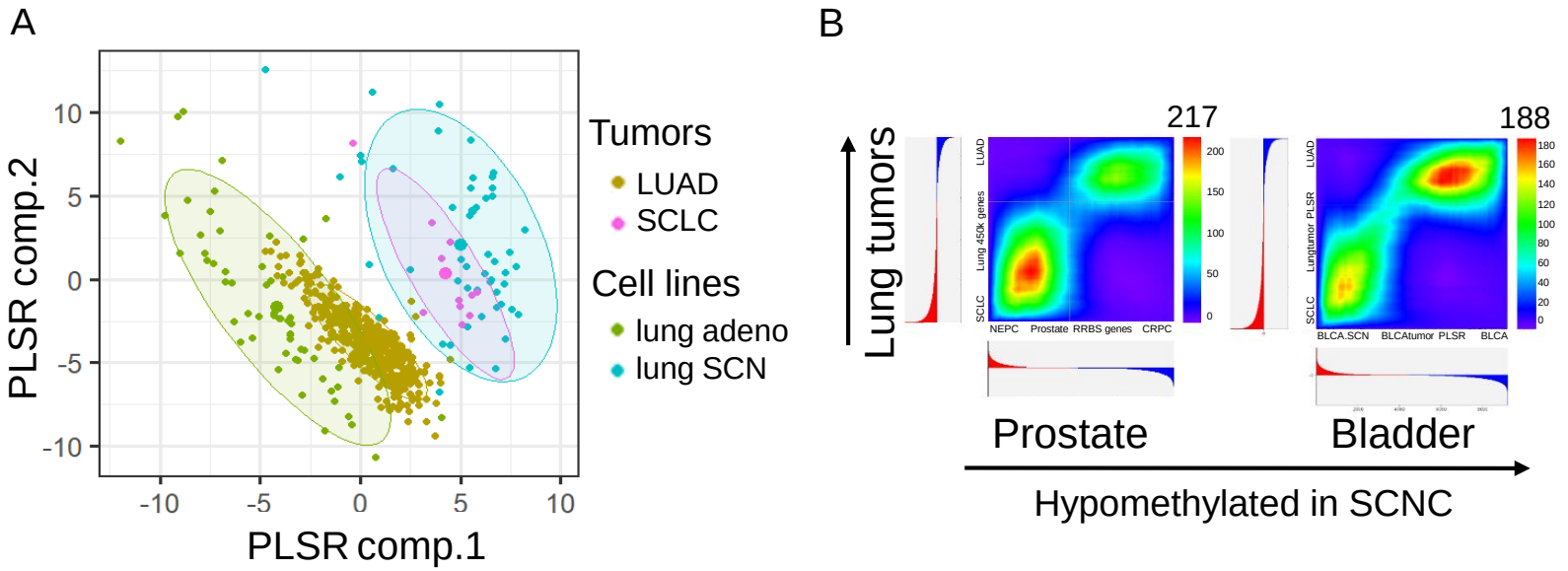
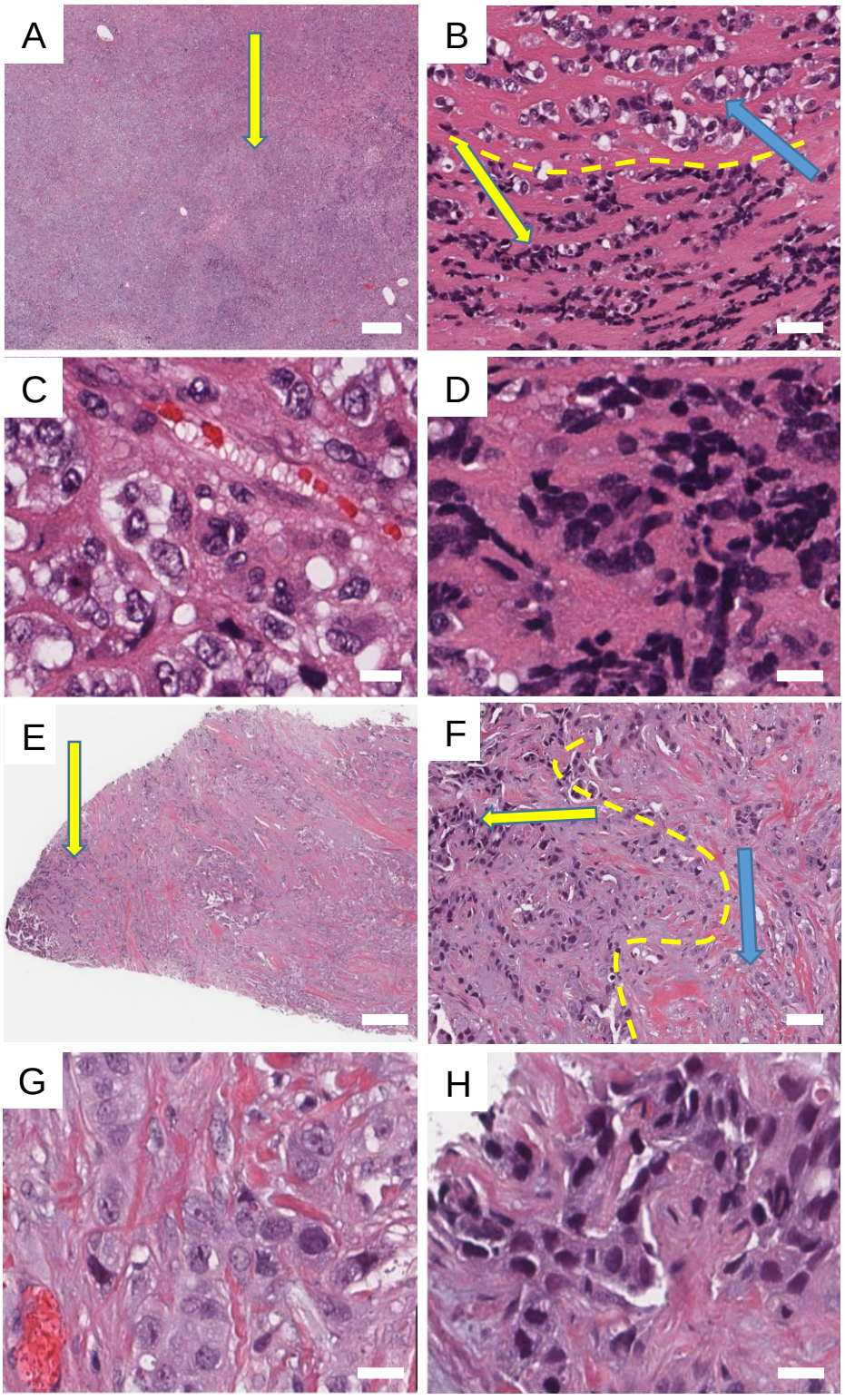
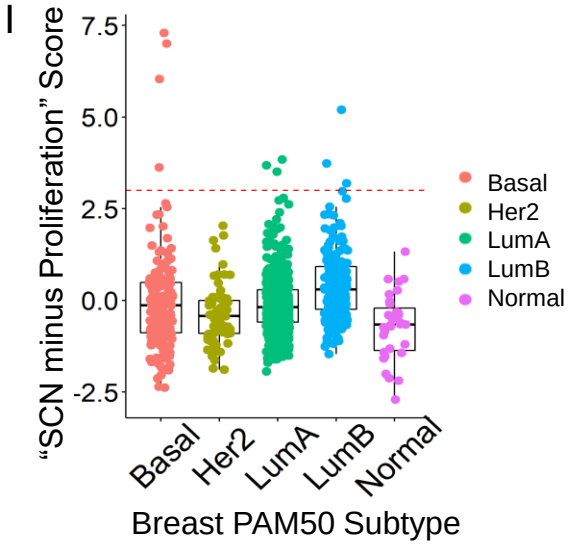
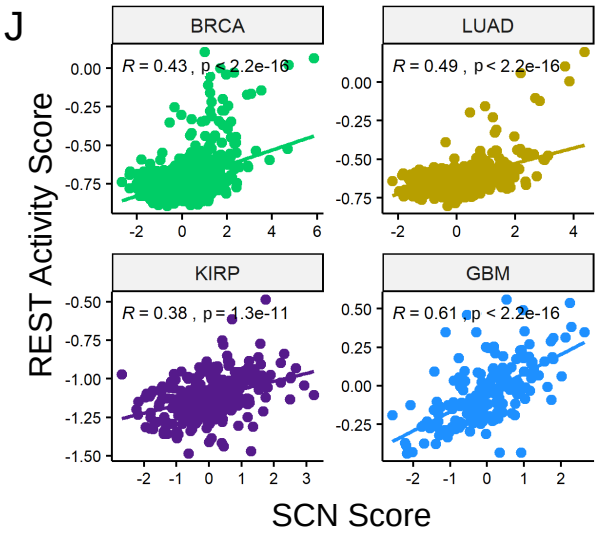
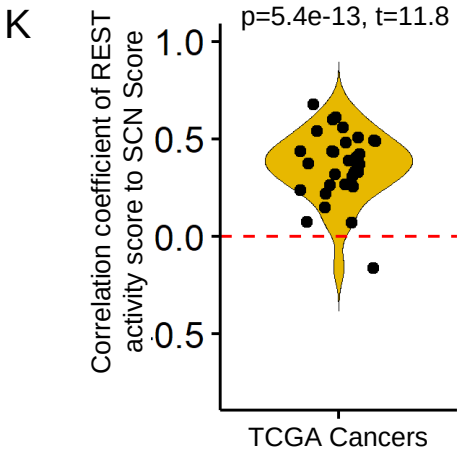



**Figure S2. SCN cancer convergence is reflected by epigenetic changes. Related to Figure 2.** (A) Lung tumor methylation projected to lung cell line methylation. Cell lines are from the same batch, mitigating concerns of batch effect in tumor data derived from two different datasets. Ellipses represent 80% confidence regions. (B) Rank signature overlap (RRHO) of PLSR component 1 loadings of matched sites averaged to gene loci across lung, prostate, and bladder. Only the subset of sites that matched across 450K microarray and RRBS sequencing platforms were used. The average rank of these ranked gene lists created independently on the 3 separate tissues were used in the enrichment analysis in Fig 2E. Numbers shown are maximum $-\log_{10}$(p values) of the RRHO heatmap.


A
B
C
D
E
F
G
H


I
"SCN minus Proliferation" Score
7.5
5.0
2.5
0.0
-2.5
Basal
Her2
LumA
LumB
Normal
Breast PAM50 Subtype


J
REST Activity Score
BRCA
R = 0.43 , p < 2.2e-16
LUAD
R = 0.49 , p < 2.2e-16
KIRP
R = 0.38 , p = 1.3e-11
GBM
R = 0.61 , p < 2.2e-16
SCN Score


K
p=5.4e-13, t=11.8
Correlation coefficient of REST activity score to SCN Score
1.0
0.5
0.0
-0.5
TCGA Cancers

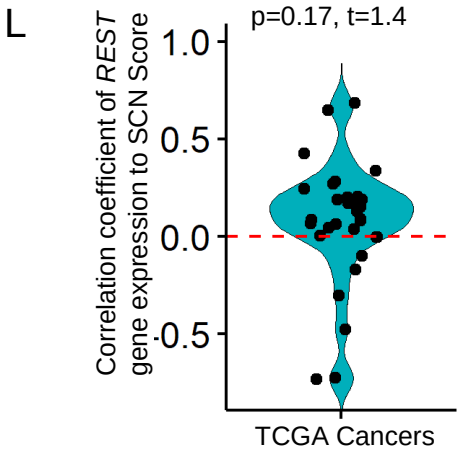

L
p=0.17, t=1.4
Correlation coefficient of REST gene expression to SCN Score
1.0
0.5
0.0
-0.5
TCGA Cancers

**Figure S3. Tumors with SCN phenotype in breast cancer. Related to Figure 4**. Tumor tissues from the TCGA breast cancer (BRCA) cohort scored by pathologist review of diagnostic slides for SCN features. (A-D) TCGA-AC-A2QH. The case was originally diagnosed as invasive ductal carcinoma. Careful examination of the digital picture available at the Cancer Digital Slide Archive website revealed focal areas of SCNC. (A) Low power view of the case. Yellow arrow indicates the junction of invasive ductal carcinoma and SCNC. Scale bar represents approximately 500 μm. (B) Medium power view of the junction of invasive ductal carcinoma and SCNC. Blue arrow points to the component of invasive ductal carcinoma, while yellow arrow points to the component of SCNC. Scale bar represents approximately 50 μm. (C) High power view of invasive ductal carcinoma region. Scale bar represents approximately 20 μm. (D) High power view of SCNC region. Scale bar represents approximately 20 μm. (E-H) TCGA-A7-A13D. The case was originally also diagnosed as invasive ductal carcinoma. Careful examination of the digital picture available at the TCGA website revealed focal areas of SCNC. (E) Low power view of the case. Yellow arrow indicates the junction of invasive ductal carcinoma and SCNC. Scale bar represents approximately 500 μm. (F) Medium power view of the junction of invasive ductal carcinoma and SCNC. Blue arrow points to the component of invasive ductal carcinoma, while yellow arrow points to the component of SCNC. Scale bar represents approximately 50 μm. (G) High power view of invasive ductal carcinoma region. Scale bar represents approximately 20 μm. (H) High power view of SCNC region. Scale bar represents approximately 20 μm. (I) Boxplots of SCN Score by PAM50 breast subtype. Dashed red line (y= 3) is threshold for calling tumor SCN-like. Lines inside boxplots represent the $25^{th}$, $50^{th}$ , and $75^{th}$ quantiles. Whiskers extend to 1.5x the interquartile range. (J) Scatter plot of REST Score vs SCN Score for 4 tumor types (R is Pearson's correlation). (K-L) Violin plot of Pearson's R from correlation of REST Score (K) or REST gene expression (L) to SCN score for TCGA cancers (dots are the individual tumor type R values). P value from one way t-test (null hypothesis mean = 0), t is the signed t statistic. Violin width is a kernel density estimation of the distribution of the data. Dashed red line at R value of zero.

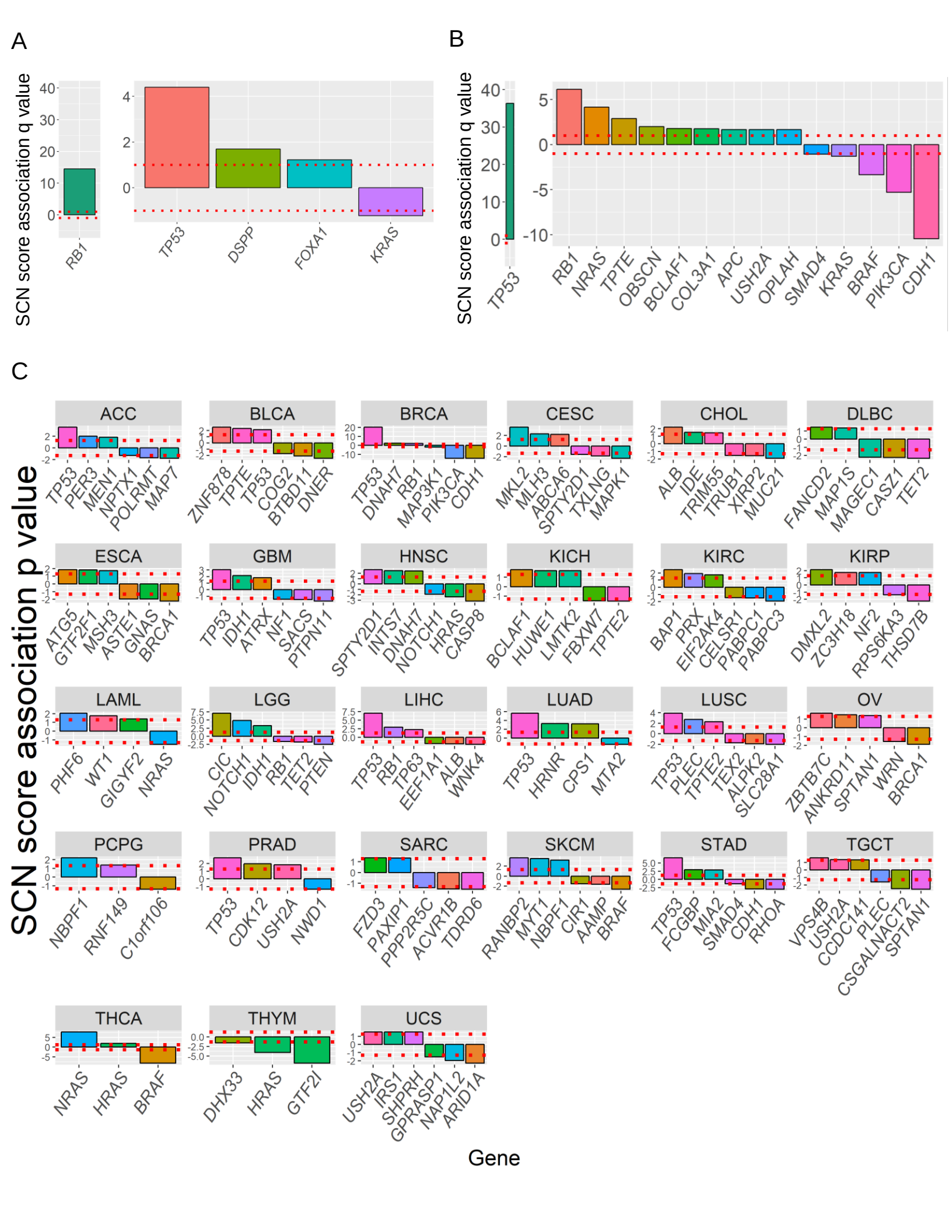

A
SCN score association q value
RB1
TP53
DSPP
FOXA1
KRAS
B
SCN score association q value
TP53
RB1
NRAS
TPTE
OBSCN
BCLAF1
COL3A1
APC
USH2A
OPLAH
SMAD4
KRAS
BRAF
PIK3CA
CDH1
C
SCN score association p value
ACC
TP53 PER3 MEN1 NPTX1 POLRMT MAP7
BLCA
ZNF878 TPTE TP53 COG2 BTBD11 DNER
BRCA
TP53 DNAH7 RB1 MAP3K1 PIK3CA CDH1
CESC
MKL2 MLH3 ABCA6 SPTY2D1 TXLNG MAPK1
CHOL
ALB IDE TRIM55 TRUB1 XIRP2 MUC21
DLBC
FANCD2 MAP1S MAGEC1 CASZ1 TET2
ESCA
ATG5 GTF2F1 MSH3 ASTE1 GNAS BRCA1
GBM
TP53 IDH1 ATRX NF1 SACS PTPN11
HNSC
SPTY2D1 INTS7 DNAH7 NOTCH1 HRAS CASP8
KICH
BCLAF1 HUWE1 LMTK2 FBXW7 TPTE2
KIRC
BAP1 PRX EIF2AK4 CELSR1 PABPC1 PABPC3
KIRP
DMXL2 ZC3H18 NF2 RPS6KA3 THSD7B
LAML
PHF6 WT1 GIGYF2 NRAS
LGG
CIC NOTCH1 IDH1 RB1 TET2 PTEN
LIHC
TP53 RB1 TP63 EEF1A1 ALB WNK4
LUAD
TP53 HRNR CPS1 MTA2
LUSC
TP53 PLEC TPTE2 TEX2 ALPK2 SLC28A1
OV
ZBTB7C ANKRD11 SPTAN1 WRN BRCA1
PCPG
NBPF1 RNF149 C1orf106
PRAD
TP53 CDK12 USH2A NWD1
SARC
FZD3 PAXIP1 PPP2R5C ACVR1B TDRD6
SKCM
RANBP2 MYT1 NBPF1 CIR1 AAMP BRAF
STAD
TP53 FCGBP MIA2 SMAD4 CDH1 RHOA
TGCT
VPS4B USH2A CCDC141 PLEC CSGALNACT2 SPTAN1
THCA
NRAS HRAS BRAF
THYM
DHX33 HRAS GTF2I
UCS
USH2A IRS1 SHPRH GPRASP1 NAP1L2 ARID1A
Gene

**Figure S4. Genetic mutations associated with the SCN phenotype. Related to Figure 4.** (A) Gene mutation association with continuous SCN score in lung (SCLC,LUAD) and prostate tumors (NEPC, CRPC, PRAD). Signed log p values are positive (negative) when associated with the SCN state (non-SCN state). Red line indicates a q value of 0.1. (B) Pan-epithelial cancer gene mutation association with SCN score in primary tumors. Mutational status is the dependent variable, conditioned on neuroendocrine score and tissue type. Dashed red line indicates a q value of 0.1. (C) Gene mutation association by individual cancer type across all TCGA cancers with data available. Dashed red line indicates a nominal p value of 0.05. Cancers not shown have no genes that met this threshold (COADREAD, UCEC, PAAD).

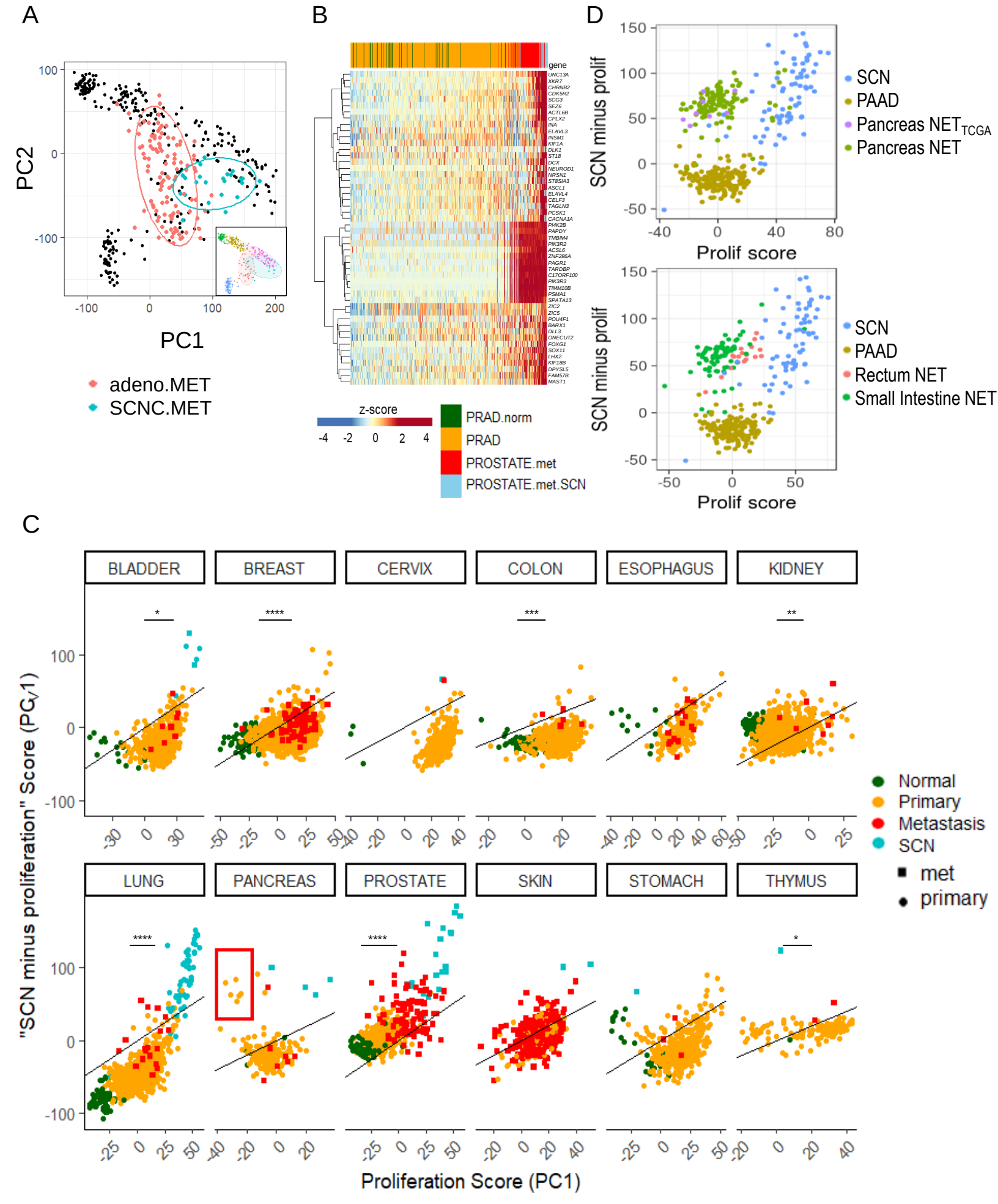

A
PC2
PC1
adeno.MET
SCNC.MET
B
gene
z-score
PRAD.norm
PRAD
PROSTATE.met
PROSTATE.met.SCN
D
SCN minus prolif
Prolif score
SCN
PAAD
Pancreas NET TCGA
Pancreas NET
Rectum NET
Small Intestine NET
C
BLADDER
BREAST
CERVIX
COLON
ESOPHAGUS
KIDNEY
LUNG
PANCREAS
PROSTATE
SKIN
STOMACH
THYMUS
"SCN minus proliferation" Score (PC1)
Proliferation Score (PC1)
Normal
Primary
Metastasis
SCN
met
primary

**Figure S5. Metastatic carcinomas display a convergent SCN trajectory with a neuronal signature. Related to Figure 5.** (A) Pan-cancer metastatic adenocarcinomas and SCNs patient tumors projected onto the SCN signature space defined by lung and prostate tumors. Projection onto the PCA framework of Fig 1A (black dots are Fig 1A samples, also shown in color in the inset). (B) Expression levels of top 50 SCN signature gene loadings (prostate and lung Fig 1A-defined) in prostate tissue and cancer samples. Samples are sorted by the sum z-score across these 50 genes. (C) Separation of proliferation signal and SCN signal. Note some SCN samples (blue) are either primary (circle) or metastatic (square) tumors. This approach highlighted a unique distinction in the pancreatic neuroendocrine tumors, in which both primary and metastatic tumors had equal expression of the neuronal program, but could be distinguished by their expression of the proliferation signature. Here, the pancreatic, cervix, stomach, and thymus blue samples are those annotated with a neuroendocrine (NE)-related term in Robinson et al., 2017. Wilcoxon-Mann-Whitney test p values are shown comparing primary (orange) and metastasis (red). (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Lines inside boxplots represent the $25^{th}$, $50^{th}$, and $75^{th}$ quantiles. Whiskers extend to 1.5 the interquartile range. (D) Additional low grade neuroendocrine tumors from the pancreas, rectum, and small intestine. A different number of genes was used due to differences in data source and processing. Thus, lung and prostate SCN, and TCGA PAAD and TCGA pancreatic NETs shown in panel C are replotted for frame of reference.

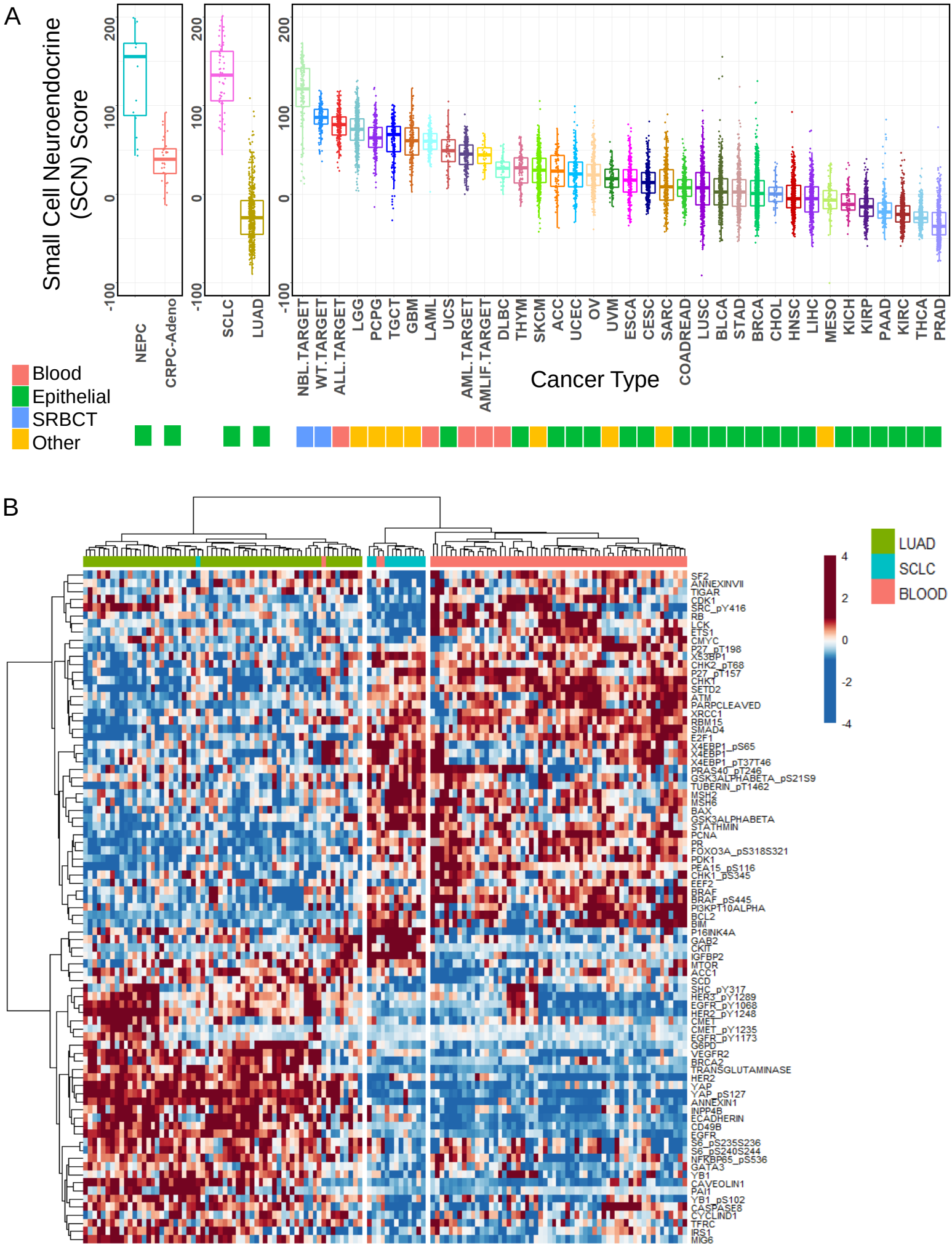

A
Small Cell Neuroendocrine (SCN) Score
200
100
0
-100
NEPC
CRPC-Adeno
SCLC
LUAD
NBL.TARGET
WT.TARGET
ALL.TARGET
LGG
PCPG
TGCT
GBM
LAML
UCS
AML.TARGET
AMLIF.TARGET
DLBC
THYM
SKCM
ACC
UCEC
OV
UVM
ESCA
CESC
SARC
COADREAD
LUSC
BLCA
STAD
BRCA
CHOL
HNSC
LIHC
MESO
KICH
KIRP
PAAD
KIRC
THCA
PRAD
Cancer Type
Blood
Epithelial
SRBCT
Other
B
LUAD
SCLC
BLOOD
4
2
0
-2
-4
SF2
ANNEXINVII
TIGAR
CDK1
SRC_pY416
RB
LCK
ETS1
CMYC
P27_pT198
X53BP1
CHK2_pT68
P27_pT157
CHK1
SETD2
ATM
PARPCLEAVED
XRCC1
RBM15
SMAD4
E2F1
X4EBP1_pS65
X4EBP1
X4EBP1_pT37T46
PRAS40_pT246
GSK3ALPHABETA_pS21S9
TUBERIN_pT1462
MSH2
MSH6
BAX
GSK3ALPHABETA
STATHMIN
PCNA
PR
FOXO3A_pS318S321
PDK1
PEA15_pS116
CHK1_pS345
EEF2
BRAF
BRAF_pS445
PI3KPT10ALPHA
BCL2
BIM
P16INK4A
GAB2
CKIT
IGFBP2
MTOR
ACC1
SCD
SHC_pY317
HER3_pY1289
EGFR_pY1068
HER2_pY1248
CMET
CMET_pY1235
EGFR_pY1173
G6PD
VEGFR2
BRCA2
TRANSGLUTAMINASE
HER2
YAP
YAP_pS127
ANNEXIN1
INPP4B
ECADHERIN
CD49B
EGFR
S6_pS235S236
S6_pS240S244
NFKBP65_pS536
GATA3
YB1
CAVEOLIN1
PAI1
YB1_pS102
CASPASE8
CYCLIND1
TFRC
IRS1
MIG6

**Figure S6. Blood cancers share expression profiles with SCNCs. Related to Figure 6.** (A) SCN score of blood and SRBCT patient tumors by projection onto Figure 1A framework (non-z-score normalized to keep on common scale). Lines inside boxplots represent the $25^{th}$, $50^{th}$, and $75^{th}$ quantiles. Whiskers extend to 1.5 the interquartile range. (B) Heatmap of significantly differentially expressed proteins between lung adeno (LUAD) and lung SCN (SCLC) cell lines (t-test p value < 0.01). Hierarchical clustering is done on measurements for lung adeno, lung SCN, and blood lines. Values were imputed for proteins lacking measurements as described in methods. Cell lines with >25% imputed values for the proteins shown were removed. (TARGET datasets: ALL.TARGET = Acute Lymphoblastic Leukemia; AML.TARGET = Acute Myeloid Leukemia (pediatric); AMLIF.TARGET = Acute Myeloid Leukemia (pediatric) Induction Failure; NBL.TARGET = Neuroblastoma (pediatric); WT.TARGET = Wilms Tumor (pediatric)

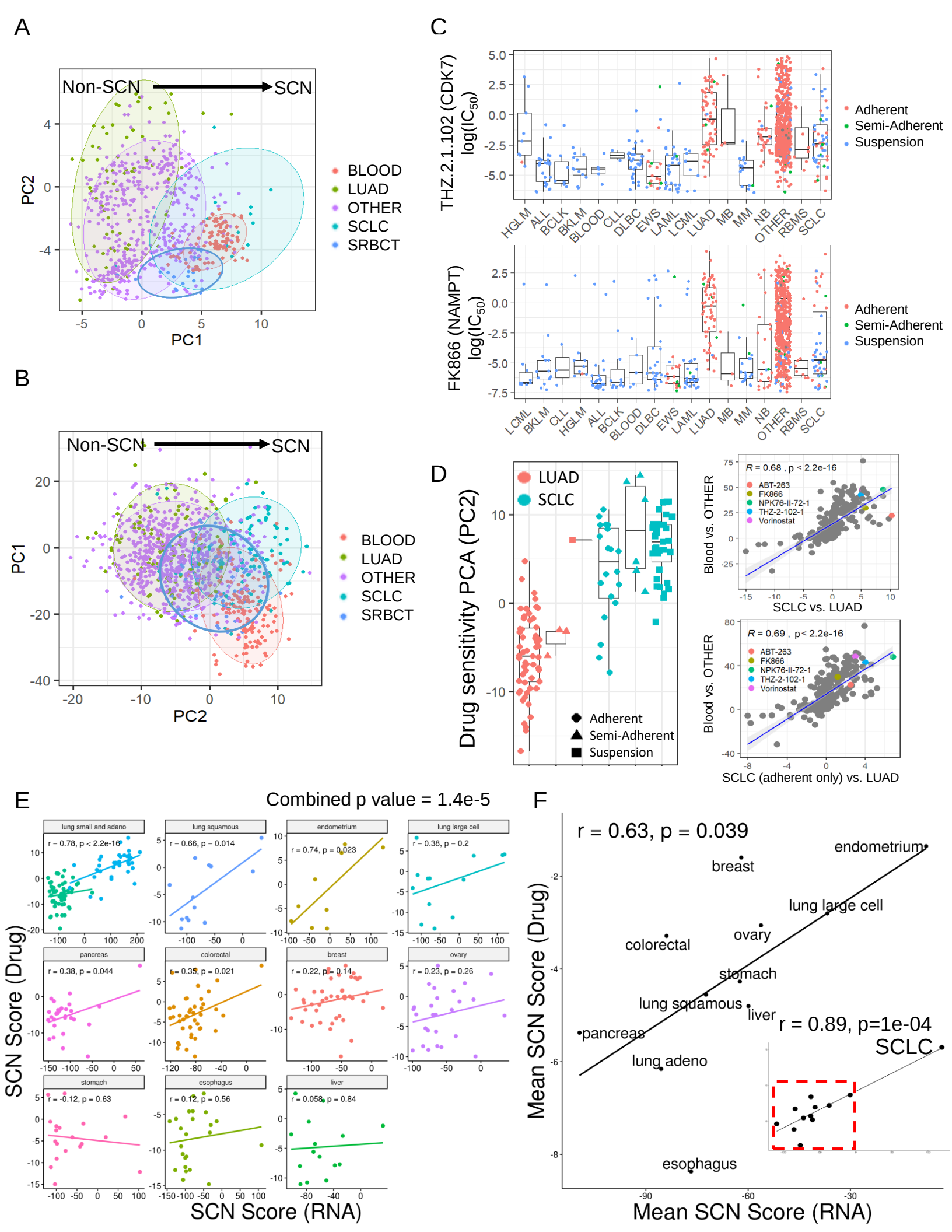

A
Non-SCN
SCN
PC1
PC2
BLOOD
LUAD
OTHER
SCLC
SRBCT
B
Non-SCN
SCN
PC1
PC2
BLOOD
LUAD
OTHER
SCLC
SRBCT
C
THZ.2.1.102 (CDK7) log(IC50)
FK866 (NAMPT) log(IC50)
HGLM ALL BCLK BKLM BLOOD CLL DLBC EWS LAML LCML LUAD MB MM NB OTHER RBMS SCLC
LCML BKLM CLL HGLM ALL BCLK BLOOD DLBC EWS LAML LUAD MB MM NB OTHER RBMS SCLC
Adherent
Semi-Adherent
Suspension
D
Drug sensitivity PCA (PC2)
LUAD
SCLC
Adherent
Semi-Adherent
Suspension
R = 0.68 , p < 2.2e-16
ABT-263
FK866
NPK76-II-72-1
THZ-2-102-1
Vorinostat
Blood vs. OTHER
SCLC vs. LUAD
R = 0.69 , p< 2.2e-16
SCLC (adherent only) vs. LUAD
E
Combined p value = 1.4e-5
lung small and adeno
r = 0.78, p < 2.2e-16
lung squamous
r = 0.66, p = 0.014
endometrium
r = 0.74, p = 0.023
lung large cell
r = 0.38, p = 0.2
pancreas
r = 0.38, p = 0.044
colorectal
r = 0.35, p = 0.021
breast
r = 0.22, p = 0.14
ovary
r = 0.23, p = 0.26
stomach
r = -0.12, p = 0.63
esophagus
r = 0.12, p = 0.56
liver
r = 0.058, p = 0.84
SCN Score (Drug)
SCN Score (RNA)
F
r = 0.63, p = 0.039
endometrium
breast
lung large cell
ovary
colorectal
stomach
lung squamous
liver
pancreas
lung adeno
esophagus
r = 0.89, p=1e-04
SCLC
Mean SCN Score (Drug)
Mean SCN Score (RNA)

**Figure S7. Shared drug sensitivities in SCNC, SRBCTs, and SCN-like epithelial cancer cell lines. Related to Figure 6.** (A) PCA of protein profiles for lung adeno (LUAD) and lung SCN (SCLC) lines, and projection of all other cell lines, with SRBCTs highlighted as a group. SRBCTs include neuroblastoma, medulloblastoma, rhabdomyosarcoma, and Ewing's sarcoma. Ellipses represent 80% confidence regions. (B) PCA of drug sensitivity $IC_{50}$ for LUAD and SCLC lines, and projection of all other cell lines, with SRBCTs highlighted as a group. (C) Individual drug sensitivities for selected drugs with high differential sensitivity (p value < 0.0001) in the SCLC versus LUAD comparison, and blood versus non-blood comparison. Individual blood cancer types are shown, which were all grouped into the BLOOD category in panels A and B. Samples marked BLOOD in this panel are blood cancer lines of unclear subtype. Drug targets shown in parentheses of axes labels. Y-axis is the log $IC_{50}$. Legend defines culture-based growth characteristic. (D) Left: PC2 (from Fig 5C) of drug sensitivity data on lung cell lines with culture-based growth characteristic indicated. Top right: Scatterplot of t-test signed log p values of SCLC (all lines) vs. LUAD (x-axis) and blood vs. other (y-axis) drug sensitivity profile comparisons (Pearson's R = 0.68). Bottom right: As in left panel but restricting SCLC samples to only those that grow adherent (Pearson's R = 0.69). The correlation coefficient of the two x-axes, "SCLC (suspension only) vs. LUAD" and "SCLC (adherent only) vs. LUAD", is R = 0.81. Note: almost all LUAD lines grow adherent (94%). (n=20 adherent SCLC, 6 semi-adherent SCLC, 38 suspension SCLC; 168 blood samples, 638 other (non-blood, non-lung, non-SRBCT), 58 adherent LUAD). Analysis of the shRNA knockdown sensitivity data likewise shows correlation between SCLC and blood knockdown sensitivity, even when only adherent SCLCs are used. The correlation of the blood vs other knockdown sensitivity signature to the SCLC vs LUAD signature is R=0.32 using only adherent SCLCs, R=0.4 using only suspension SCLCs, and R=0.43 using all SCLCs (n=11 adherent, n=2 semi-adherent, n=10 suspension SCLCs). (E) Correlation (Pearson) between drug sensitivity SCN score and RNA expression SCN score for epithelial cancer cell lines. Cell lines with larger RNA and drug scores are more SCN-like. (Combined p value for all epithelial cancers except LUAD and SCLC using weighted-Stouffer's test). The first panel shows the correlation coefficient and p values for both LUAD and SCLC. The individual values are r = −0.14, p = 0.32, and r = −0.51, p = 0.0012, respectively. (F) Mean drug sensitivity SCN score versus mean expression SCN score for epithelial tissue cell lines. Inset, inclusion of the mean from SCLC cell lines (red circle region corresponds to the main graph). (* p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001). EWS: Ewing's Sarcoma (SRBCT); MB: Medulloblastoma (SRBCT); NB Neuroblastoma (SRBCT); RBMS: Rhabdomyosarcoma (SRBCT); ALL: Acute Lymphoblastic Leukemia; BKLM: Burkitt Lymphoma BCLK: B-cell Leukemia ; CLL: Chronic lymphoblastic leukemia ; HGLM Hodgkin's Lymphoma ; MM Myeloma. Lines inside boxplots represent the 25$^{th}$, 50$^{th}$, and 75$^{th}$ quantiles. Whiskers extend to 1.5 the interquartile range.

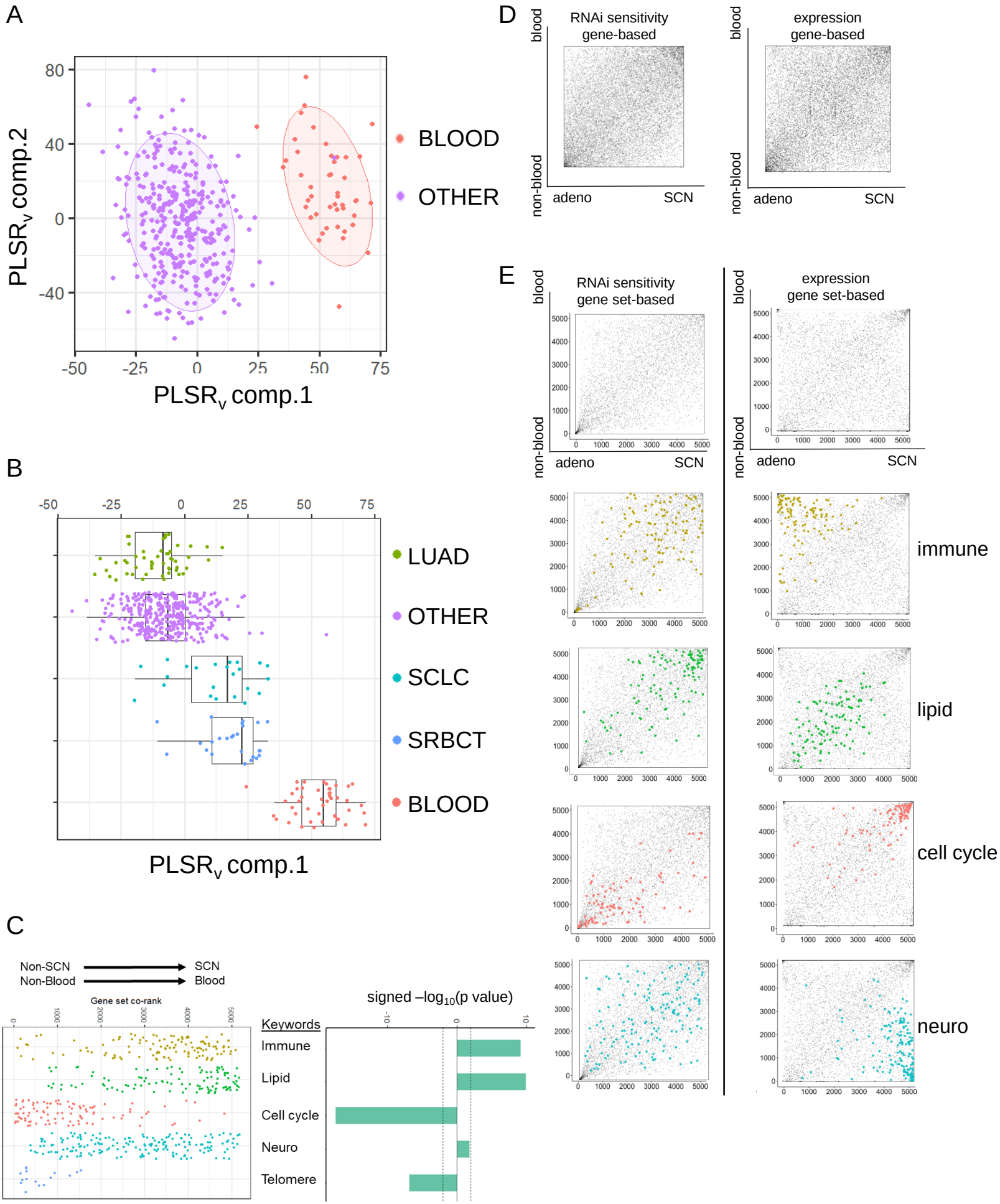

A
80
40
0
-40
-50
-25
0
25
50
75
PLSR$_v$ comp.2
PLSR$_v$ comp.1
BLOOD
OTHER
B
-50
-25
0
25
50
75
LUAD
OTHER
SCLC
SRBCT
BLOOD
PLSR$_v$ comp.1
C
Non-SCN
Non-Blood
SCN
Blood
Gene set co-rank
0
1000
2000
3000
4000
5000
Keywords
Immune
Lipid
Cell cycle
Neuro
Telomere
signed –log$_{10}$(p value)
-10
0
10
D
blood
non-blood
adeno
SCN
RNAi sensitivity
gene-based
expression
gene-based
E
RNAi sensitivity
gene set-based
expression
gene set-based
immune
lipid
cell cycle
neuro

**Figure S8. Validation of shared vulnerabilities based on genome-scale functional RNAi screens. Related to Figure 7.** (A) Varimax-rotated PLSR ($PLSR_V$) of blood versus non-blood cell lines (non-blood = all cell lines except SCLC , LUAD, and SRBCT). Ellipses represent 80% confidence regions. (B) Projection of LUAD, SCLC and SRBCT cell lines onto $PLSR_V$ component 1 of panel A. Lines inside boxplots represent the $25^{th}$, $50^{th}$, and $75^{th}$ quantiles. Whiskers extend to 1.5 the interquartile range. (C) Enrichment or de-enrichment of categories of gene sets, based on co-ranked blood and SCLC RNAi sensitivity gene set signatures. Dotted line in the p value bar plot indicates nominal Kolmogorov-Smirnov p value = 0.01. (D-E) RRHO scatter plots of blood and lung SCN sensitivities by genes (D) and by gene set (E). Gene set RRHO scatter plots are subcategorized and colored by immune, lipid, neuro, and cell cycle gene sets, with all other gene sets colored gray (left: sensitivity based, right: expression based).