

Cell Reports, Volume 33

Supplemental Information

SCLC-CellMiner: A Resource for Small Cell

Lung Cancer Cell Line Genomics and Pharmacology

Based on Genomic Signatures

Camille Tlemsani, Lorinc Pongor, Fathi Elloumi, Luc Girard, Kenneth E. Huffman, Nitin Roper, Sudhir Varma, Augustin Luna, Vinodh N. Rajapakse, Robin Sebastian, Kurt W. Kohn, Julia Krushkal, Mirit I. Aladjem, Beverly A. Teicher, Paul S. Meltzer, William C. Reinhold, John D. Minna, Anish Thomas, and Yves Pommier

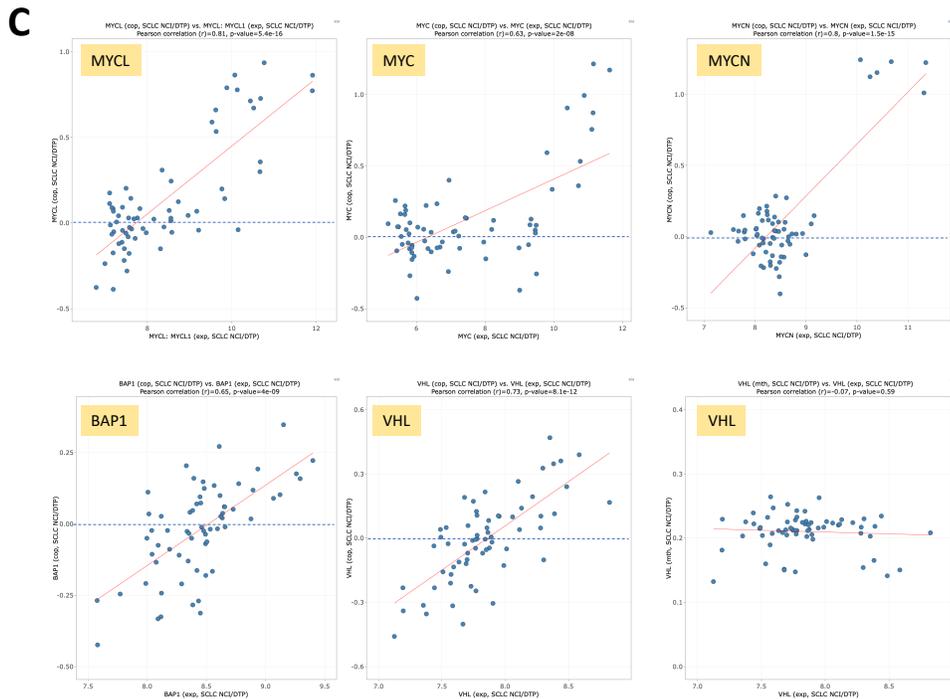
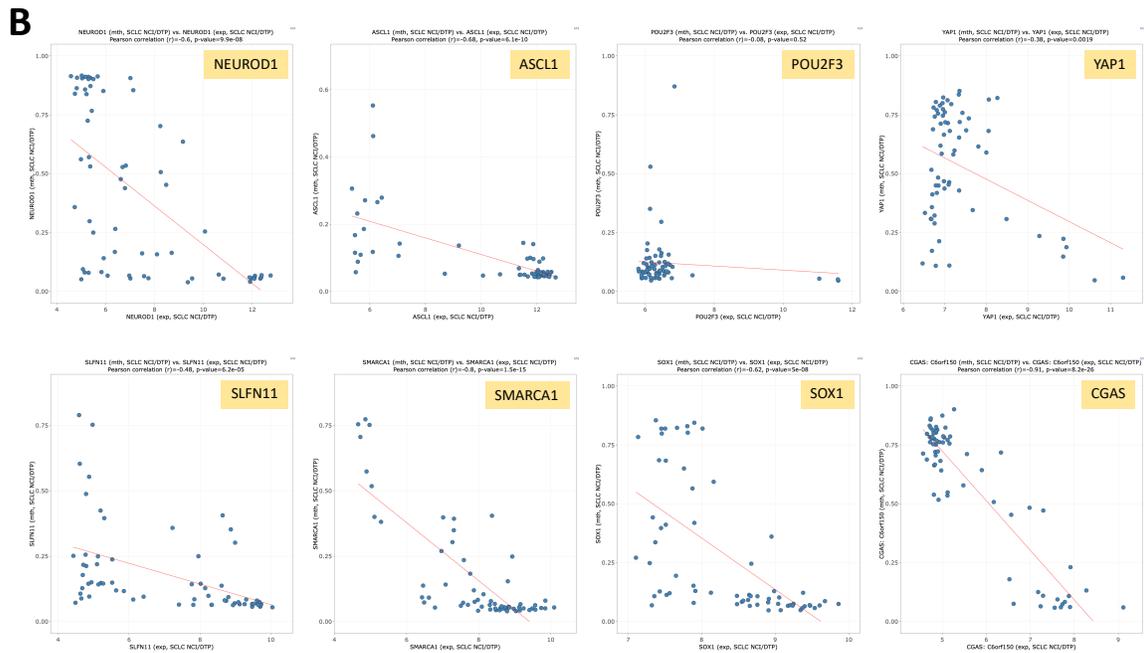
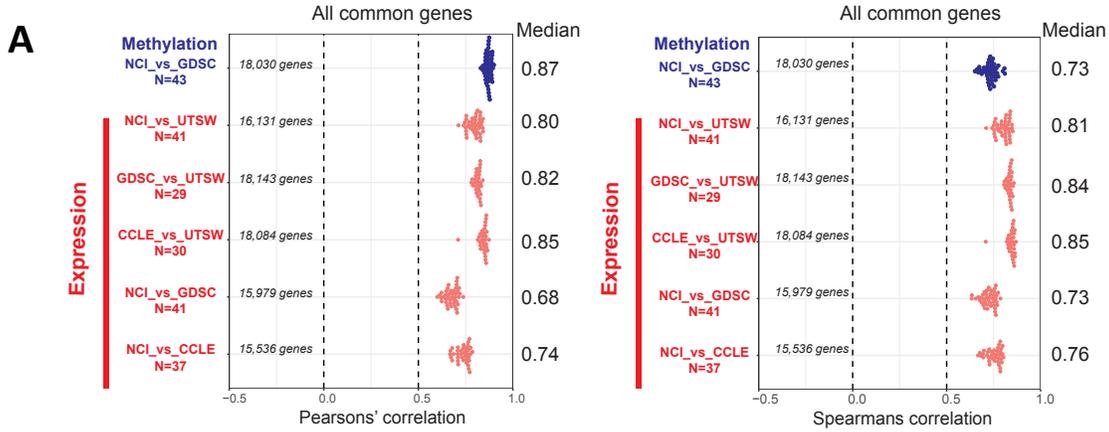


Figure S1, Related to Figure 2: Reproducibility and correlations between data sources by cell line

(A) Pearson's (left) and Spearman's (right) correlation of cell lines using common genes between data sources. Each dot is a cell line. (B) Snapshots from SCLC-CellMiner (<https://discover.nci.nih.gov/SclcCellMinerCDB>) plotting DNA methylation (y-axis) vs. gene expression (x-axis). The Pearson correlations of *NEUROD1*, *ASCL1*, *POU2F3*, *YAP1*, *SLFN11*, *SMARCA1*, *SOX1* and *CGAS* genes in the NCI-SCLC dataset are -0.60, -0.68, -0.08, -0.38, -0.48, -0.80, -0.62 and -0.91, respectively. (C) Snapshot from SCLC-CellMiner plotting gene copy number (y-axis) vs. expression (x-axis). Pearson's correlations are 0.81, 0.63, 0.80, 0.65 and 0.73, respectively. The bottom right plot shows the lack of correlation between promoter methylation (y-axis) and expression (x-axis) for *VHL*.

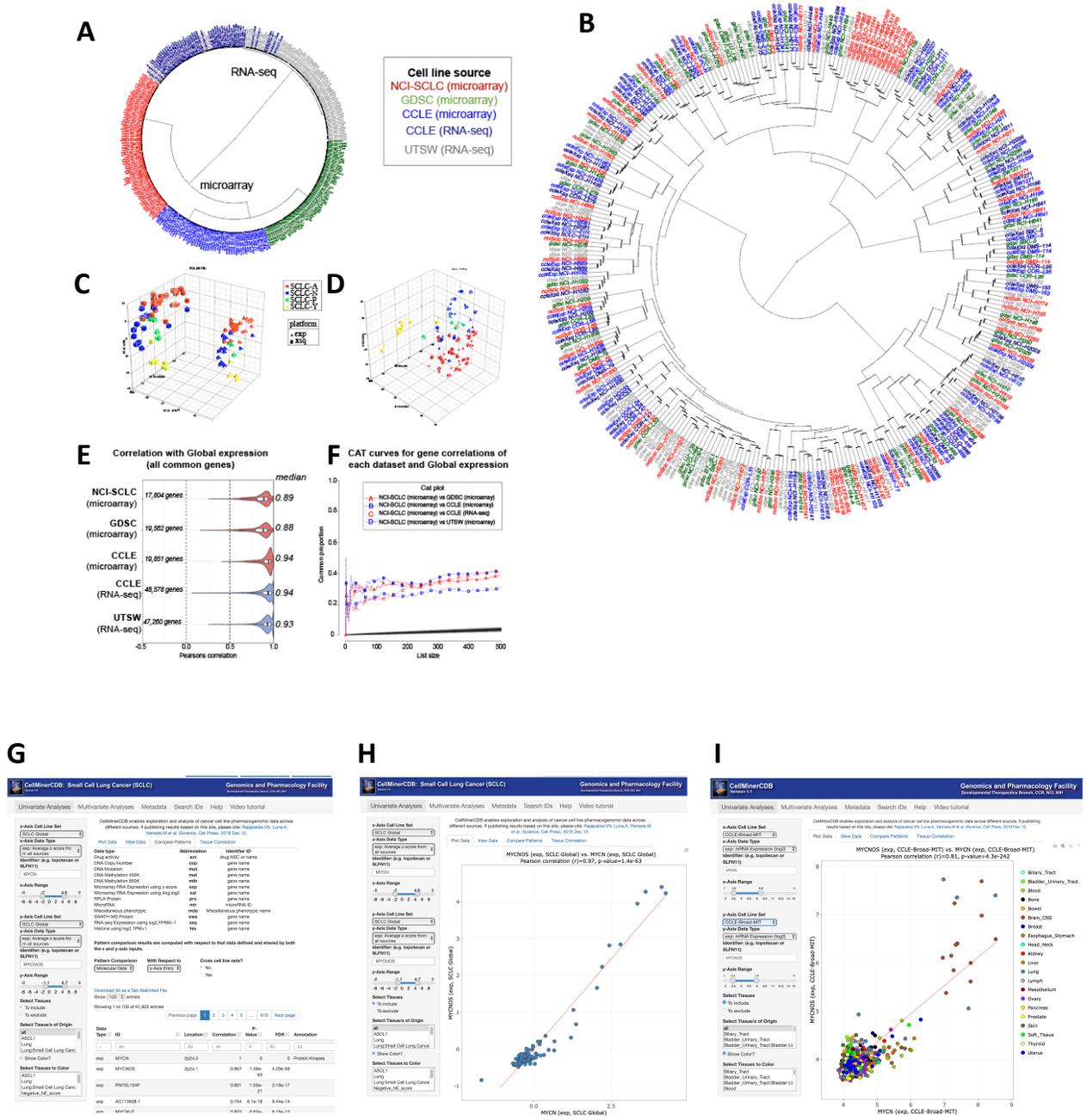


Figure S2, Related to Figure 3: Normalized expression data validation and reproducibility of the “Global” dataset (116 SCLC cell lines) using Z-score

Circos plot visualization of hierarchical clustering distance of cell lines using (A) unnormalized data and (B) Z-score normalized data. Cell lines clustered based on data origin (RNA-seq or microarray) and database source before normalization. After Z-score normalization, common cell lines in different datasets clustered together. Principle component analysis (PCA) plots (C) before and after (D) Z-score normalization between microarray and RNA-seq data in the CCLE dataset. Each point represents a cell line. Rectangles are the cell lines according to RNA-seq value and triangles the cell lines according to microarray expression values. Cell lines are color-coded according to the NAPY classification: SCLC-A cell lines in red, SCLC-N cell lines in blue, SCLC-P in green and SCLC-Y in yellow. The x-axis represents the first principal component, the y-axis the second component and the z-axis the third principal component. (E) Reproducibility between the *SCLC-Global* and the individual datasets. Pearson correlations between the *SCLC-Global* and the indicated individual data sources for matched cell lines were 0.89,

0.88, 0.94, 0.94 and 0.93 for GLOBAL/NCI, GLOBAL/GDSC, GLOBAL/CCLE (microarray), GLOBAL/CCLE (RNA-seq) and GLOBAL/UTSW (RNA-seq) expressions, respectively. **(F)** CAT-plot showing that genes with high correlation ranking in one comparison have similarly high correlation ranking in other comparisons. **(G-I)** SCLC-CellMiner examples for “Compare Patterns”, correlation between MYCN and MYCNOS expression. Each panel is a snapshot from SCLC-CellMiner and CellMinerCDB. **(G)** SCLC-CellMinerCDB snapshot of correlation between *MYC* expression in all SCLC cell lines and all the other genes can be easily found through the “Compare Patterns” function in the “Univariate Analysis” section (<https://discover.nci.nih.gov/ScLcCellMinerCDB>). The example shows that *MYCN* expression is highly correlated with *MYCN* antisense *MYCNOS* expression (Pearson correlation = 0.967, p-value = 1.39×10^{-63}). **(H)** SCLC-CellMiner Snapshot showing *MYCNOS* expression (y-axis) vs. *MYCN* expression (x-axis) in the SCLC-Global cell lines. **(I)** Findings in SCLC cell lines can be readily compared to other cancer cell lines using CellMinerCDB (<http://discover.nci.nih.gov/cellminerfdb>). Shown is the high correlation between *MYCN* and *MYCNOS* expression is also present across additional cancer subtypes.

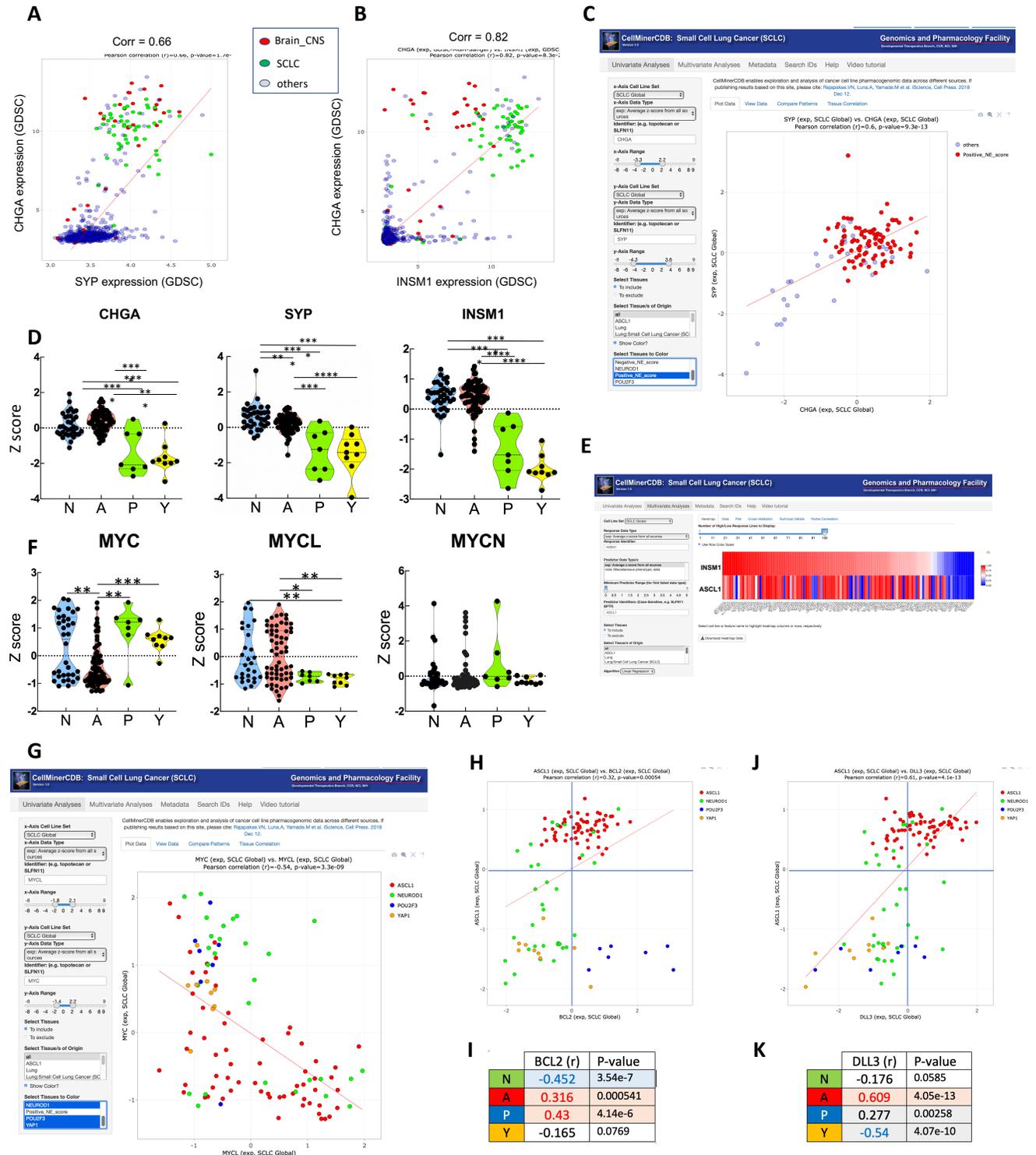
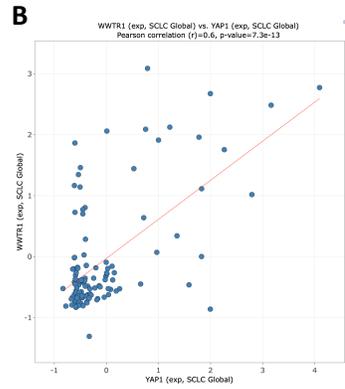
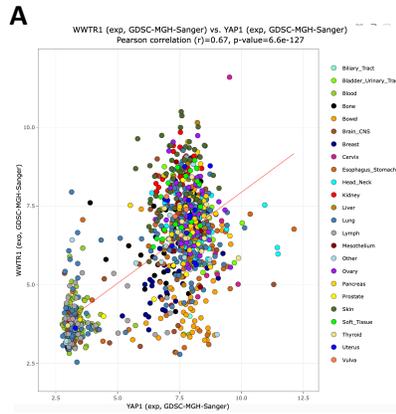


Figure S3, Related to Figure 4: Expression characteristics of key markers involved in SCLC carcinogenesis

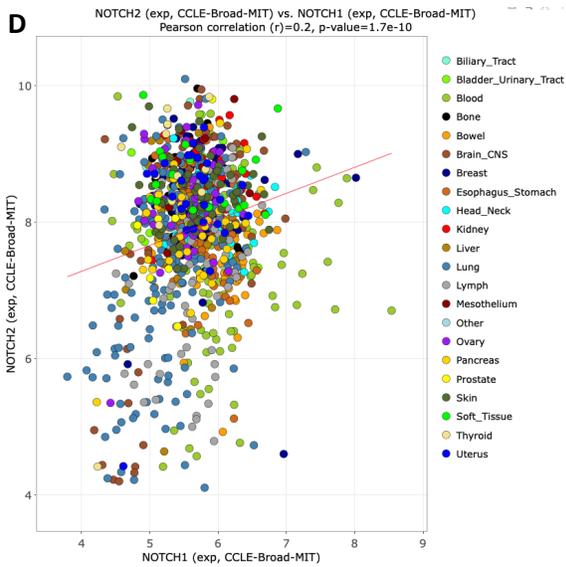
Chromogranin (CHGA), synaptophysin (SYP) and insulinoma-associated protein 1 (INSM1) are neuroendocrine markers used in routine practice to diagnose SCLC and NE tumors. The images for each panel are snapshots from the SCLC-CellMiner (<https://discover.nci.nih.gov/ScleCellMinerCDB>) and CellMiner (<http://discover.nci.nih.gov/cellmineradb>) websites. Panels (A) and (B) highlight that *CHGA*, *SYP* and *INSM1* expressions are highly correlated with Pearson correlations at 0.66 and 0.82, respectively. High expression of these

three neuroendocrine markers is mainly found in SCLC (green points) and brain tumor (red points) cell lines. Using “Univariate Analysis” function in SCLC-CellMiner website, the plot in panel (C), showing SYP expression (y-axis) and CHGA expression (x-axis), highlights that high expression of both neuroendocrine markers is found only in the cell lines with a high NE score. Similarly, panel (D) shows the expression of *CHGA* (left), *SYP* (middle) and *INSM1* (right) according the NAPY classification. The expression of the three neuroendocrine markers is higher in the SCLC-N and -A cell lines (blue and red, respectively) compared to the SCLC-P and -Y cell lines (green and yellow, respectively). Each point represents a cell line. (E) Snapshot of the “Multivariate Analyses” tool of SCLC-CellMiner using the Linear Regression option and showing that *ASCL1* expression is highly associated with *INSM1* expression. (F) Left: *MYC* is constantly highly expressed in the SCLC-P and -Y cell lines while its expression is bimodal in the SCLC-N cell lines and generally low in the SCLC-A cell lines. Middle: *MYCL* expression has an opposite behavior with low expression in all SCLC-P and -Y cell lines and a bimodal distribution in the SCLC-N and -A cell lines. Right: *MYCN* is less frequently expressed in all cell lines, especially the SCLC-Y subset. (G). Snapshot of *MYC* (y-axis) and *MYCL* expression (x-axis) in the 116 SCLC cell lines of SCLC-Global showing their mutually exclusive expression (correlation = - 0.54). High *MYC* expression is mainly found in the SCLC-P and -Y cell lines (blue and yellow points, respectively) while high *MYCL* expression is mainly in the NE SCLC-A and -N cell lines (red and green points, respectively). Correlations between *ASCL1* expression (y-axis) with *BCL2* (H) and *DLL3* (J) expression (x-axis) in the 116 cell lines of *SCLC-Global*. High *BCL2* and *DLL3* expressions are mainly found in the SCLC-A cell lines (red). Of note, SCLC-P cell lines also have high *BCL2* expression. Pearson correlations of each NAPY gene (*NEUROD1*, *ASCL1*, *POU2F3* and *YAP1*) vs *BLCL2* (I) and *DLL3* (K). Values in red represent significantly positive Pearson correlations and values in blue significantly negative correlations.



C

	N	A	P	Y
NOTCH1	-0.002	-0.49 (1.84e-8)	0.45 (4.62e-7)	0.39 (1.68e-5)
NOTCH2	-0.132	-0.42 (3.07e-6)	0.23 (1.49e-2)	0.56 (6.96e-11)
NOTCH3	-0.173	-0.40 (8.55e-6)	0.44 (8.96e-7)	0.62 (2e-13)
NOTCH4	-0.002	0.25 (0.00784)	-0.169	-0.091
REST	0.221 (0.0171)	-0.55 (5.48e-10)	0.51 (8.08e-7)	0.76 (2.05e-19)



E

Tissue of origin	Cell lines with complete observations	Pearson correlation	P-value
All	All	All	All
ALL	1036	0.2	1.7e-10
Brain_CNS	86	0.58	5e-9
Lung	186	0.39	2.6e-8
Lymph	64	0.36	0.0038
Thyroid	12	0.64	0.026
Pancreas	44	0.32	0.037
Esophagus_Stomach	64	-0.25	0.044
Uterus	27	0.32	0.11
Prostate	8	0.58	0.14
Mesothelium	11	0.45	0.16
Ovary	52	0.19	0.17
Biliary_Tract	8	-0.45	0.27
Bowel	62	-0.11	0.39
Bladder_Urinary_Tract	27	-0.15	0.46
Soft_Tissue	20	-0.12	0.61
Blood	116	-0.044	0.64
Bone	29	-0.086	0.66
Breast	59	-0.035	0.79
Head_Neck	34	-0.045	0.8
Kidney	36	0.04	0.82
Liver	28	0.03	0.88
Skin	62	0.016	0.9

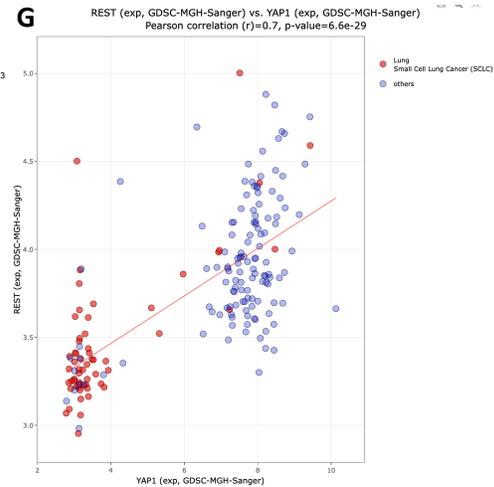
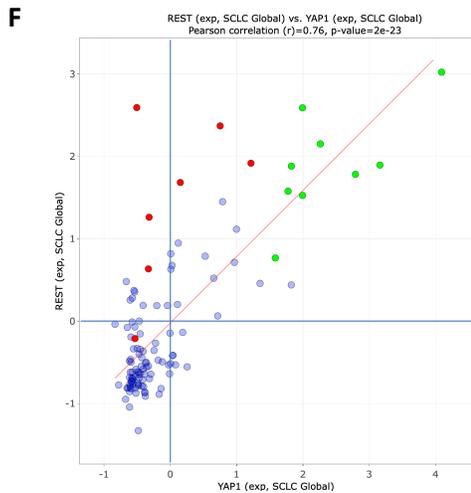


Figure S4, Related to Figure 5: *YAP1* and *NOTCH* pathway co-expression

Snapshots (<http://discover.nci.nih.gov/cellminerfdb>; <https://discover.nci.nih.gov/SclcCellMinerCDB>) showing that *YAP1* expression is highly correlated with the expression of its heterodimeric partner TAZ (encoded by the *WWTR1/TAZ* gene) across the (A) 986 cell lines of the GDSC and among the (B) 116 SCLC cell lines of SCLC-Global. (C) Pearson correlations between the NOTCH/REST pathway gene expression and the NPY genes (*NEUROD1*, *ASCL1*, *POU2F3* and *YAP1*). Expression of the NOTCH genes (except *NOTCH4*) is negatively correlated with *ASCL1* expression and positively correlated with *POU2F3* and *YAP1* expressions. (D) Snapshot from CellMinerCDB (<http://discover.nci.nih.gov/cellminerfdb>) showing correlated expression of *NOTCH2* (y-axis) and *NOTCH1* (x-axis) across histological subtypes in the CCLE dataset. (E) Snapshot of the tabular output from CellMinerCDB showing Pearson correlations between *NOTCH2* and *NOTCH3* expressions across cell lines from the listed histological subtypes. (F) Snapshot from SCLC-CellMiner showing *REST* (y-axis) vs. *YAP1* (x-axis) expression in the 116 SCLC cell lines of SCLC-Global (<http://discover.nci.nih.gov/cellminerfdb>). (G) Same as A except for both subtypes of lung cancers (SCLC + NSCLC) in the GDSC dataset (<https://discover.nci.nih.gov/SclcCellMinerCDB>). The plots highlight that non-NE cell lines have a high *REST* expression (green and red point in panel (F)) and that high *YAP1* expressing lung cancer cell lines (G) also have high expression of *REST*.

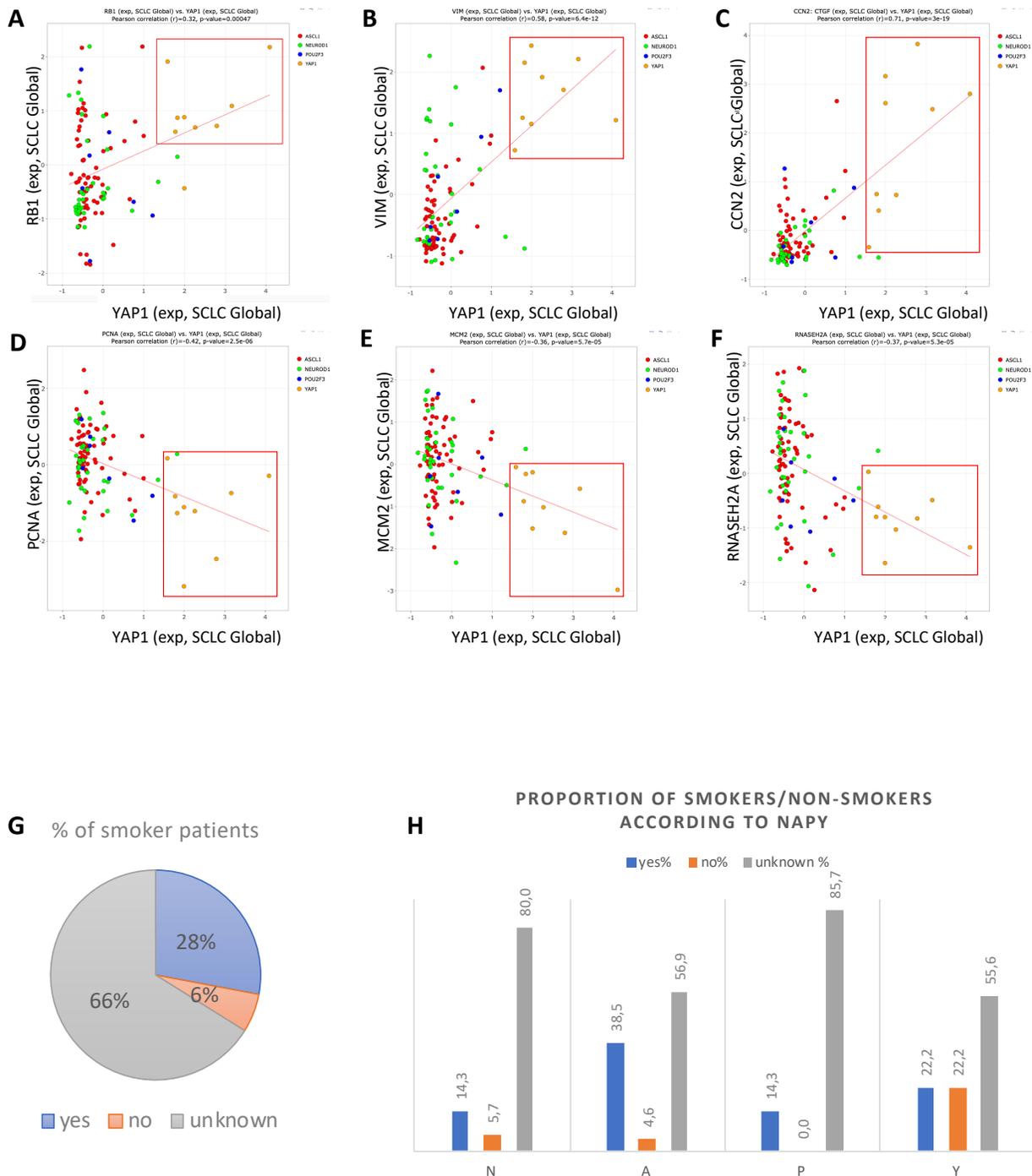


Figure S5, Related to Figure 5: Examples of correlations between *YAP1* expression and expression of key cancer genes and distribution of smoking status

Snapshots (<https://discover.nci.nih.gov/ScicCellMinerCDB>) of *YAP1* expression (x-axis) vs. the indicated cancer genes (y-axis) in the 116 cell lines of SCLC-Global. Each point represents a cell line (red: SCLC-A, green: SCLC-N, blue: SCLC-P and yellow: SCLC-Y). Panels (A), (B) and (C) show highly positive Pearson correlations between *YAP1* and *RB1*, *VIM* and *CCN2*. Panels (D), (E) and (F) show highly negative Pearson correlations between *YAP1* and the replication-associated genes *PCNA*, *MCM2* and *RNASEH2A*. The red rectangles encompass the SCLC-Y

cell lines. **(G)** Global distribution of the cell lines based on patient information (smoking status). **(H)** Distribution of the cell lines across the NAPY subgroups based on smoking status.

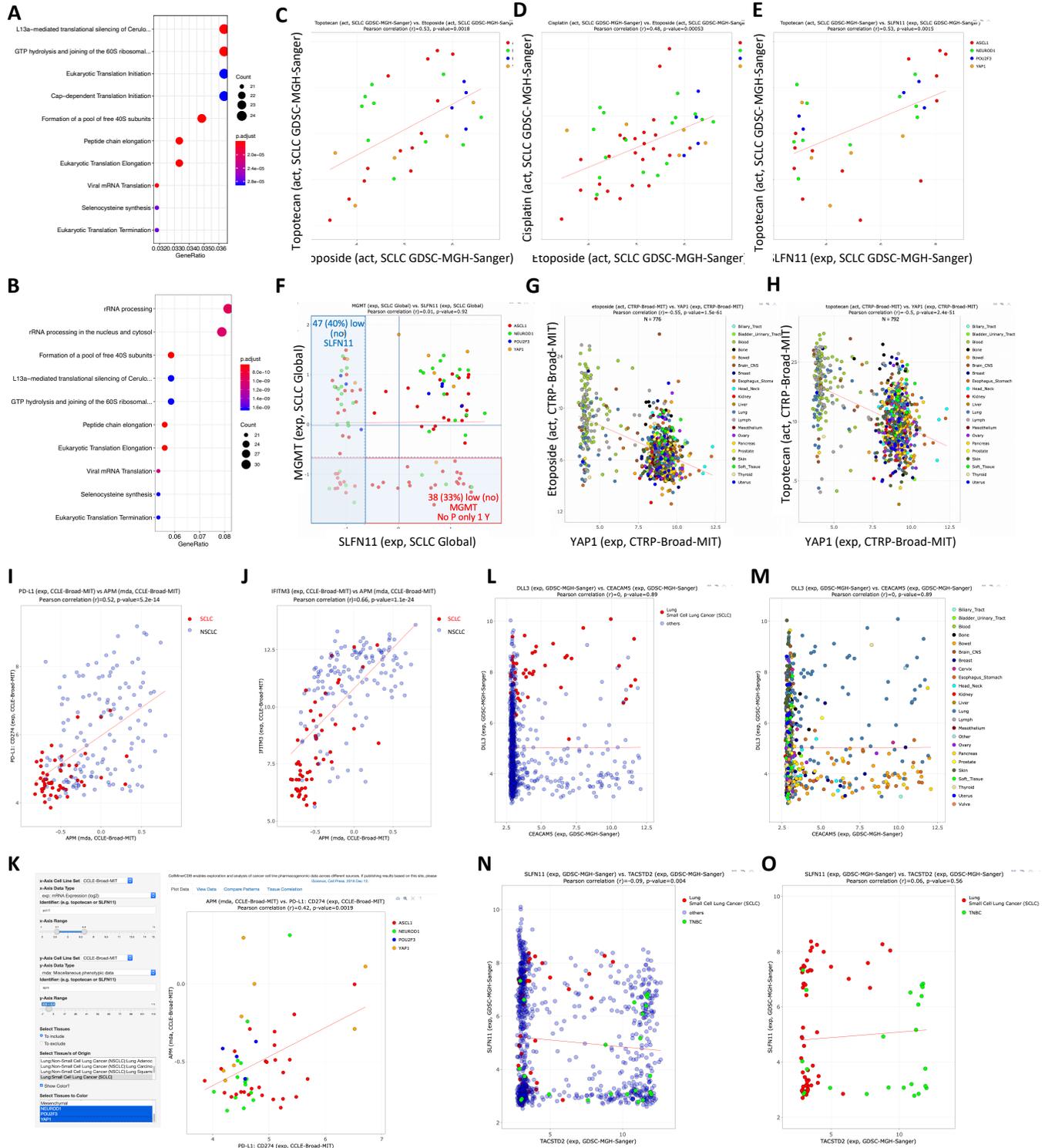


Figure S6, Related to Figure 6: Drug sensitivity of the SCLC cell lines

(A-B) Pathway analyses for the differentially expressed genes between sensitive and non-sensitive cell lines according the heatmap analysis shown in Figure 6A. (A) Pathway analysis when considering all genes. (B) Pathway analysis when considering the genes overexpressed in the most sensitive cell lines group according to Figure 6A.

(C) Correlation between the activity of etoposide (-log IC50) (x-axis) and topotecan (y-axis) across the SCLC cell lines. (D) Correlation between the activity of etoposide (-log IC50) (x-axis) and cisplatin (y-axis). (E) Correlation between *SLFN11* expression (x-axis) and the activity of topotecan (y-axis) in the SCLC cell lines. Each point represents a cell line (red: SCLC-A cell lines, green: SCLC-N cell lines, blue: SCLC-P cell lines and orange: SCLC-Y cell lines). (F) Broad range of bimodal distribution for *SLFN11* and *MGMT* expression in the 116 cell lines of *SCLC-Global*. 40% of the cell lines (47/116) have a low expression of *SLFN11* and 33% (38/116) low *MGMT* expression. Among them, there is no SCLC-P cell line and only one SCLC-Y cell line. (G-H) *YAP1* expression (x-axis) is negatively correlated with the activity of etoposide (log -IC50) (G) and negatively correlated with the activity of topotecan (H) across different histological subtypes (<http://discover.nci.nih.gov/cellminerfdb>). (I-K) We used a list of the 18 genes constituting the APM signature included as metadata in the CellMiner websites (<https://discover.nci.nih.gov/CellMinerCDB> and <https://discover.nci.nih.gov/SclcCellMinerCDB>). (I-J) Correlation between the APM score (x-axis, source: <https://elifesciences.org/articles/49020>) and the expression of *PDL1* (I) and *IFTIM3* (J). Each point represents a cell line (red: SCLC and blue: NSCLC cell lines). Note the lower APM score for the SCLC compared to the NSCLC cell lines. (K) *PD-L1* expression (x-axis) vs. APM score (y-axis) in the 116 cell lines of SCLC-Global. Note that only few cell lines have high APM score. Among the cell lines with highest APM score, most are SCLC-Y. Each point represents a cell line (red: SCLC-A, green: SCLC-N, blue: SCLC-P and orange: SCLC-Y cell lines). (L-M) Snapshot of CellMinerCDB (<https://discover.nci.nih.gov/cellminerfdb>) showing *CEACAM5* expression (x-axis) vs. *DLL3* expression (y-axis) in the 986 cell lines of the GDSC dataset. Each point represents a cell line. (L) SCLC are in red and the other cell lines in blue. Most SCLC cell lines have high *DLL3* expression but only a subset has high expression of *CEACAM5*. (M) Cell lines are represented according to their histological subtype. SCLC cell lines (light blue) have both high *DLL3* and *CEACAM5* expression. (N) Snapshot of CellMinerCDB (<https://discover.nci.nih.gov/cellminerfdb>) showing the expression of *TACSTD2* (x-axis) vs. *SLFN11* (y-axis) in the 986 cell lines of the GDSC dataset and in the subset of 59 SCLC and 24 triple-negative breast cancer (TNBC) cell lines of the GDSC dataset (O). Each point represents a cell line (red: SCLC, green: TNBC and blue: other cell lines).

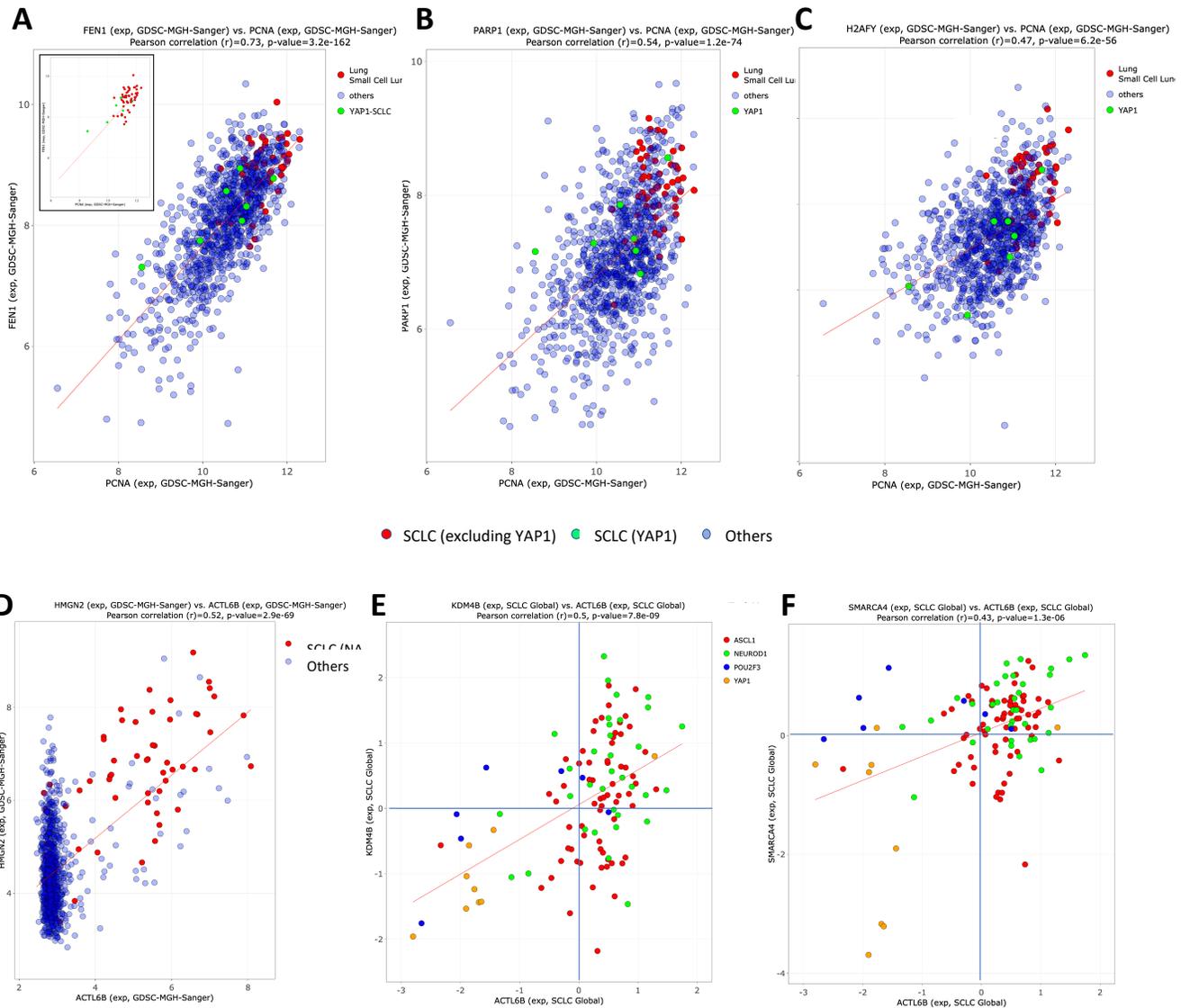


Figure S7, Related to Figure 3: Overexpression of selected DNA replication and epigenetic regulatory genes in SCLC cell lines

Snapshot of CellMinerCDB (<https://discover.nci.nih.gov/cellminercdb>) showing the expression of *PCNA* (x-axis) vs *FEN1* (A), *PARP1* (B) and *H2AFY* (C) (y-axis) in all histological subtypes of cell lines from GDSC dataset. Each point represents a cell line (red: SCLC, green: SCLC-Y cell lines and blue: other cell line subtypes). Note that DNA replication genes are highly expressed in SCLC cell lines except for the SCLC-Y cell lines. (D) Snapshot of CellMinerCDB (<https://discover.nci.nih.gov/cellminercdb>) showing high expression of *ACTL6B* (x-axis) and *HMGN2* (y-axis) in most of the 59 SCLC cell lines (green) and 86 brain cell lines (red) of the GDSC. (E-F) Snapshot of SCLC-CellMiner (<https://discover.nci.nih.gov/ScLcCellMinerCDB>) showing co-expression of *ACTL6B* (x-axis), *KDM4B* (E) and *SMARCA4* (F) (y-axis) among the 116 cell lines of *SCLC-Global* (green: SCLC-N, red: SCLC-A, blue: SCLC-P and orange: SCLC-Y cell lines). Note that among the SCLC cell lines, most of the cell lines expressing low *ACTL6B* are non-NE (SCLC-P and SCLC-Y).