SCLC-TumorMiner: A Directly Accessible Genomics Resource for Precision Oncology – Big Data for Small Cells

Fathi Elloumi[1,*], Anjali Dhall[1,*], Daiki Taniyama[1], Augustin Luna[1,2], Yasuhiro Arakawa[1,3], Sudhir Varma[1,4], Jeffrey Wang[1], Anisha Tehim[1,5], Mark Raffeld [6], Kenneth Aldape[6], Christophe Redon[1], Roshan Shresta[1], William Reinhold[1], Mirit Aladjem[1], Jaydira Del Rivero[1], Nitin Roper[1], Anish Thomas[1] and Yves Pommier[1,#]

[1] Laboratory of Molecular Pharmacology and Developmental Therapeutics Branch, Center for Cancer Research, NCI, NIH
[2] NLM, NIH
[3] Department of Clinical Pharmacology and Therapeutics, The Jikei University School of Medicine, Tokyo, Japan
[4] HiThru Analytics LLC
[5] Cornell University, NY
[6] Laboratory of Pathology, Center for Cancer Research, NCI, NIH

[*] Equal contributions

To whom correspondence should be addressed: pommier@nih.gov; fathi.Elloumi@nih.gov

SUMMARY

RNAseq and DNAseq are fast, cost-effective and quantitative methods to dissect cancer cells. However, for each sample, they generate thousands of data points, and comparing patient samples multiplies the complexity. To handle these difficulties, we developed TumorMiner, a web-based tool for clinicians and basic researchers. Here we present our analyses and website for Small Cell Lung Cancer (SCLC). SCLC_TumorMiner includes 235 samples from untreated and relapse patients across the NCI, the University of Rochester, Tongji University and the University of Cologne (https://discover.nci.nih.gov/SclcTumorMinerCDB/). SCLC_TumorMiner allows the molecular classification of tumors based on the canonical NAPY classification and NMF, genomic network analyses exemplified by the Myc and Notch pathways, and the identification of risk-factors and predictive gene expression biomarkers for cell surface and intracellular targets such as DLL3, TROP2, SEZ6, CEA, TRPM5, SLFN11 and proapoptotic and multidrug resistance genes. The architecture of TumorMiner allows its extension to proteomic databases, and to other cancers and institutions worldwide to elucidate cancer pathways, achieve precision medicine and serve as a medical assistant software.

INTRODUCTION

Small cell lung cancers are rapidly proliferative tumors often diagnosed with metastases (Extensive-stage; ES-SCLC). SCLC accounts for approximately 13%–15% of all lung cancer cases and remains one of the most lethal malignancies. Survival rate is less than one year for most patients, and 95% of patients die within 5 years despite initial response to cisplatin-etoposide combination therapy. Upon relapse, tumors evolve and become resistant to chemotherapy reflecting profound changes in their biology and genomic network.

In addition to mutations of the tumor suppressors TP53 and RB1 in 80% of SCLC, selective activation of transcription factors characterizes the disease. SCLCs have been divided into two groups. 70-80% are neuroendocrine (NE) driven by two pioneer neurogenic transcription factors, ASCL1 (Acheate-scute homolog 1) and NEUROD1 (Neurogenic differentiation factor 1). The remaining non-NE SCLCs are driven by two other transcription factors POU2F3 (POU class 2 homeobox 3) and YAP1 (Yes-associated protein 1). Consequently, SCLCs have been classified in 4 groups: SCLC-A, SCLC-N, SCLC-P and SCLC-Y [1]. Additionally, an Inflamed/Immune subset has been identified as SCLC-I [2]. Multiple genomic studies have been conducted in the past 10 years, confirming the NAPY and NAPI classifications in untreated patients. The common histopathological classification of SCLC is presently based on immunohistochemistry for NE markers: synaptophysin (encoded by the *SYP* gene), chromogranin (encoded by *CHGA*) and neural cell adhesion molecule (encoded by *NCAM1*). Immunohistochemistry of ASCL1, NEUROD1, POU2F3 and YAP1 [3] is not routinely performed.

Immune checkpoint (PD-L1) inhibitors have recently been added to chemotherapy as standard of care for 1st line treatment of ES-SCLC. Atezolizumab (Tecentrip®) was approved in 2019 in combination with carboplatin and etoposide [4], and durvalumab (Imfinzi®) in 2020 in combination with etoposide and either carboplatin or cisplatin [5]. However, immune checkpoint inhibitors only provide a modest overall survival benefit of 2-3 months [5,6].

The most significant therapeutic recent breakthrough is Tarlatamab®, a bispecific T-cell engager (BiTE®) targeting the cell surface glycoprotein Notch ligand DLL3 (Delta-like ligand 3), a transcriptional target of ASCL1 [7]. Tarlatamab was granted FDA approval in May 2024 for ES-SCLC non-responsive to platinum-based therapy or at relapse [8]. However, Tarlatamab is only active in 50% of patients who need to be hospitalized to control treatment-related immune hyper-reactions, and it is prescribed without mandatory evaluation of DLL3 expression.

The substantial benefit provided by Tarlatamab and immune checkpoint inhibitors is fueling the development of novel therapeutic strategies targeting not only DLL3 [9], but also other proteins expressed at the surface of SCLC cells [10]. Two main approaches are actively pursed beyond T-cell engagers, Chimeric antigen receptor T-cell (CAR-T) therapies such as AMG119 [11] and antibody drug conjugates (ADCs). The efficacy of ADCs is not only dependent on the expression of their cell surface target(s) but also on the type of cytotoxic payload delivered to tumor cells. The three main classes of payloads are tubulin inhibitors, as in Rova-T [12], topoisomerase I (TOP1) inhibitors, as in Enhertu (Trastuzumab deruxtecan) and Trodelvy (Sacizutumab govitecan), and DNA binders. Despite their potency, tubulin inhibitors have two main limitations: they are substrates for drug efflux transporters, and they are neurotoxic. The advantage of TOP1 inhibitors is their selectivity for replicating cells and the availability of a testable biomarker, Schlafen 11 (SLFN11) [13], which has a wide range of expression in SCLC cells [7] and tumors [14] and acts as a lethal factor by irreversible blocking stressed replication forks [15,16].

RNA sequencing (RNA-seq) is a well-established, sensitive and routinely used method. However, even when only applied to protein-coding genes, it generates large datasets, which are

difficult to analyze instantaneously in a clinical setting. To address this issue, we developed the SCLC_TumorMiner database integrating 235 SCLC patient samples (154 untreated and 81 relapse) from our NCI clinic and the published datasets of the University of Cologne (UoC) [17], Tongji University (TU) [18], and the University of Rochester (RU) [19] (Table 1, Table S1). The interface of TumorMiner is an extension of on our widely used SCLC_CellMinerCDB website including the genomics and drug responses of 118 SCLC cell lines from multiple open sources (https://discover.nci.nih.gov/rsconnect/SclcCellMinerCDB) [7]. Thus SCLC_TumorMiner and SCLC_CellMinerCDB can be used to compare tumor samples with cell lines.

Here we demonstrate that SCLC_TumorMiner allows instantaneous: 1/ genome-wide analyses of the transcriptome and mutations of patient samples; 2/ Pathway analyses based on the hypothesis that genes functioning in a given genetic network have cross-correlated expression; 3/ Patient sample classification using both NAPY and/or NMF (non-negative matrix factorization); and 4/ Identification of prognosis and predictive biomarkers for targeted therapies.

RESULTS

Introduction to SCLC_TumorMiner

The SCLC_TumorMiner website is built on the architecture of our cancer cell line CellMiner cross-database websites (https://discover.nci.nih.gov/). Thus, users familiar with SCLC_CellMinerCDB (https://discover.nci.nih.gov/SclcCellMinerCDB/) will have no difficulty navigating back and forth between the tumor samples and cell line databases [7].

Figure 1 (https://discover.nci.nih.gov/SCLC_TumorMinerCDB/) introduces the website with a snapshot of the default page display. The upper part provides tabs for each of the website tools and applications. They are numbered 1-10 for convenience on the figure. The "2 variable Analyses" tab (numbered "1" in Figure 1A) allows the display of different data types or parameters on each of the axes: "xsq" (RNA-seq), "sig" (Signatures), "mut" (gene mutations), "mda" (Miscellaneous data), "hll" (Hallmark gene sets) and "act" (Drug Response). These parameters can be displayed on each axis independently of each other. For instance, plotting ASCL1 expression vs. NE (Neuro Endocrine) signature shows the expected high correlation (Pearson correlation (r)=0.79, p-value= e-51) (Figure S1A).

Univariate analysis data can be downloaded as Tab-delimited format. Figure S1B compares the expression distribution of NEUROD1, ASCL1, POU2F3 & YAP1 in the samples from untreated and relapse patients. It demonstrates a loss of the bimodal distributions for ASCL1 and POU2F3 in the relapse samples while YAP1 expression is also significantly reduced in the relapse samples (Figures S1A-B and 1). As result the assignment of samples to each of the NAPY groups cannot be applied to more than 50% of the relapse samples while, as reported previously [1], it discriminates the samples from untreated patients (Figure 1B)

The "Multivariable Analyses" tab (labeled "2" in Figure 1A) shows as default the Heatmap "Display" of ASCL1 expression (xsq: RNA-seq) vs. the expression of key genes driving the biology of SCLC: INSM1 & CHGA used as immunohistochemistry NE markers, MKI67 (encoding Ki67 used as proliferation marker), as well as the SCLC pioneer transcription factors NEUROD1, POU2F3 & YAP1, and the oncogene MYC (Figure 2A). Each row displays the selected genes and the columns the 235 individual patient samples as a Heatmap). Users can explore different inputs and outputs by entering chosen gene names in the "Response Identifier" or "Predictor Identifiers" windows (left panels of Figure 2A). Examples of "Multivariate

130 Analyses" for the NAPY genes are shown in Figure 2A-E. They will be discussed in the next
131 "Network and Pathway Analyses" section.
132 In the "Multivariate Analyses" tab, the default visualization setting is "Display". Users can
133 also choose the "Linear Regression" tab (found in bottom left corner of the website under the label
134 "Algorithm"; Figure 2A). This feature computes the correlations between each of the "Predictive
135 Identifiers" and the "Response Identifier". Figure S2A-C shows the results for ASCL1 for the
136 genes entered in Figure 2A. The "Plot" option demonstrates the high predictive value of the
137 combined expression of INSM1, CHGA, MKI67, NEUROD1, POU2F3, YAP1 and MYC as
138 predictors of ASCL1 expression (Figure S2A) (Pearson correlation r = 0.89; p-value = 6.8e-83;
139 numbers displayed on top of the graph). The contribution of each of the 7 predictor genes are
140 available under the "Technical Details" tab (Figure S2B), demonstrating the prominent
141 contribution of INSM1, CHGA as positive predictors and NEUROD1, POU2F3 and MYC as
142 negative predictors. Users can also choose "Lasso" (Least Absolute Shrinkage and Selection
143 Operator), a machine-learning and statistical method that identifies genomic biomarkers predicting
144 the expression of the "Response Identifier". For YAP1 expression, the top gene identified is the
145 Neuron Restrictive Silencing Factor REST, a key downstream effector of the Notch pathway [7].
146 The "Expression Distribution" tab (labeled "3" in Figure 1A) displays the expression of
147 any chosen gene across the 4 SCLC_TumorMinerCDB datasets (the two untreated NCI & RU
148 datasets and the two relapse TU & UoC datasets). Schlafen 11 (SLFN11) expression is displayed
149 as default on the website (Figure S3A). SLFN11 expression tends to be suppressed in relapse
150 samples, which can explain the resistance of relapse patients to DNA-targeted chemotherapies
151 (such as etoposide, topotecan, cisplatin and lurbinectedin). Users can display and download data
152 for any gene from each of the 4 datasets by inserting the corresponding gene HUGO name in the
153 "Gene symbol" box of the website.
154 The "Mutation Variants" tab (labeled "4" in Figure 1A) displays mutations for up to 5
155 genes at a time. Colors reflect the types of mutations, and the variants can be viewed and
156 downloaded using the "View variants" tab (Figure S3B). As expected TP53 is the most frequently
157 mutated gene (81% of the 114 samples with Exome-sequencing data). The second most frequently
158 mutated gene is TTN, which encodes Titin. This is because titin is the largest human protein, and
159 TTN the largest target for exome sequencing, thereby reflecting the high mutation burden of
160 SCLC. Accordingly, TTN gene mutation is correlated with Tumor Mutation Burden (TMB)
161 (Pearson correlation = 0.37, p-value = 6e-05 in the "2-variable analyses" tool of
162 SCLC_TumorMiner; not shown in the figures). As expected, the next most mutated gene is RB1
163 with an overall mutation frequency of 55% (Figure S3B). Although inactivation of SLFN11 leads
164 to resistance to chemotherapy [15,20,21], its mutation frequency is low (only 4% of all 114 sequenced
165 samples; Figure S3B) (13% vs. 2% in the untreated).
166 The "Prognosis Biomarkers", "Predictive Biomarkers" and "MyPatient" tabs (numbered 5,
167 6 & 7 in Figure 1A) will be discussed in the remaining sections of the manuscript with examples
168 displayed in Figures 5, 7 and S6.
169
170 **Genomic networks and canonical pathways of SCLC patient samples**
171 We previously reported [7] that for the SCLC cell lines, the expression of genes functioning together
172 in a given pathways are cross-correlated (see Fig. 5 in [7]). The main pathways for SCLC comprise
173 the neuroendocrine (NE) pathways driven prominently by ASCL1 and less frequently by
174 NEUROD1, and the non-NE pathways driven by POU2F3 and YAP1, the MYC genes and genes
175 of the Notch pathway [1,7].

As in the SCLC cell lines (https://discover.nci.nih.gov/SclcCellMinerCDB/) [7] and consistent with prior studies of clinical samples, the group of NE-SCLC is prominent across our four clinical datasets based on the expression of the two genes used as immunochemistry NE markers, INSM1 and CHGA (Figures 2A). As expected, ASCL1 expression is highly correlated with the NE markers, while NEUROD1

Next, we used the "Multivariable Analysis" tool to investigate the networks/pathways of each of the 4 NAPY genes (Figure 1B). The approach is an extension of our previous genomic network analyses in the 118 SCLC cell lines for ASCL1 and YAP1 (Figure 5 in ref. [7]). Here we also analyzed NEUROD1 and POU2F3 (Figures 2C-E) in the tumor samples and compared the 235 tumor samples to the 118-cell line collection ((Figure S2C). Figure 2A shows that multiple samples express both ASCL1 and NEUROD1 and that NEUROD1-expressing samples do not closely match INSM1 or CHGA. Figure 2A also shows a significant overlap between YAP1, POU2F3 and MYC expression.

Analysis of the ASCL1 coexpression network [7] (Figure 2B) shows that most of the samples expressing ASCL1 also express INSM1 (Pearson correlation r = 0.72; p-value = 3.6e-38 in the "2 variable analysis"). As expected, NKX2-1 and PROX1, the transcriptional coactivators of ASCL1 [7,22] (Figure 5 in [7]) show high coexpression (Pearson correlations r = 0.67 & 0.55; p-values = 3.2e-32 and 9.3e-20, respectively using the "2 variable analysis" tool). The downstream transcriptional target of ASCL1, DLL3 (encoding the Delta Like Canonical Notch Ligand 3), a validated SCLC therapeutic target, is highly expressed in approximately two-third of the samples, and its expression is highly correlated with ASCL1 (Pearson correlation r = 0.82; p-value = 5.7e-58 in the "2 Variable Analyses"; highest correlate with ASCL1 expression in "Compare Patterns" of the "2 Variable Analyses"). *SEZ6* (encoding Seizure 6-like), another ASCL1 target gene and an emerging cell surface target for Antibody Drug Conjugates is also highly correlated with both ASCL1 and DLL3 (Figure 2B). Of note, SEZ6 expression is highly correlated with SEZ6L (SEZ6 homolog like) in spite of different chromosome locations (17q11.2 and 22q12.1, respectively), suggesting that both genes function in the ASCL1 pathways. The "Compare Analyses" tool of SCLC_TumorMinerCDB identifies additional genes associated the ASCL1 genetic network, such as GRP (Gastrin Releasing Peptide) (4th most highly correlated gene after DLL3 and SEC11C (SEC11 Homolog C, Signal Peptidase Complex Subunit)). Taken together, the 9 "Predictor genes" listed in Figure 2B are highly predictive of ASCL1 expression (r = 0.87; p-value = 6.5e-73; Figures S2C) and comprise a significant part of the ASCL1 network/pathways.

Analysis of the NEUROD1 coexpression network shows that NEUROD1 expression only accounts for a fraction of the NE samples as many samples have high INSM1 expression without expressing high levels of NEUROD1 (Figure 2C, compare the 1st and 2nd rows). NKX2-2 (NK2 transcription factor related, locus 2 (Drosophila)), which is involved in the morphogenesis of the central nervous system and a known regulator of NEUROD1 in pancreatic islet cells together with the Neurogenin genes (NEUROG1 and NEUROG3) [23] are co-expressed with NEUROD1 (Figure 2C). The coexpression of PAX6 (Paired Box Homeotic Gene-6), which encode a transcription factor important for the development of the eyes, nose, pancreas and the central nervous system, is logical as it has been shown to control NEUROD1 expression [24]. Using the "2 variable Analysis" tool and the "Compare Patterns" tools of SCLC_TumorMinerCDB (see Figure 1A; red dotted box), we found coexpression of two genes encoding cell surface proteins and potential targets for immune therapies, *GFRA1* (Glial Cell Line-Derived Neurotrophic Factor Receptor Alpha) and *SEMA6A* (encoding one of the semaphorins) [25] (Figure 2C). The SCLC cell lines [7] also express significant levels of SEMA6A with high correlation with NEUROD1 (p = 0.57 with p-value =

222 2.7e-11). Whether NEUROD1 directly drives the expression of GFRA1 and SEMA6A remains to
223 be determined. Taken together, the 8 NEUROD1 "Predictor genes" are highly predictive of
224 NEUROD1 expression (r = 0.85; p-value = 2.6e-65; Figures S2A-C) and could be used to establish
225 a prominent NEUROD1 subtype in tumor samples. They also identify a significant part of the
226 NEUROD1 network/pathway.
227 Analysis of the POU2F3 network shows the expected inverse expression between POU2F3
228 and INSM1 (Figure 2D, compare the 1st and 2nd rows) as POU2F3 SCLC cells are derived from
229 the Tuft cells of the bronchial epithelium and are non-NE [26-28]. Figure 2D shows that the tumor
230 samples expressing POU2F3 also coexpress 4 transcription factors known to drive POU2F3
231 expression [26,29]: ASCL2, COLCA2 (also known as OCAT-2 or POU2AF3), C11orf53 (also known
232 as OCAT-1 or POU2AF2) and SOX9. Consistent with the Tuft lineage of the POU2F3-driven
233 SCLC (SCLC-P), POU2F3 is also highly correlated with the Tuft-ionocyte progenitor cell-specific
234 genes FOXI1 and AVIL [27,28], and with TRPM5 (Transient Receptor Potential Cation Channel
235 Subfamily M Member 5), which encodes a cell surface ion channel protein involved in immune
236 defense, and was recently shown to be highly expressed in bronchial Tuft cells [30]. This suggests
237 that TRPM5 could be considered as cell surface target for SCLC-P tumors. Figure 2D also shows
238 that the POU2F3 network/pathway is coordinated with the NOTCH pathways as POU2F3
239 expression is highly correlated with REST and its transcriptional activator NOTCH1 (Pearson
240 correlations r = 0.67 and 0.46; p-values = 9e-22 and 2.7e-9, respectively in the "2 variable analysis"
241 for the untreated samples). MYC expression is also highly correlated with POU2F3 (Pearson
242 correlation r = 0.74; p-value = 5.6e-28 in the "2 variable analysis" in the untreated samples), which
243 is consistent with independent reports [18,28,31]. Taking together the expression of the 11 genes
244 displayed in Figure 2D could be used as predictor for POU2F3 expression and for classifying a
245 given tumor sample as SCLC-P (Pearson correlation = 0.94; p-value = 3.3e-112) (Figure S2C).
246 Analysis of the YAP1 network confirms that samples with high YAP1 expression lack
247 INSM1 expression. Consistent with the SCLC cell lines analyses (Figure 5 in [7]), TAZ (encoded
248 by *WWTR1*), the transcriptional co-activator partner of YAP1 is coexpressed with YAP1 (Pearson
249 correlation r = 0.81; p-values = 1.24e-56). Similarly, TEAD2 and TEAD3, the downstream
250 transcription factors of YAP/TAZ are coexpressed with YAP1. Figure 2E also shows that the
251 downstream transcriptional targets of YAP1 (see Figure 5C-D in [7]) including SMAD6, VGLL4,
252 PTPN14, and the cell adhesion genes CCN1, CTNNA1 and CTNNA2 are coexpressed with YAP1.
253 Also consistent with the SCLC cell line data (see Figure 5C-D in [7]) the NOTCH pathway (REST,
254 NOTCH1 and NOTCH2) is co-expressed in the SCLC-Y tumor samples. Together the expression
255 of the 12 genes displayed in Figure 2E is a statistically significant predictor for YAP1 expression
256 and classification of a given tumor sample as SCLC-Y (Pearson correlation = 0.89; p-value = 1.9e-
257 81) (Figure S2D).
258 Analysis of the Notch network using REST, the transcriptional target and effector of the
259 Notch pathway (Figure 2F) shows, as expected, coexpression of REST with the 4 Notch receptors,
260 NOTCH1-4 and with YAP/TAZ, consistent activation of the Notch pathway by YAP1-mediated
261 transcription [32]. By contrast, the Notch ligands (JAG1/2, DLL3, DLL1, DLL4 and DLK1) are
262 significantly suppressed in the Notch receptor-positive samples and coexpressed with DLL3 and
263 ASCL1 (Figure 2F), consistent with the conclusion that the YAP/Notch/REST network controls
264 the NE fate of SCLC cells [32] and vice-versa as DLL3 suppresses Notch signaling in SCLC [33].
265 The Myc pathway (including MYC, MYCL and MYCN) is activated in 52% of the
266 untreated patient samples where MYCL dominates (36.4% of the samples) and is linked to ASCL1
267 expression (SCLC-A) (Figures 2B & 2G). By contrast MYC is highly expressed in a smaller

268 fraction of the untreated patients (15.6%) and primarily in the SCLC-P samples (Figures 2D) [34].
269 Consistent with the fact that POU2F3 disappears in the relapse samples and that ASCL1 expression
270 is also attenuated in the samples of relapse patients, the expression of both MYCL and MYC is
271 reduced in the relapse samples; with a greater reduction for MYCL at relapse (Figures 2G).
272 To test the robustness and conservation of the 5 networks displayed in Figure 2 (ASCL1,
273 NEUROD1, POU2F3, YAP1 and Notch) we compared their predictive value in TumorMiner and
274 CellMinerCDB. Figure S2C shows that the 5 genomic networks are highly conserved in tumor
275 samples and cell lines.
276
277 **Classification of SCLC tumor samples**
278 Although all the samples from the untreated patients can be assigned to the NAPY classification,
279 less than half (47%) of the relapse samples can be classified using RNA-seq expression levels > 5
280 (log2(FPKM+1)) (Figure 1B). To overcome this limitation, we first performed cluster analyses
281 using 12 genes by adding to the 4 NAPY genes (highlighted in red in Figure 3) the two NE genes
282 (INSM1 and CHGA), MYC and REST and 4 innate immune genes of the major histocompatibility
283 complex (MHC): HLA-A, HLA-B, HLA-C and B2M (Beta-2-Microglobulin) that are part of the
284 recently defined inflamed signature of SCLCs (SCLC-I) [2]. This 12 gene algorithm successfully
285 classified the untreated patients in 2 groups (Figure 3A; see annotations under the heat map). The
286 largest is the SCLC-NE group (≈ 70% of the patient samples) with high expression of ASCL1,
287 INSM1, CHGA and/or NEUROD1, and the 2nd the POU2F3 group (SCLC-P). As discussed above
288 the SCLC-P also includes samples with high MYC, REST and to a lesser extent YAP1 expression.
289 The inflamed group characterized by high MHC expression (HLA-A/B/C and B2M) is more
290 diffuse (≈ 50% of the patient samples) and tends to be excluded from the SCLC-NE group and to
291 encompass the SCLC-P group.
292 For relapse patient samples, the 12 gene signature did not provide clear clusters (Figure
293 3B), and the ASCL1, NEUROD1, INSM1 and CHGA genes failed to form defined clusters. The
294 only visible separation is for the MHC genes with ≈ 50% of the samples MHC positive (SCLC-I)
295 (Figure 3B; see annotation under the heatmap).
296 To overcome the difficulty of classifying relapse patient samples, we performed NMF
297 (Non-Negative Matrix Factorization) analyses [2] using the top-5000 protein coding genes most
298 variably expressed in the samples from untreated and relapse patients (Figure 3C, Table S2 and
299 S3). Among those genes, 65% are common for the untreated and relapse samples (Figure S5A).
300 Cophenetic analysis showed that 4 clusters gave a satisfactory division of the samples (Figure
301 S5B-C). The top 30 selectively expressed genes for each of the 4 clusters for the samples of
302 untreated patients (NMF1u, NMF2u, NMF3u and NMF4u) are displayed in Figure 3D. Similarly,
303 the top 30 selectively expressed genes for the clusters of relapse patient samples (NMF1r, NMF2r,
304 NMF3r and NMF4r) are displayed in Figure 3E. Overall, only a few genes are common in the
305 NMF clusters of untreated and relapse patient samples: 5 out of 30 between NMF1u and NMF1r,
306 3 out of 30 in NMF2u and NMF2r, and 2 out of 30 in NMF3u and NMF3r (see diagonal lines
307 between panels D and E in Figure 3) See (Table S4 and S5).
308 Pathway analyses for each of the clusters show that NMF1u consists of genes involved in
309 protein, RNA metabolism and cell cycle division (Figure S5A), and includes the NE genes ASCL1
310 (Figure 3H) and GRP (coding for the gastrin releasing peptide; see above) (Figure 3D). The NE
311 enrichment of NMF1u is easily visible through the mapping ASCL1, NEUROD1, INSM1 and
312 CHGA in the NMFu clusters (Figure 3F). NMF2u also shows a strong NE enrichment (Figure 3F),
313 which is confirmed by Reactome pathway analysis (Figure S5A). NMF3u coincide with POU2F3

expression (Figures 3F and 3H), and with immune and interleukin pathways (Figure S5A) as well as with the MYC pathway (Figures 3D; MYC and MYB annotated in red). NMF4u partially extends to NMF3u and shows positive enrichment for immune, cytokine and interferon pathways (Figures 3D and S5A). It is also enriched for YAP1 (Figure 3H).

Pathway analyses for the 4 relapse NMF clusters show that NMF1r includes genes and pathways related to NE differentiation and cell cycle progression (Figures 3E and S5B) as well as ASCL1 (Figure 3H). NMF2r and NMF1r overlap with genes involved in cell metabolism as well as NE pathways (including NEUROD1) (Figures 3E, 3H and S5B). NMF3r includes genes involved in metabolism and complement. It also includes YAP1 (Figure 3H). NMF4r consists specifically of immune genes (Figures 3E and S5B) and some YAP1-expressing samples (Figures 3G and 3H).

Figure 4A summarizes the Reactome pathway analyses across the samples. It demonstrates the similarities between the untreated and relapse NMF1 and NMF2 clusters with enrichment for NE and cell cycle pathways. It also demonstrates the prominent enrichment for immune pathways both in the untreated and relapse NMF 3 and 4 clusters. The complete pathway enrichment results for untreated and relapse SCLC are provided in Table S6 and Table S7, respectively.

NMF clusters coincide with molecular signatures

Having performed NMF classification, we derived the NMF signatures (4 untreated and 4 relapse; see Methods) and integrated them into SCLC_TumorMiner. This feature enables direct explorations and correlation analyses with other parameters included on the website. Figure 4 explores the correlations between the NMF signatures and the transcriptional signatures included in SCLC_TumorMiner: NE, RepStress, CES, Epithelial, Stroma, Immune, APM, Ychrom. The signatures were computed as the mean expression of the genes included in each signature (see Method section). Figure 4B tabulates the correlations between each NMF and the 8 signatures (Pearson correlation coefficients obtained by comparing NMFs and individual signatures). The colors reflect the direction and intensity of the correlations (red for highly positive and blue for highly negative). Consistent with the top gene signatures (see Figure 3D-E) and the Reactome pathway analyses (see Figure S5A-B), the NMF1u, NMF2u, NMF1r and NMF2r signatures are highly correlated with the NE signature as well as with the RepStress and CES (Chromosome Centromere and Kinetochore score) signatures. They are also significantly devoid of the Immune, APM (Antigen Presentation Machinery) and Stroma signatures (Figure 4B). Conversely, the NMF3u, NMF3r, NMF4u and NMF4r are enriched for the Immune, APM, Stroma and epithelial signatures and devoid of the NE, RepStress and CES signatures.

Figure 4C shows examples of the significant correlations NMF2r and NMF4r with 4 of the 8 signatures as snapshots of "2 Variable Analyses". Together these results demonstrate the functional significance of the NMF classification for SCLC patient biopsies.

Prognosis biomarkers and risk factors

The SCLC_TumorMiner website is designed to assess prognosis biomarkers (Figures 5 and S6). Users can test risk factors based on gene expression, mutations, clinical metadata, gene and NMF signatures. The default display on the website is for SLFN11 (Schlafen 11) (Figure S6A) whose expression is known to predict response to chemotherapeutic agents targeting DNA replication [15,20]. SLFN11's wide range of expression in SCLC samples (Figure 6E and S8) makes it a quantitative biomarker. Taking the 4 datasets together, high SLFN11 expression is associated with significantly better survival (Figures S6E), which agrees with a recent publication [35] demonstrating

preferential response of SLFN11-expressing cells to the DNA-targeted chemotherapies used to treat SCLC.

The prognostic value of each of the eight NMF gene signatures can be explored in the "Prognosis Biomarkers" tool of SCLC_TumorMiner. NMF4r (inflammatory-immune) is a good prognosis biomarker for the entire set of 205 patients (Figure 5A-C). The good prognosis value of the APM, immune and Stroma signatures (Figures 5D and S6B) also indicates that tumors with high immune and stroma infiltration respond better to treatments.

Notably, none of the NAPY genes show significant predictive value, even in the untreated patients despite their differential expression (see Figure 1). Similarly, MYC, MYCL, REST and NOTCH do not show significant predictive value. On the other hand, high expression of some genes is associated with high risk in untreated patients (Figures 5E). For instance, consistent with a recent study reporting that high expression of Serine and Arginine Rich Splicing Factor 1 (SRSF1) gene is linked with poor survival in SCLC patients [36], we also observe that high expression of SRSF1 is a risk factor for the untreated patients (HR 2 groups = 1.7; p-value = 0.02), while its significance in the relapse patients (NCI + UR datasets) is less (HR 2 groups = 1.3; p-value = 0.19).

In addition to univariate analyses, user can perform multivariable analyses by selecting combination of multiple genes or signatures. Figures 5E-F shows the example of combining MKI67, which encodes the proliferation biomarker Ki67, an established poor prognosis biomarker, with SEZ6 and SRSF1 in the 150 samples of untreated patients. High expression of the 3 genes together shows significantly poor prognosis (HR 2 groups = 1.7; p-value = 0.01). Additional predictive biomarkers can be searched using the Lasso analysis option (highlighted in yellow at the bottom left in Figure 5A).

Predictive biomarkers for cell surface targets

ADCs are rapidly changing cancer treatments by allowing tumor-targeted delivery of chemotherapeutic agents, which otherwise exhibit dose-limiting toxicity. Thirteen ADC have been approved in the past 6 years while the FDA only approved one small molecule, Lurbinectedin for metastatic SCLC. The most recently FDA-approved ADC (January 2025) is Datroway® (Datopotamab Deruxtecan) delivering the TOP1 inhibitor Deruxtecan to cancer cells expressing TROP2 (TACSTD2) [37]. In parallel, CAR-T cells and bi- and tri-specific T-cell engagers are in active development. Among the T-cell engagers, Imdelltra® (Tarlatmab-dlle) is approved for the treatment of ES-SCLC [8] without DLL3 biomarker(s) for patient selection. Although the DLL3-targeted ADC with a tubulin inhibitor payload, Rovalpituzumab teserine (ROVA-T) showed clinical activity [12], it was discontinued due to its dose-limiting toxicity [38].

Figures 6A-B displays the expression of clinical relevant cell surface targets (Table S8). In the untreated patient samples, DLL3 shows a broad range of expression with high expression in multiple NE samples (Figure 6A and 6C and S7A) belonging to the NMF1u & NMF2u groups (Figure 6C). This result is consistent with those of the SCLC cell lines [7]. Notably, we find reduced DLL3 expression at relapse (Compare Figures 6C and D; see also Figure S7A), consistent with the reduced ASCL1 expression at relapse (see Figure 1A). Together these results highlight the potential of measuring DLL3 expression as biomarker for DLL3-targeted therapies and suggest that DLL3-targeted therapies should be moved to front-line treatment.

Because TROP2 ADCs, Datroway® and Trodelvy® , are FDA-approved for triple-negative breast cancer (TNBC), and Trodelvy® has been granted Breakthrough Therapy Designation for ES-SCLC patients [39], we looked at TACTSD2 (TROP2) expression across our 235 samples

(Figures 6A-D and S7A). Like DLL3, TROP2 expression is low in relapse tumors (Figure 6 and S7A). Only the samples of untreated patients with low NE score and high APM score highly express TROP2 (Figure 6A). Accordingly, TROP2 expression is highest in the NMF3u (POU2F3-endriched) & NMF4u (Immune-enriched) clusters (Figures 6 and S7A). Further analyses in the SCLC cell lines of CCLE also show that none of the 50 SCLC cell lines express TROP2 above log2(FPKM+1) >5 while 24% of the cell lines expressed DLL3 above this threshold. These results suggest potential utility of Trodelvy® or Datroway® in newly diagnosed patients with non-NE (POU2F3), NMF4 and APM (Inflammatory/immune) signatures.

Figures 6 and S7A display the expression of 6 additional cell surface targets with high expression [log2(FPKM+1) >5] in at least 10 samples. DLK1 (Delta Like Non-Canonical Notch Ligand 1) is highly expressed many of the untreated patient samples with NE signature (Figures 6A and S7A). It is also expressed in relapse tumors although less than in the untreated group (Figure 6A-D). NMF analyses show DLK1 expression in the NMF1u and NMF2u and in approximately 50% of the samples in the NMF4u group (Inflammatory and Immune group) (Figure 6C). These results are encouraging as a DLK1-ADC is being developed and tested in NE tumors [40].

Two of the carcinogenic embryonic antigens (CEACAM5 and 6), which are targeted with several ADCs in clinical trials for colon, gastric cancers and NSCLC, are highly expressed in both the untreated and relapse patient samples with NMF1u and NMF2r signatures (Figures 6A-D and S8). These results suggest the potential of CEACAM5/6-targeted ADCs in SCLC patients. It also suggests the potential value of measuring CEA in the blood of SCLC patients as a biomarker [41,42].

Claudin 18 (CLDN18) is also the target of ADCs in clinical trials for pancreatic cancer. Figure 6A shows that it is highly expressed in a few untreated patients but not in relapse patients.

As discussed above, SEZ6, a downstream target of ASCL1 [43] shows high expression across untreated and relapse samples matching the NE signature (Figures 6A-D and S7A). Notably, the overlap between SEZ6 and DLL3 is only partial. We also found that the B7-H3 (encoded by CD276, and belonging to the immunoglobulin superfamily), a common ADC cell surface target, is expressed in a broad range of SCLC samples (Figures 6A-D and S7A). These results validate the potential of SEZ6- and B7-H3-targeted ADCs in SCLC [44].

Predictive biomarkers for chemotherapy and ADC payloads

Because ADCs act by delivering chemotherapeutic payloads, we explored the intracellular targets and determinants of response to these payloads. Figures 6E-J and S7B-C summarize the expression of intracellular drug targets (Figure E-F), drug efflux transporters (Figure G-H) and proapoptotic genes (Figure I-J) known to drive response to chemotherapeutic drugs.

Schlafen 11 (SLFN11) shows a broad expression range (over 8-fold) both in the untreated and relapse patient samples, in agreement with SLFN11's wide distribution across the SCLC cell lines irrespective of the NAPY subgroups [7] (https://discover.nci.nih.gov/SclcCellMinerCDB/). Consistent with SLFN11being an immune- and interferon-response gene [15], the NMF4 inflammatory/immune subgroups show significantly high expression of SLFN11 both in the untreated and relapse patient samples (Figure 6E and F). Comparing SLFN11 between the untreated vs. relapse samples shows significant suppression of SLFN11 expression in the relapse patient samples (two-tailed $p < 0.0001$) (Figure S7B and S8). The broad range of SLFN11 expression suggests that further studies are warranted to confirm the relevance of SLFN11 to predict response of patients treated with topotecan and TOP1 inhibitor payload ADCs [3,14,45].

451 Since TOP1 is the primary cellular targets of topotecan and TOP1-ADC, and TOP2 the
452 target of etoposide, we looked at their expression (Figure 6E-F). Both are highly expressed in the
453 samples from untreated and relapse patients with a slightly lower expression in the
454 inflammatory/immune subgroups (NMFu and NMFr). Thus, TOP1 and TOP2 levels are irrelevant
455 as predictive biomarkers.
456 Lack of O-6-Methylguanine-DNA Methyl-Transferase (MGMT) is well-established as
457 predicting the activity of DNA alkylating agents, and foremost of temozolomide [46], which is
458 occasionally used in SCLC [47]. Figure 6 (right panels E & F) shows low expression MGMT in some
459 samples, suggesting the potential of measuring MGMT expression for temozolomide treatments.
460 Drug efflux transporters of the ABC family are a major cause of resistance to tubulin
461 inhibitors that are commonly used as ADC payloads. Yet, measuring ABC transporters is not done
462 in clinical routine. Figure 6G-H displays the expression of the main ABC transporter genes.
463 ABCB1, which encodes P-glycoprotein (PgP) has notably low expression both in the untreated
464 and relapse samples, while ABCC2 and ABCC3 show significant overexpression in the NMF3r
465 cluster (Figure 6H). Together, these results suggest that measuring ABC transporters should be
466 considered for ADCs with anti-tubulin payloads.
467 Because broad resistance of cancers to chemotherapy is dependent on their reduced
468 apoptotic threshold, we explored the apoptotic genes. Figures 6I-J and S7C shows global
469 suppression of proapoptotic gene expression in relapse patient, which may account, in part for the
470 overall resistance of relapse patients to chemotherapy.
471
472 **Analysis of individual patients using the "MyPatient" tab of SCLC_TumorMiner**
473 One of the long-term goals of SCLC_TumorMiner is to provide an interface for classifying tumors
474 and providing prognosis and rational therapies to a given patient (Figure 7A).
475 As discussed above (see Figure 1B), untreated patient samples can be assigned to at least
476 one of the NAPY groups based on high expression cut-off of log2(FPKM+1) > 5. In our dataset,
477 as expected, most samples are prominently NE (74% ASCL1, 20% NEUROD1), with 14% of the
478 samples dual ASCL1 & NEUROD1. The POU2F3 group represents 10% of the samples, which is
479 consistent with prior publications [1]. The YAP1 group is the smallest with 8% of the samples. We
480 found only 2 samples (1%) with both ASCL1 & YAP1 (see Figures 1B). Thus, we conclude that
481 the NAPY classification can effectively classify samples from untreated patients.
482 The situation is different for the relapse samples (see Figures 1B and S9A). Less than half
483 of the samples (46%) can be assigned to the NAPY groups: 31% to the ASCL1 group, 14% to the
484 NEUROD1 group (i.e. a total of 45% are NE), only 1 sample to the POU2F3 group, and no sample
485 to the YAP1 group. Thus, the NAPY classification is deficient for relapse patient samples.
486 To overcome this limitation, based on our data, we developed open-access NMF classifiers
487 (listed in Table S5). We first selected the genes with a 2-fold selectivity compared to the genes in
488 the other 3 NMF groups. This led to a list of genes in each NMF group:128 for NMF1r, 49 for
489 NMF2r, 299 for NMF3r and 293 for NMF4r. The same was done for the untreated samples
490 providing 54 genes for NMF1u, 137 for NMF2u, 85 for NMF3u and 386 for NMF4u (Table S4).
491 Figure S9B shows the mean average scores for each NMFr across the 81 samples from relapse
492 patients. A threshold of 3.8 was chosen to differentiate the samples assigned to each NMF class.
493 This approach enables the computation of mean average NMF scores for individual samples. We
494 implemented this approach to score automatically individual samples in the "MyPatient" tab of
495 SCLC_TumorMiner (Figure 7B). Upon typing a patient identifier (code), SCLC_TumorMiner
496 automatically displays key parameters for classifying the patient sample.

497 The "My Patient" tab also displays predictive biomarkers for ADCs, BiTEs or CAR-T
498 therapies (Figure 7C). This is done by the TumorMiner software, which queries a dataset listing
499 46 cell surface targets for ADCs currently FDA-approved and in clinical development (Table S8).
500 The website displays the cell surface targets expressed above a threshold of 5 (log2(FPKM+1)).
501 For example, in Figure 7C, the displayed patient tumor sample (ID: S00035T) overexpresses
502 SEZ6, NCAM1, DLK1, FN1 and SLFN11, implying that the patient could be candidate for ADCs
503 comprising SEZ6, NCAM1, DLK1 or FN1 antibodies and a TOP1 payload based on high SLFN11
504 expression.
505
506 These results highlight the translational potential of the SCLC_TumorMiner platform
507 for individualized patient care. While traditional NAPY classification remains useful for untreated
508 tumors, its limited applicability to relapsed samples underscores the importance of alternative
509 classifiers. By implementing NMF-based transcriptomic classifiers and integrating biomarker-
510 based predictions for targeted therapies, the "MyPatient" module enables real-time, personalized
511 tumor profiling. This facilitates not only tumor subtype identification but also actionable
512 therapeutic insights, bridging molecular characterization with clinical decision-making in SCLC.
513 As the platform evolves with more comprehensive datasets, it holds promise for guiding precision
514 medicine strategies in both newly diagnosed and treatment-refractory patients.
515
516 **DISCUSSION**

517 Mining cancer patient clinical and genomics data is critical for understanding the state of the
518 disease, finding biomarkers, predicting disease outcome or drug response, and ultimately shifting
519 patient care to a logic-based and flexible paradigm. During the last decade, new tools have emerged
520 such as cBioPortal (http://cbioportal.org), Xena (https://xena.ucsc.edu)), GEPIA2
521 (http://gepia.cancer-pku.cn/), UALCAN (https://ualcan.path.uab.edu/) or CARIS
522 (https://www.carislifesciences.com/) offering different functions for data visualization and
523 analyses mining The Cancer Genome Atlas (TCGA) or public datasets. Here we present "SCLC-
524 TumorMiner" a new web portal for mining patient clinical, genomics and treatment response data
525 focusing on SCLC, which was not included in TCGA. SCLC_TumorMiner hosts samples from
526 both relapse and untreated patients. Its architecture and displays are constructed on the architecture
527 of SCLC_CellMinerCDB (https://discover.nci.nih.gov/SclcCellMinerCDB), which enables users
528 to compare tumor samples with a broad collection of SCLC cell lines derived primarily from
529 untreated patients [7,48].
530 SCLC_TumorMiner (https://discover.nci.nih.gov/SclcTumorMinerCDB/) contains a
531 battery of unique tools (see Figure 1A) allowing instantaneous display of omics data for individual
532 genes and individual patient samples (*2-Variable Analyses*, *Prognosis Biomarkers* and *Predictive*
533 *Biomarkers*) or for multiple genes to explore molecular networks (*Multivariable Analyses* and
534 *Compare Patterns* tools). It integrates 8 pathway *Signatures* relevant to SCLC biology (NE,
535 RepStress, CES, APM, Immune, EMT, Stroma, Y-Chrom), 50 Hallmark signatures and 46 cell
536 surface targets relevant to the current development of Antibody Drug Conjugates (ADCs),
537 Bispecific T-cell engagers (BiTEs) and CAR-T cell therapies. Data can be downloaded from the
538 SCLC_TumorMiner website for further analyses, as demonstrated in this report by the different
539 plots comparing gene expression in untreated and treated patients and in samples from different
540 NMF groups (see Figures 2G, 3H, 6C-J, S1B-C and S7).
541 A biologically relevant feature of TumorMiner is its ability to explore the logic of
542 molecular networks/pathways based on the premise that individual components of molecular

543 pathways are quantitatively linked and significantly correlated [7]. Figures 2 and S2 demonstrate
544 that each of the NAPY genes has its own network (genomic identity) across tumor samples and
545 that these networks are highly conserved across patient samples and cell line models (Figure S2C),
546 implying their robustness. A notable finding is the interconnection between the Notch, the
547 YAP/TAZ and the ASCL1-NE pathways whereby expression of the YAP/TAZ pathway is tightly
548 linked to activation of the Notch pathway; and vice-versa that activation of the NE (ASCL1-SEZ6)
549 pathway is associated with expression of the Notch ligands (DLL3, DLK1, DLL1, DLL4 and
550 JAG2) and negatively correlated with the Notch receptors and their transcription effector REST
551 (see Figure 2F). These results suggest a feedback loop involving the non-NE pathways including
552 YAP/TAZ and the Notch receptors, and the NE pathway including ASCL1 and the Notch ligands
553 [32,33]. Whether and how YAP/TAZ functions upstream of the Notch transcription factor REST
554 remains to be established.
555 Cancer cell plasticity and heterogeneity are two outstanding characteristics of SCLC
556 [10,19,28,49-52]. This is exemplified by the fact that at relapse, patients become resistant to
557 chemotherapy after initially responding often dramatically. Our analyses reveal several key
558 differences between the samples of untreated and relapse patients.
559 One such difference is the loss of the bimodal expression of NEUROD1, ASCL1, POU2F3
560 and YAP1 in the samples from relapse patients (see Figure 1). This could have, at least two
561 explanations, which are not mutually exclusive: 1/ differentiated cancer cells expressing ASCL1
562 (NE cells) or POU2F3 (Tuft-ionocyte-like cells) are eliminated by the 1st line chemotherapy (see
563 Figures 1A and S1B), leaving behind less differentiated resistant cancer cells; 2/ chemotherapy is
564 less effective toward the less differentiated and more plastic tumor cells with basal cell
565 characteristics [28], which remain in the tumor and become prominent at relapse. Both explanations
566 are consistent with the conclusion that SCLC tumors are endowed with intratumoral heterogeneity
567 (ITH) and plasticity [10,19,28,49-51].
568 Another notable difference is the significant downregulation of MYCL in relapse tumors
569 (see Figures 2G and S1C). This reduction is likely consequential to the fact that ASCL1 is a
570 transcriptional driver of MYCL [53] (as visualized in Figures 2C) [53] and that ASCL1 expression is
571 significantly suppressed in the relapse tumor samples (see Figures 1A and S1B).
572 Prognosis biomarkers (i.e. Risk factors) are also different between relapse and untreated
573 patient samples. While MKI67, SEZ6 and SRSF1 [which encodes a splicing factor associated with
574 poor prognosis [36]], are high risk factors for untreated patients, they are not significant in relapse
575 patients (see Figures 5 and S6C).
576 With respect to the chemoresistance of the relapse tumors, it is likely multifactorial. Indeed,
577 while the expression of SLFN11 and of proapoptotic genes (BAD, BAX, BID, NOXA and PUMA)
578 is high in chemotherapy-naïve patient samples, it is significantly reduced in relapse tumors (see
579 Figures 6 and S7), suggesting that the initial response of patients to various chemotherapies is
580 driven, in part by SLFN11 and pro-apoptotic genes. Conversely, resistance of the relapse patients
581 may be related to the downregulation of the cell death and SLFN11 pathways. Whether these
582 pathways are epigenetically suppressed in relapse tumors and could be reactivated by HDAC and
583 EZH2 inhibitors remains to be tested [45,54,55].
584 Cell surface targets for bispecific antibodies T-cell engagers (BiTEs), antibody drug
585 conjugates (ADCs), and CAR-T are in full development. Regarding the FDA-approved therapies
586 targeting DLL3 and TROP2, we find that expression of DLL3, the target of Tarlatamab
587 (Imdelltra®) is generally low at relapse, consistent with DLL3 being a direct transcriptional target
588 of ASCL1, which is significantly suppressed at relapse [7,53]. This suggests moving Tarlatamab to

589 1st line therapy. Also, since IHC analyses apparently fail to predict Tarlatamab responses [56], adding
590 RNAseq to IHC could identify patients who respond (approximately 40%) to Tarlatamab [8,56] and
591 the other DLL3-targeted therapies in development including other BiTEs, CAR-T cells or ADCs
592 [9,44]; thereby avoiding unnecessary toxicities to non-responders and limiting treatment cost.
593 Surprisingly, expression of TROP2 (encoded by *TACSTD2*) is generally low both in the untreated
594 and relapse patient samples. This questions the potential of Trodelvy® (Sacituzumab govitecan) or
595 Datroway® (Datopotamab deruxtecan) in SCLC patients irrespective of their disease stage.
596 Based on our observations, SEZ6 appears promising for two of the ADCs in clinical
597 development: ABBV-011 and ABBV-706. Yet, like DLL3, SEZ6 is a downstream transcriptional
598 target of ASCL1 [43,44] (see Figure 2) and is therefore likely to be targeted more effectively in
599 untreated patients. Our systematic screening for other cell surface targets reveals that CEA
600 (carcinoembryonic antigen) encoded by *CEACAM5* and *CEACAM6* is highly expressed in multiple
601 samples from relapse patients, suggesting the potential of CEA-ADCs in development [57-59]. An
602 advantage of targeting CEA is that, in addition to measuring CEACAAM5/6 transcripts and by
603 IHC in tumor samples, clinical trials could include CEA blood sampling [60,61]. Of note, we find that
604 TRPM5, which encodes a sodium channel transporter, is selectively expressed in the POU2F3-
605 expressing tumors and that GFRA1 (Glial Cell Line-Derived Neurotrophic Factor Receptor Alpha)
606 and SEMA6A (encoding one of the semaphorins) [25] are specifically expressed in NEUROD1
607 expressing SCLC (Figure 2), suggesting they might be considered as novel cell surface targets for
608 SCLC.
609 As we consider moving toward precision medicine, multivariable analyses will be needed.
610 For instance, for a given cell surface target, clinicians may have to choose between an ADC with
611 a TOP1 or a tubulin payload or a T-cell engager (BiTE). One could propose guiding payload choice
612 with at least two biomarkers: SLFN11 and ABC drug efflux transporters. If the tumor expresses
613 high SLFN11 and drug efflux transporters, TOP1 inhibitors will likely be superior to a tubulin
614 inhibitor ADC, and conversely a tubulin inhibitor ADC payload would be logical for tumors with
615 low SLFN11 and ABC transporters expression. For immune checkpoint inhibitors and BiTEs, one
616 could consider testing the predictive value of high APM and Immune signatures, which would
617 reflect the presence of immune cells in the tumor.
618 Ultimately, we envision that the "*MyPatient*" tab of SCLC_TumorMiner could be
619 developed for clinical practice as bulk RNA-seq and DNA-seq are readily performed on sample
620 biopsies. Sequencing results (FASTQ files) and clinical parameters could be loaded into
621 SCLC_TumorMiner. The "*My_Patient*" tool would provide a medical interface displaying a
622 patient's genomic parameters, dictating which therapy would be most appropriate and predicting
623 risk-factors. Based on our analyses, RNA-seq values can readily be set to separate samples. RNA-
624 seq expression value (log2(FPKM+1) > 5 and <3 discriminate between high and low gene
625 expression, respectively. Tumor_Miner allows the display and comparison of the molecular data
626 of the given patient in the context the historical cohort of SCLC patients. Additionally, the
627 Tumor_Miner architecture and interface could be readily expanded by integrating protein
628 expression (proteomics) and epigenetic (DNA methylome) parameters as we have already done in
629 our CellMiner websites (https://discover.nci.nih.gov/).
630

631

632 **LIMITATIONS OF THE STUDY**

633 Adding more samples with annotated data would further increase the statistical power of
634 SCLC_TumorMiner. However, our experience has been that material transfer agreements for
635 sharing FASTQ files and clinical information between institutions and individual researchers
636 remain a limiting factor. SCLC was not included in TCGA because surgical tumors biopsies are
637 not commonly done in SCLC patients. Obtaining high quality samples in sufficient quantity should
638 be a priority to obtain the necessary omics and pathological data required to precisely analyze
639 individual patients in a disease which remains currently lethal within months. Different genomic
640 platforms were used in our SCLC_TumorMiner database. This creates batch effects, which are
641 difficult to correct and can lead to erroneous conclusions unless one is aware of it. One solution is
642 to harmonize the technical platforms and avoid changing them without the absolute need to do so.
643 Although navigating websites remains a challenge, the rapid advance of Artificial Intelligence (AI)
644 is likely to humanize the website interface.
645

646 **MATERIAL AND METHODS**

647 **Data content and Infrastructure**
648 Patient clinical and genomics data were collected from our DTB clinic at the Center for Cancer
649 Research, NCI-NIH, from the University of Rochester (RU), the University of Cologne (UoC) and
650 Tongji university (TU). Our genomics patient data currently includes a total of 235 samples from
651 212 patients with RNA-seq and/or whole exome seq (WES) samples. The NIH samples were
652 obtained from patients enrolled in IRB-approved protocols and had consented to the analysis of
653 their tumors.
654 Clinical data from NIH and collaborators were gathered and manually curated by a team
655 of DTB physicians. They integrate patient demographics, survival, treatment with response and
656 sample information including sample type, biopsy site, tumor stage and pathology indicators (see
657 Table S2). NCI patient samples were correlated with treatment response, and for each patient we
658 selected for the first drug(s) or treatment(s) post-sampling as well as its response class (CR:
659 complete response, PR: partial response, PD: progressive disease or SD: stable disease). Due to
660 the current limited number of treatments, we decided that the response will be of value 1 (good
661 responder: PR, CR) or 0 (bad responder: SD, PD).
662 All genomics patient data were processed with our on-site data-handling pipelines using
663 the raw data. We generated 235 RNA-seq and 114 WES normalized data. Treatment response were
664 available for 21 patients. Table S9 shows details by disease, data sets and experiments.
665 We also computed for each sample gene expression Hallmark enrichment and genomics signatures
666 scores such as the NE, RepStress, EMT, APM, the level of stroma and immune infiltrate
667 contamination or a new Y chromosome signature scores. Table S10 presents the list of signatures
668 as well as the phenotypic data quarriable on TumorMiner.
669 SCLC-TumorMiner has similar design as the CellMinerCDB tool and websites
670 (https://discover.nci.nih.gov/) dedicated to mining pharmaco-genomics data across broad ranges
671 of cancer cell line sets and databases. SCLC-TumorMiner is implemented as a shiny application
672 and hosted using Posit Connect on CBIIT servers at
673 https://discover.nci.nih.gov/SclcTumorMinerCDB/.

674
675

676 **RNA sequencing**
677 The NIH/NCI and Rochester university (RU) samples were made from FFPE samples using the
678 Illumina access protocol whereas the Universities of Cologne (UoC) and Tongji university (TU)
679 were from Frozen using PolyA protocol. All the SCLC samples were processed with other diseases
680 samples such as Adrenal cortical, Pancreatic NET or GINET samples by running the NCI CCBR
681 RNAseq pipeline (https://github.com/skchronicles/RNA-seek.git). STAR (2.7.6a) was run to map
682 reads to hg38 reference genome (release 36. Then RSEM was used to have gene expression values
683 in log2(FPKM+1). We applied the "RemoveBatchEffect" function from the package Limma to
684 remove the impact of the library preparation protocols (polyA, total RNA or access) and keeping
685 in consideration the diseases differences.
686
687 **Mutations**
688 Somatic mutations were based on paired tumor-normal samples and ran with the CCBR exome
689 sequencing pipeline (https://github.com/mtandon09/CCBR_GATK4_Exome_Seq_Pipeline).
690 Briefly, BWA MEM (version 0.7.17) was run to map reads to the hg38 reference genome. Then
691 Mutect2 in GATK 4.2 was used to call the variants. Variants with 6 or more total filtered reads
692 were selected for display. We used Annovar (hg38 version) to annotate the mutations, including
693 their predicted effect on amino-acid sequence and whether they are predicted to be deleterious, as
694 well as their prevalence in the ExAC normal population. A mutation was called deleterious if its
695 predicted effect on the RNA sequence was frameshift, nonsense or splicesense, or a missense
696 mutation with a SIFT score <0.05 or PolyPhen2 HDIV score>=0.85. We selected all deleterious
697 variants for the mutation gene level score.
698 The fraction of alternate alleles (the mutation ratio) was used as the probability that any
699 cell has that mutation. Furthermore, we assumed that the probability that any mutation is present
700 in a gene is independent of the probability of any other mutation on that gene. Given those
701 assumptions, we calculate the overall mutation probability for that gene in any particular cell as:
702 1-(1-p1)*(1-p2)*...*(1-pn)
703 where p1, p2, ..., pn are the mutation ratios for each individual mutation on that gene. 1-pj is the
704 probability that jth mutation is not present for a particular cell. Thus, the product is the
705 probability that none of the mutations are present and 1 minus the product is the probability that
706 at least one mutation is present. The expression gives a number that is between 0 and 1, which we
707 multiply by 100 to get the mutation score for that gene.
708
709 **Tumor Mutation Burden (TMB)**
710 The TMB was computed using the R package MAFtools [62] based on Mutect2 somatic variants that
711 have variant allele frequency (VAF) >= 5%

712 **Genomic data for survival**
713 If a patient has multiple samples, we select the primary tumor sample (if available) for survival
714 analysis or we choose one based on biopsy site. The list of selected samples is downloadable from
715 the website.
716
717 **Unsupervised Clustering Approach**
718 We implemented unsupervised clustering based on Non-Negative Matrix Factorization (NMF)
719 algorithm using log-transformed normalized RNA-seq data to classify SCLC patients into
720 molecular subtypes. NMF uses dimensionality reduction approach in which it decomposes the

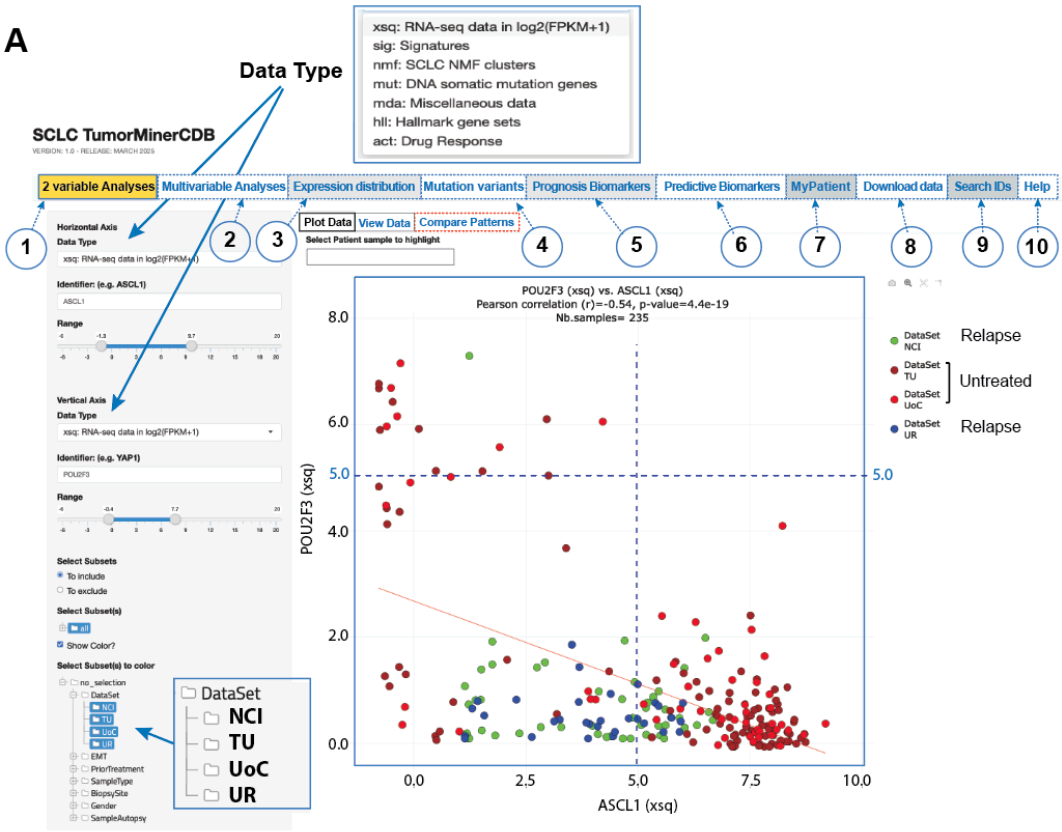
721 gene expression matrix X (n × m), into the product of two lower-dimension non-negative matrices
722 (See Equation 1), where n represents number of patients and m is number of genes.
723 $$X \approx WH \quad (1)$$
724 with X = n × m matrix, W = n × k matrix, and H = m × k, where k is a latent factor; optimal number
725 of factors (k) was determined based on matrix stability using cophenetic correlation coefficient.
726 We selected only protein-coding genes for both untreated and relapse SCLC samples, then we
727 selected the top 5000 highly variable genes for both datasets using a non-parametric approach,
728 Median Absolute Deviation (MAD) using "mad" function in R (version 4.2.3). After which, we
729 performed consensus NMF clustering evaluating values of k ranging from 2 to 8, based on
730 cophenetic correlation coefficient. k = 4 provided most stable and robust clusters for both untreated
731 and relapse SCLC datasets (See Figure S4). Clustering was performed using "NMF" library in R
732 (version 4.2.3).
733
734 **Pairwise Differential Expression Analysis**
735 After defining the NMF subgroups, differential gene analyses were performed using the "Limma"
736 package in R (version 4.2.3). To determine the up or down-regulated genes across the NMF-
737 subgroups, we performed pairwise comparisons for untreated and relapse SCLC separately. This
738 included all possible combinations of comparisons: NMF1 vs NMF2, NMF1 vs NMF3, NMF3 vs
739 NMF4, NMF2 vs NMF3, NMF2 vs NMF4, NMF3 vs NMF4. In addition, each subgroup was
740 compared with the aggregated sum of other subgroups, for instance: NMF1 vs Others
741 (NMF2+NMF3+NMF4), NMF2 vs Others (NMF1+NMF3+NMF4), NMF3 vs Others
742 (NMF1+NMF2+NMF4) and NMF4 vs Others (NMF1+NMF2+NMF3).
743 Subsequently, we utilized these pairwise comparisons to identify subgroup-specific genes
744 for each NMF cluster for both untreated and relapse samples. Significantly upregulated genes in
745 NMF1 in comparison to NMF2, NMF3 and NMF4 with $\log_2$FC >1 and p-value < 0.05 are listed
746 as NMF1-specific genes. Similarly, this criterion was implemented for the selection of specific
747 genes for the NMF2, NMF3 and NMF4 subgroups. The subgroup-specific genes identified using
748 this approach were used to construct NMF-derived gene expression signatures.
749
750 **Gene Set Enrichment Analysis (GSEA)**
751 To investigate the functional relevance of genes we performed pre-ranked Gene Set Enrichment
752 Analysis (GSEA). From previous analysis, we obtained pairwise comparisons for each NMF
753 groups and used t-statistic parameter, which provided magnitude and direction of differential
754 expression to rank the genes. Ranked genes were then analyzed using the fGSEA R package
755 (version 4.2.3) with REACTOME gene sets [https://www.gsea-
756 msigdb.org/gsea/msigdb/human/collections.jsp#C5]. fGSEA is a fast implementation of the
757 GSEA algorithm providing normalized enrichment scores (NES) and adjusting for multiple testing
758 using permutation-based p-values and false discovery rate (FDR).
759
760 **Statistical Methods**
761 Correlations, scatter plots, heatmaps, survival analysis, regression model predictions, and pathway
762 enrichment scores were generated using the R-Project for Statistical Computing based mostly on
763 the R packages: ggplot2, plotly, heatmaply, survival, glmnet, GSVA and GSEABase. These
764 packages are downloadable from the comprehensive R archive network (CRAN) or Bioconductor
765 websites.
766

767 **Table 1.** SCLC samples and Datasets
768

		Untreated		Relapse		
		UoC	**TU**	**NCI**	**RU**	**Total**
Number of samples		56	98	50	31	235
Genomics	RNA Seq	56	98	50	31	235
	WES	0	98	16	0	114
Prior Treatment	Yes	0	0	50	31	81
	No	52	98	0	0	150
	Unknown	4	0	0	0	4
Biopsy Site	Lung	55	98	7	10	170
	Liver	0	0	21	1	22
	Lymph Node	0	0	10	11	21
	Pleura	1	0	1	4	6
	Adrenal Gland	0	0	4	0	4
	Mediastin	0	0	3	0	3
	Others	0	0	4	5	9
Tumor Stage	Stage 1	25	30	0	0	55
	Stage 2	7	26	0	0	33
	Stage 3	17	41	2	2	62
	Stage 4	7	1	48	14	70
	Unknown	0	0	0	15	15

769
770
771
772
773
774
775
776
777
778

Figures with legends



B

	NEUROD1	ASCL1	POU2F3	YAP1	ASCL1 +NEUROD1	ASCL1 +YAP1
Untreated [1]	20% (31)*	74% (114)*	10% (16)*	8% (12)*	14% (22)*	1% (2)*
Relapse [2]	14% (11)*	31% (26)*	1% (1)*	0	2% (2)*	0

* Number of patient samples are in parenthesis

[1] All untreated samples can be assigned to a NAPY group

[2] Only 47% of relapse samples can be assigned to a NAPY group

Figure 1. Introduction to SCLC_TumorMinerCDB (https://discover.nci.nih.gov/SclcTumorMinerCDB/). **A**. Snapshot of the introductory "2 variable Analyses" display (1). Both the horizontal and vertical axes offer a choice of "Data Type" shown in the upper inset. Users must type their choice in the "Identifier" box (here ASCL1 for the horizontal axis and POU2F3 for the vertical axis). Other options include choosing the "Range" for the axes and the "Select Subsets" to "include" or "exclude". The + sign and folder indicate that further selections can be made such as choosing the "Dataset" to display. To display more than one dataset, users need to press the Command key while clicking the individual datasets (lower inset). Other features of SCLC_TumorMinerCDB include: (2) "Multivariable Analyses" with examples shown in panel B and Figure 2; (3) "Expression Distribution" across the 4 Datasets with example shown in Figure S1; (4) "Mutation variants"; (5) "Prognosis Biomarkers" with example shown in Figures 5 and S4; (6) "Predictive Biomarkers" with example shown in Figure 6; (7) "My Patient" to display the genomic characteristics of a selected patient with example shown in Figures 7 and S6. Additional features include the ability to "Download Data" for further analyses (8) as shown in Figures 7, S1, S4 and S5, "Search IDs" (9), as well as a detailed "Help" section (10). **B**. NAPY distribution of samples from untreated and relapse patients. Samples were assigned to each NAPY group based on gene expression log2(FPKM+1) > 5.0 (see horizontal and vertical dotted lines in panel A).

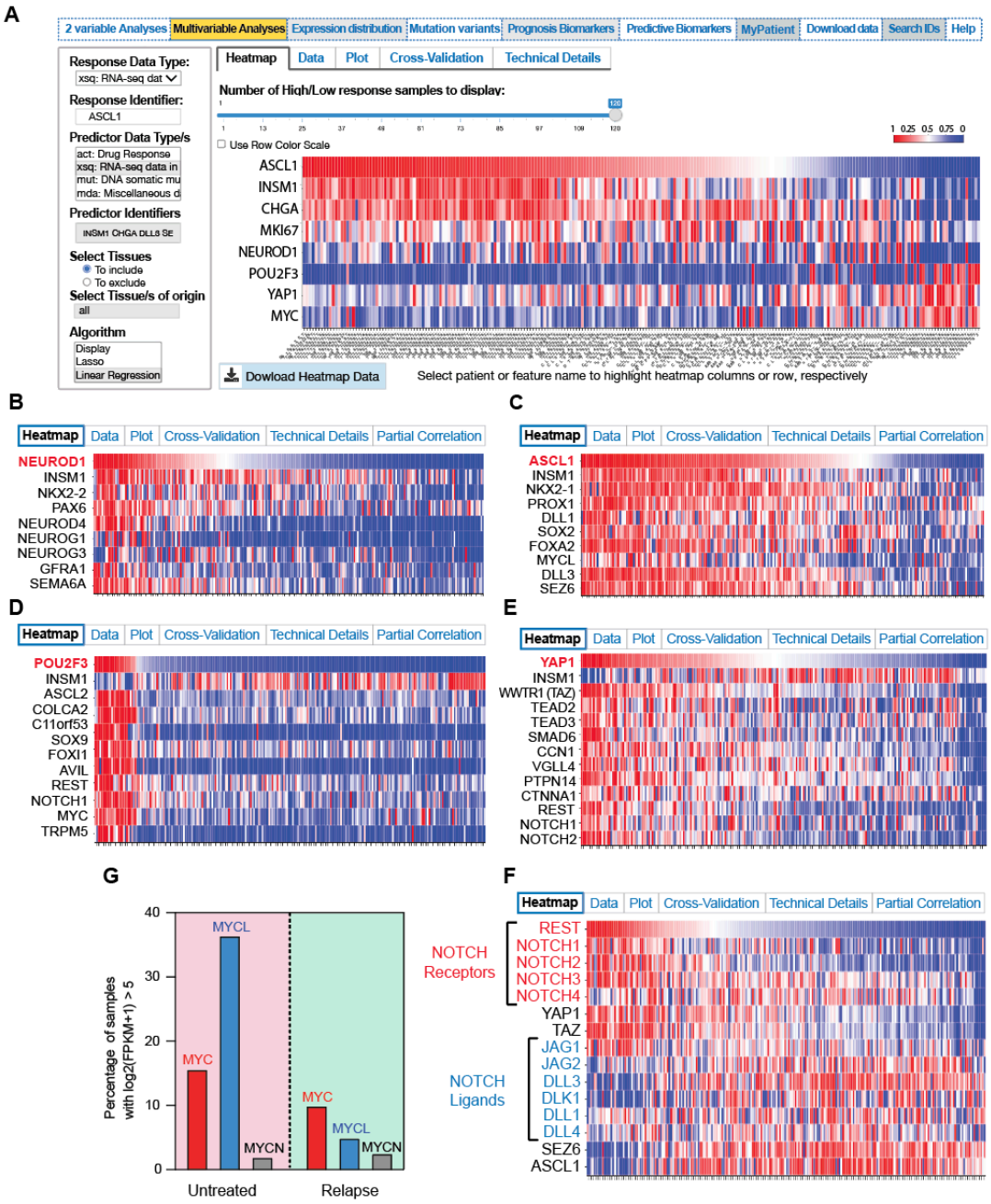


Figure 2. Multivariable Analyses showing co-regulation of the neuroendocrine (INSM1, CHGA), NAPY (NEUROD1, ASCL1, POU2F3 and YAP1), NOTCH (NOTCH1, NOTCH2, REST) and MYC (MYC, MYCL and MYCN) pathways. **A**. Default display for the "Multivariable Analyses" tab. Color scale goes from red to blue to reflect expression levels for each gene (rows) and patients (columns). Users can choose different data types (upper inset of panel A) and then type the "Response Identifier" to display as the first row on the heatmap. "Predictor Identifier (s)" need to be typed with a space between each. "Tissues of origin" can be selected. The default algorithm (bottom left) is "Display". Users can also perform Linear and Lasso analyses like in CellMinerCDB (https://discover.nci.nih.gov/cellminercdb/). **B-E**. Multivariable analyses for the NAPY genes used as "Response Identifiers". INSM1 is included in each analysis as a NE biomarker. **F**. Multivariable analysis for the Notch pathway. **G**. Differential expression of the 3 MYC genes in samples from untreated and relapse patients. Linear regression analyses for each of "Response Identifiers" (in red, 1st rows in panels B-F) show that the "Predictor Identifiers" are highly significant (Pearson correlations > 0.85 and p-values > 2.5e-65; see Figure S2C).

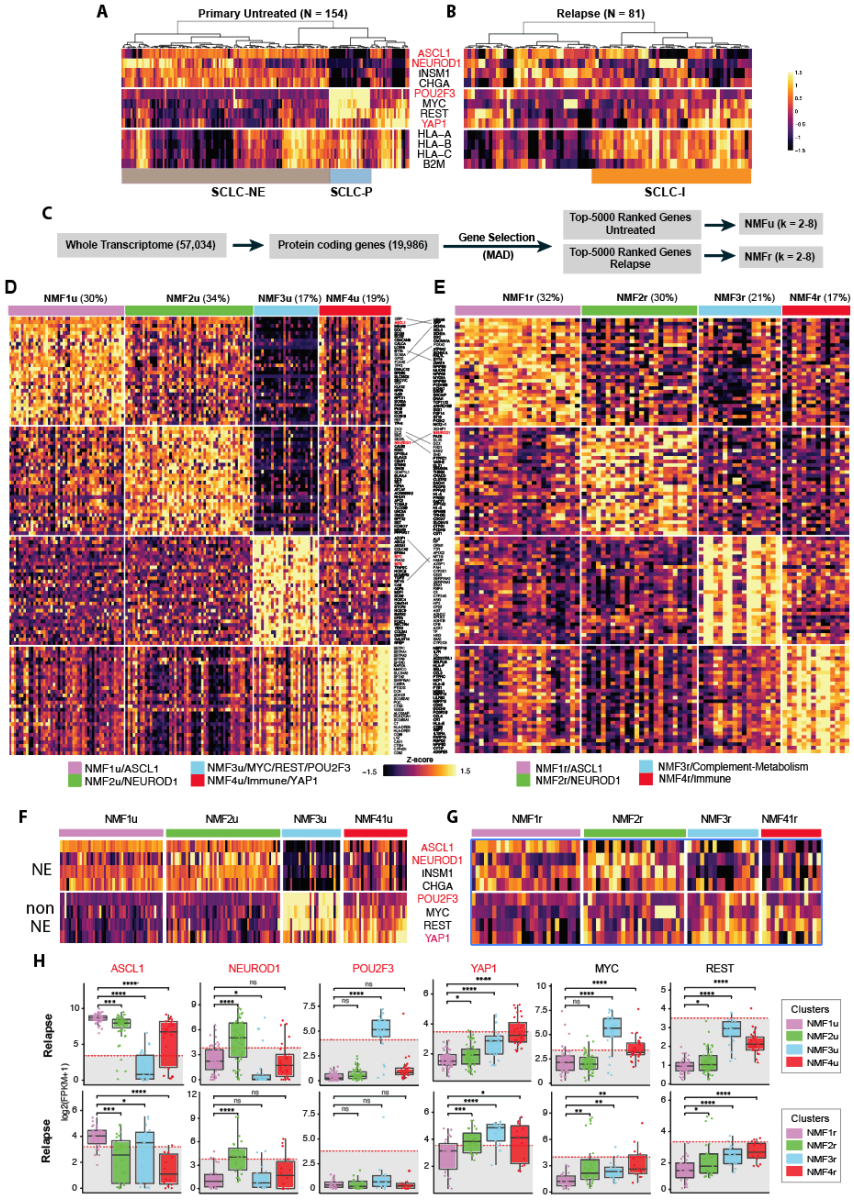
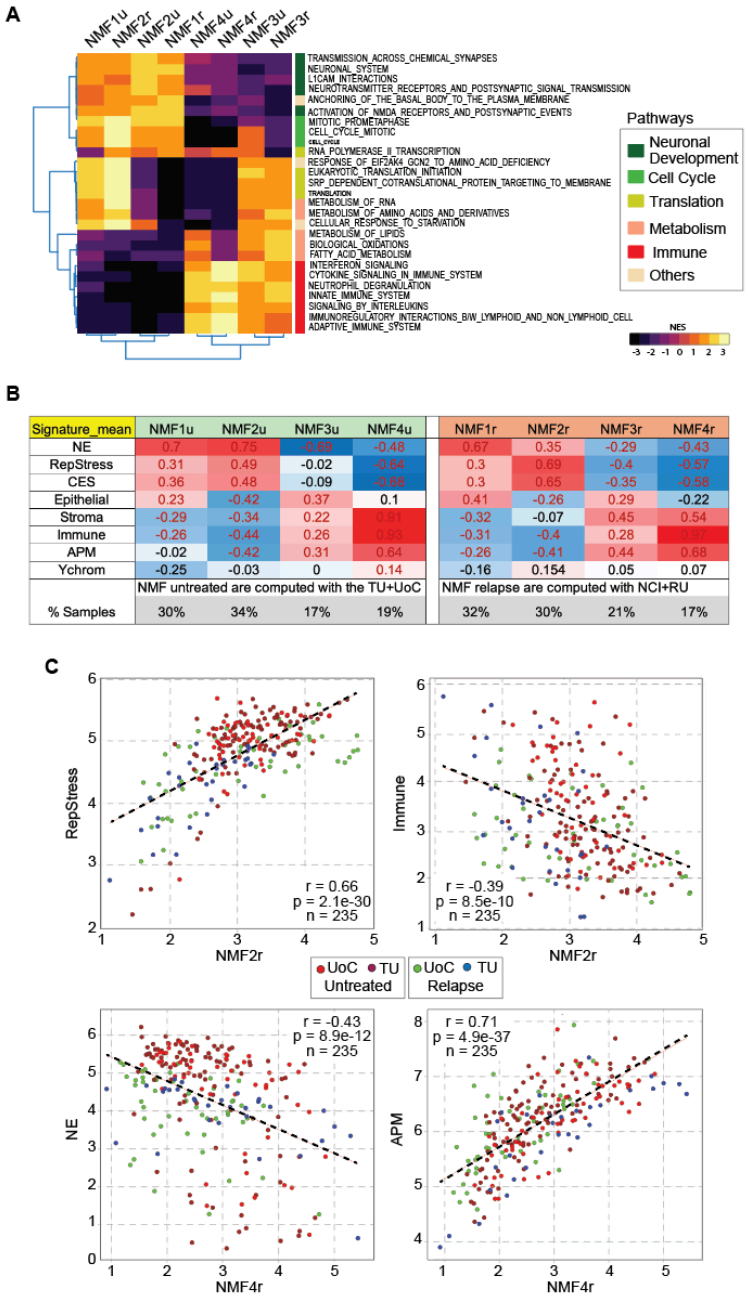
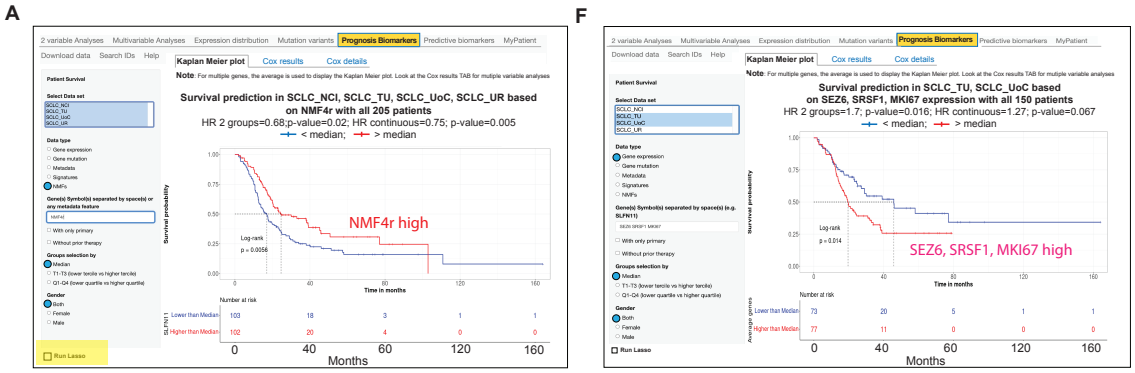


Figure 3. Classification of the untreated (primary) and relapse 245 SCLC samples comparing untreated (UoC and TU datasets) and relapse samples (NCI and RU). **A-B**. Hierarchical clustering of the untreated and relapse samples based on the expression of the four genes of the NAPY classification (NEUROD1, ASCL1, POU2F3 and YAP1 in red) and of genes reflecting commonly used biomarkers for clinical neuroendocrine (NE) classification (INSM1 and CHGA) as well as MYC, REST and 4 exemplary innate immune genes (HLA-A, B, C and B2M). **C**. Flow chart summarizing the NMF approach for selecting 5,000 protein-coding transcripts based on MAD (Median Absolute Deviation) for the untreated and relapse samples. **D-E**. Display of the NMF clusters for the untreated and relapse samples (D and E, respectively). The percentages of samples within each NMF are indicated in parenthesis. NMFs for untreated and relapse are abbreviated NMFu and NMFr, respectively. The rows show the top 30 genes for each NMF with log fold-change > 1 and p-value < 0.05. Only 8 genes (13%) are common across untreated and relapse samples for the NMF1 and NMF2 clusters, as indicated by the oblique lines. **F-G**. NMF distribution of the canonical NAPY genes (red) and reference NE genes as well as MYC and REST. **H**. Box plots comparing the differential expression of the NAPY genes, MYC and REST across the 4 untreated (top) and relapse NMFs (bottom). Statistical comparisons were performed by Student t-test. **** indicates $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Shaded area represent low expression (log2(FPKM+1) <3).



Signature_mean	NMF1u	NMF2u	NMF3u	NMF4u	NMF1r	NMF2r	NMF3r	NMF4r
NE	0.7	0.75	-0.69	-0.48	0.67	0.35	-0.29	-0.43
RepStress	0.31	0.49	-0.02	-0.84	0.3	0.69	-0.4	-0.57
CES	0.36	0.48	-0.09	-0.88	0.3	0.65	-0.35	-0.58
Epithelial	0.23	-0.42	0.37	0.1	0.41	-0.26	0.29	-0.22
Stroma	-0.29	-0.34	0.22	0.91	-0.32	-0.07	0.45	0.54
Immune	-0.26	-0.44	0.26	0.93	-0.31	-0.4	0.28	0.67
APM	-0.02	-0.42	0.31	0.64	-0.26	-0.41	0.44	0.68
Ychrom	-0.25	-0.03	0	0.14	-0.16	0.154	0.05	0.07
	NMF untreated are computed with the TU+UoC				NMF relapse are computed with NCI+RU			
% Samples	30%	34%	17%	19%	32%	30%	21%	17%

Figure 4. Genomic signatures associated with each of the NMF's from the untreated and relapse patient biopsy samples. **A**. Cluster image map of the 8 NMFs (4 untreated and 4 relapse) and top Reactome pathways. **B**. Correlation between each of the NMF's and the eight genomic signatures of SCLC_TumorMinerCDB for the samples from untreated (left) and relapse (right) patients. Correlations were obtained using the "2 variable Analyses" of the SCLC_TumorMinerCDB website ("Plot Data" or "View Data" tabs for each of the NMF on the horizontal axis). For the untreated samples the corresponding NMF's were computed after limiting the analysis using "Select Subsets" for the untreated Datasets: TU + UoC. Conversely for the relapse samples, analyses were done using the corresponding relapse samples: NCI + UR. **C**. Snapshots of "2 Variable Analyses" using the 4 datasets (inset) and displaying the correlations between the NMF2r signature and the RepStress (upper left) and the Immune signatures (upper right). The bottom 2 panels show the correlations between NMF4r and the NE (bottom left) and APM (bottom right) signatures.



B

	NMF untreated			
	NMF1u	NMF2u	NMF3u	NMF4u
Untreated	0.023	0.089	0.640	0.590
Relapse	0.089	0.140	0.190	0.350
All	0.061	0.510	0.049	0.220

C

NMF relapse			
NMF1r	NMF2r	NMF3r	NMF4r
0.200	0.045	0.120	0.045
0.440	0.630	0.400	0.095
0.080	0.390	0.170	0.022

D

	Signatures		
	APM	Immune	Stroma
Untreated	0.049	0.011	0.200
Relapse	0.078	0.130	0.560
All	0.027	0.004	0.012

E

	Genes				
	SLFN11	SEZ6	SRSF1	MKI67	SEZ6+SRSF1+MKI67
Untreated	0.190	0.052	0.018	0.170	0.011
Relapse	0.870	0.880	0.430	0.470	0.960
All	0.006	0.170	0.820	0.064	0.490

Blue = Low risk Red = High risk

Number are p-values for Hazard Ration (HR) 2 groups

Figure 5. Exploration of prognosis biomarkers. Panels A and F are screen snapshot of the SCLC_TumorMinerCDB website. **A**. High NMF4r expression is a good prognosis biomarker for all 4 Data sets. **B-E**. Tabulated results for the prognosis value of each of the NMFs (panels B-C), for 3 signatures (panel D) and genes (panel E) for each of the patient group. Combined expression of SEZ6, SRSF1 and MKI67 is a bad prognosis biomarker for the combined untreated Data sets (TU and UoC). Numbers are p-values for Hazard Ratio (HR) and colors highlight significant value ($p < 0.05$) with low-risk parameter in blue and high-risk in red. **F**. Gene expression of more than one gene can be tested. Here expression of SEZ6, the splicing factor SRSF1 and the proliferation marker Ki67 are poor prognosis biomarkers in untreated patients (TU and UoC). The option "Run Lasso" highlighted in yellow at the bottom left of panel A enables the unbiased search of novel prognosis biomarkers among selected Gene Sets lists that are displayed upon clicking "Run Lasso". Users are invited to run this feature on the website.

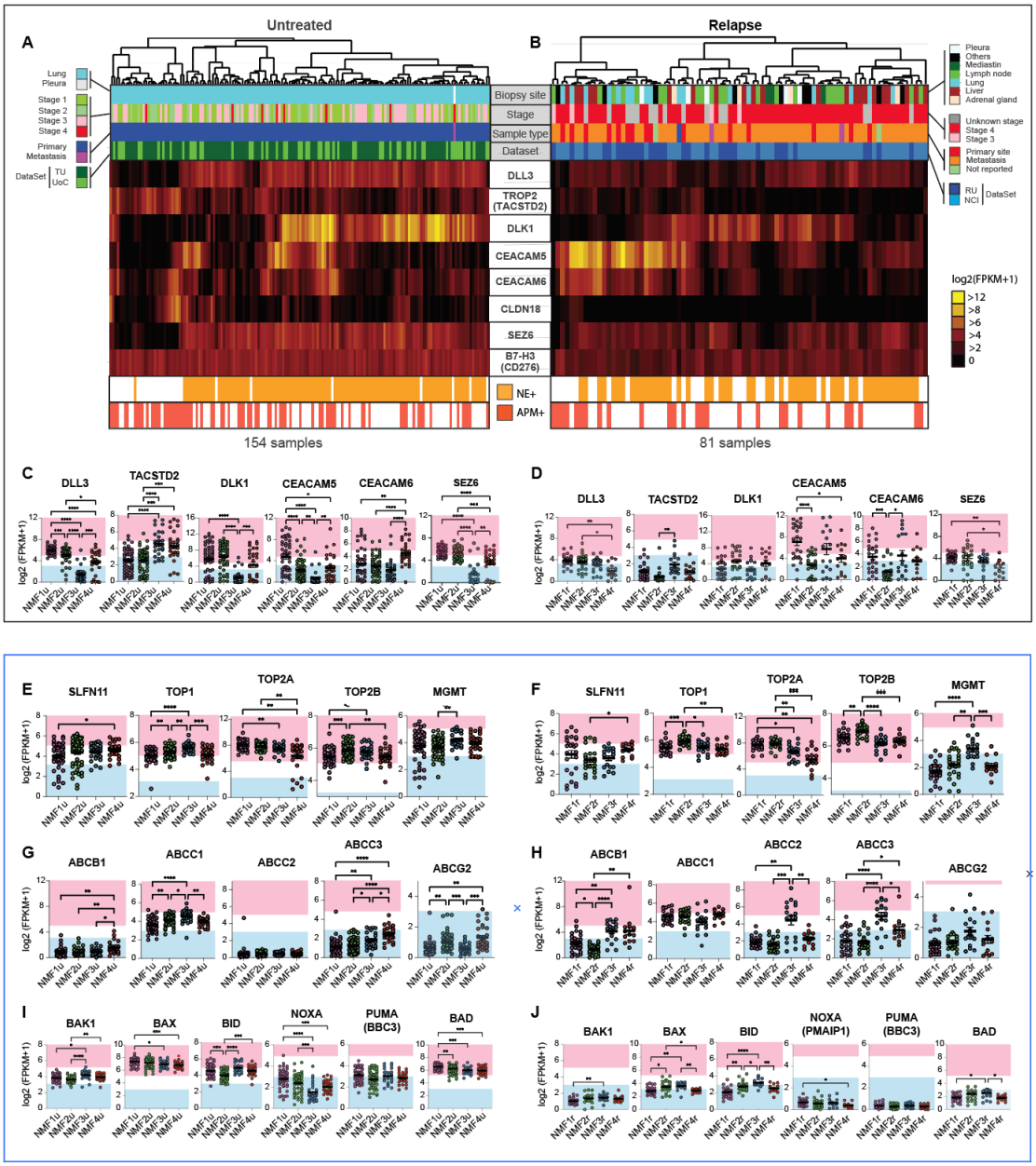


Figure 6. Predictive biomarkers with focus on ADCs for untreated patients (left panels A, C, E, G & I) versus patients after relapse (right panels B, D, F, H & J). **A-B**. Heatmaps showing the differential expression of 8 selected ADC targets. Samples were clustered based on the ADC target surface receptor genes listed between panels A and B. DLL3 and TROP2 (TACSTD2) are both FDA-approved targets. DLK1, CEACAM5, CEACAM6, CLN18, SIZ6 and B7-H3 are displayed because of their selective expression in subsets of patients. The top parts of the figure show for each patient the biopsy site, clinical stage, sample type and dataset. The bottom part shows for each patient whether the sample is enriched for the neuroendocrine (NE) and antigen-presenting-machinery (APM) signatures. **C-D**. Relationship between the listed ADC antibody targets and the NMF subgroups (NMFu and NMFr stand for the NMF's of samples from untreated and relapse patients, respectively. The blue-shaded areas include samples with low gene expression (log2(FPKM+1) <3) and the pink-shaded areas samples with high expression (log2(FPKM+1) >5). **E-F**. Expression of intracellular drug targets/biomarkers in the different NMF subgroups. **G-H**. Expression of drug efflux transporters in the different NMF subgroups. **I-J**. Expression of major pro-apoptotic genes in the different NMF subgroups. Statistical comparisons were performed by Student t-test. **** indicates $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

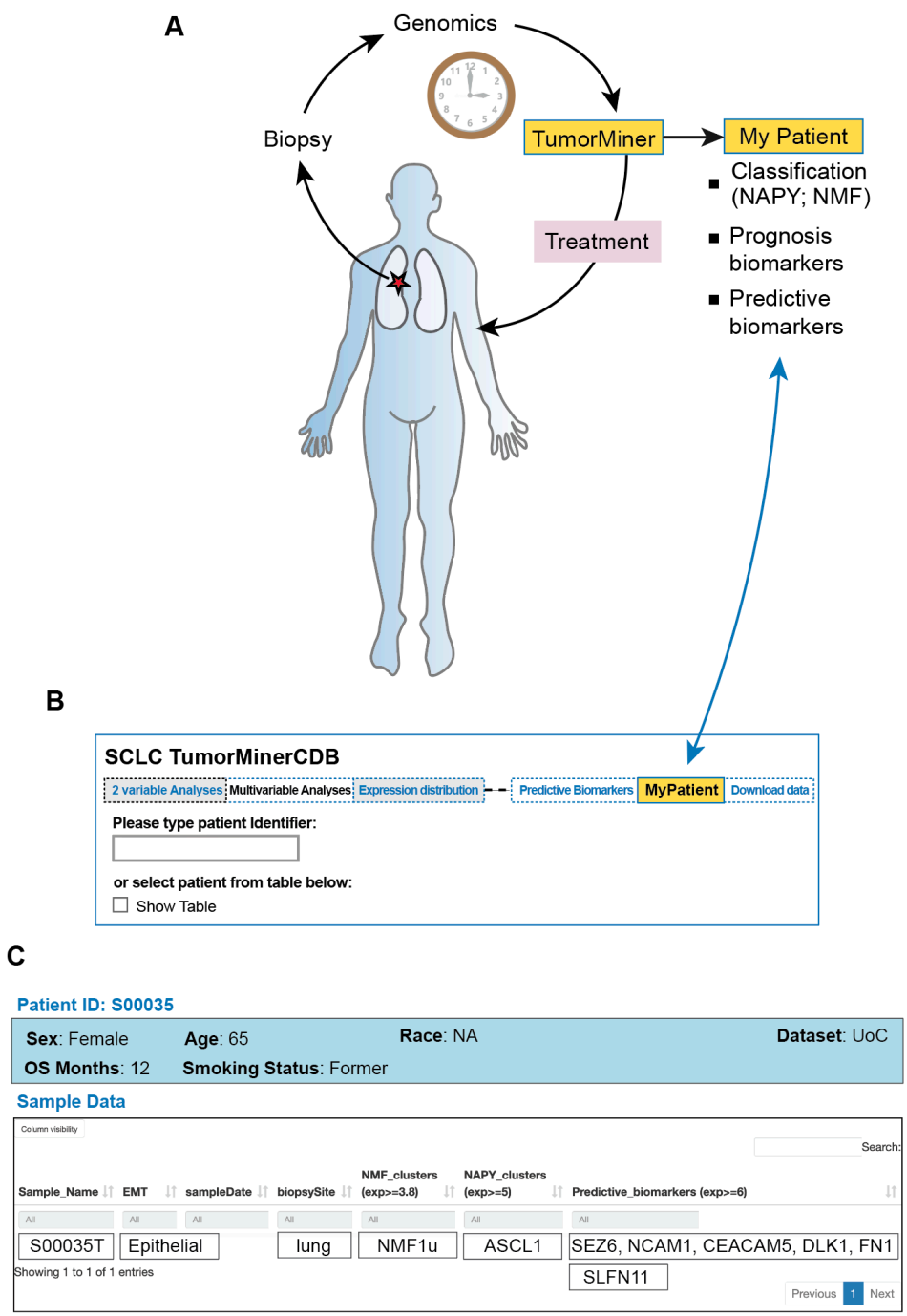


Figure 7. MyPatient module of SCLC_TumorMinerCDB. **A**. Proposed use of the "MyPatient" module. Genomic analyses are performed immediately after biopsy and uploaded into TumorMiner to provide results guiding the patient treatment as genomic classification, prognostic and predictive biomarkers are automatically displayed using "MyPatient" tab. **B**. Snapshot of the "MyPatient" tab in SCLC_TumorMinerCDB. Users are prompted to enter a "patient identifier" or select a patient from the available table. **C**. Snapshot of the "MyPatient" tab in SCLC_TumorMiner following the selection of a given patient. Here is an example for patient "S00035".

Supplemental Figures

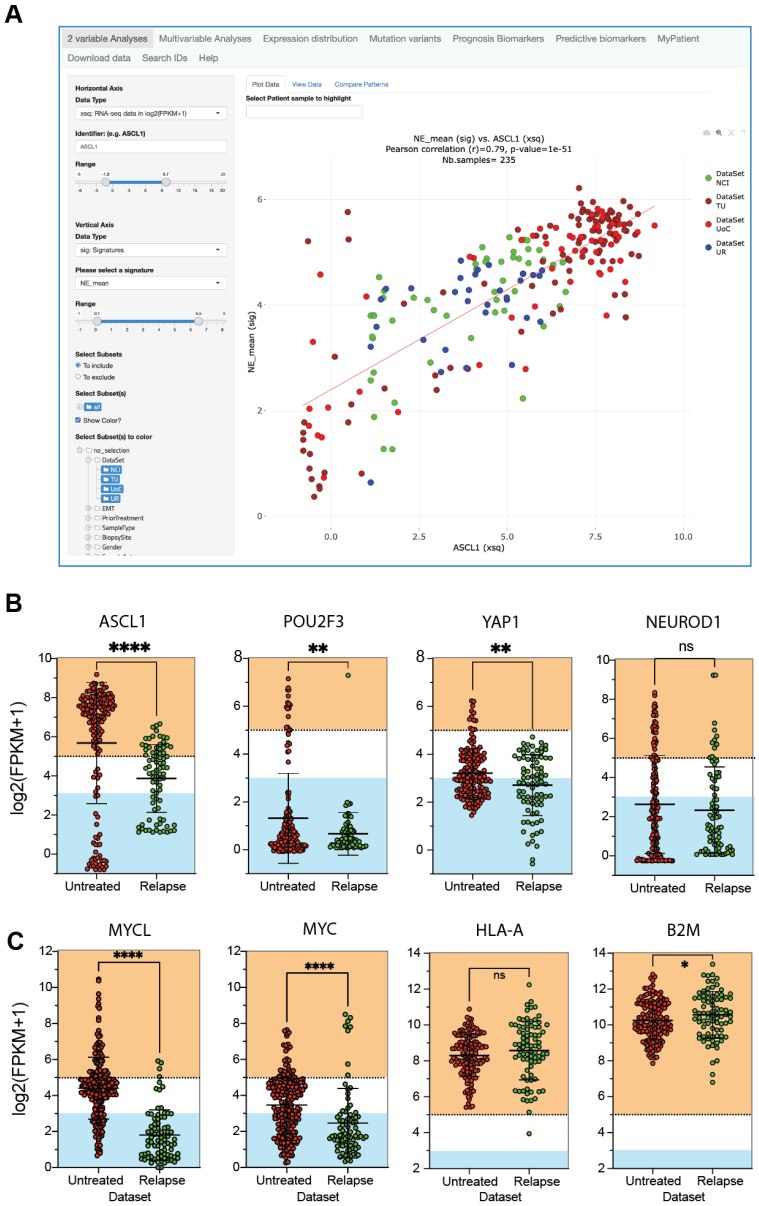
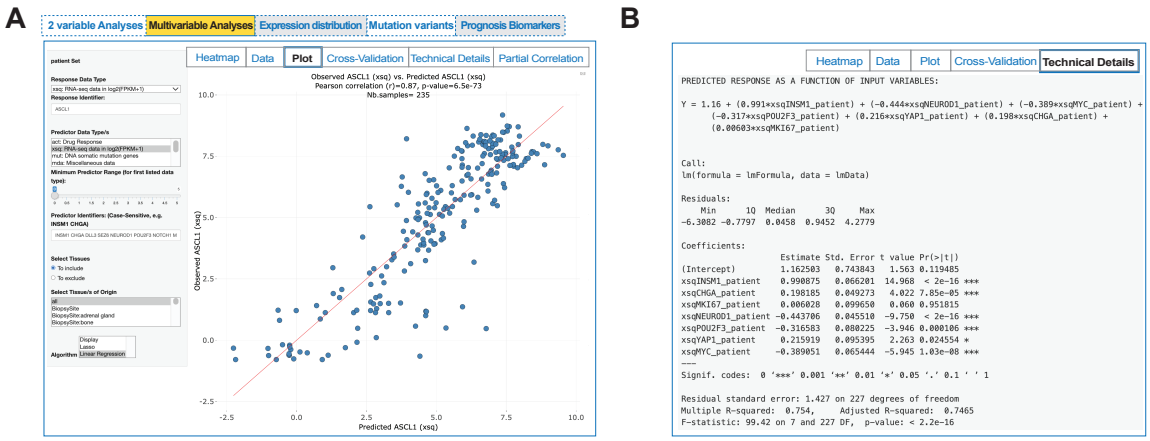


Figure S1 (related to Fig 1). **A**. Example of a "2 Variable Analysis" snapshot of the websites displaying the correlation between ASCL1 expression and the NE signature. **B**. Distribution and differential expression of the NAPY genes across the untreated (n = 154) and relapse (n = 81) patient samples. Colored areas define samples with significant expression ((log2(FPKM+1) > 5.0 in pink) and low expression (log2(FPKM+1) <3 in blue). Statistical comparisons were performed by Student t-test. **** indicates $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Data were downloaded from the SCLC_TumorMinerCDB website using the download tabs. **C**. Differential expression of the two MYC genes (MYCL, MYC) and of the two innate immune response genes HLA-A and B2M across the untreated (n = 154) and relapse (n = 81) patient samples. Colored areas define samples with significant expression ((log2(FPKM+1) > 5.0 in pink) and low expression (log2(FPKM+1) <3 in blue). Statistical comparisons were performed by Student t-test. **** indicates $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Data were downloaded from the SCLC_TumorMinerCDB website using the download tabs.



C

Predictive value for listed Response Identifiers

Response identifier	ASCL1 (panel A)	ASCL1 (panel B)	NEUROD1 (panel C)	POU2F3 (panel D)	YAP1 (panel E)	REST (panel F)
Predition all Tumors	r=0.87; p=6.5e-73	r=0.91; p=8.2e-92	r=0.85; p=2.6e-65	r=0.94; p=3.3e-112	r=0.89; p=1.9e-81	r=0.88; p=7.9e-77
Prediction all Cell lines	r=0.80; p=1.4e-26	r=0.92; p=3.8e-43	r=0.77; p=1.2e-23	r=0.96; p=1.1e-61	r=0.82; p=1.9e-29	r=0.90; p=4.7e-43

(r) = Pearson correlation

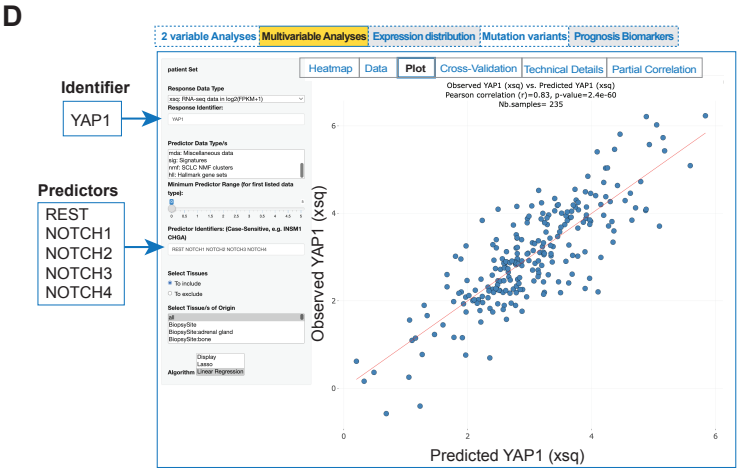
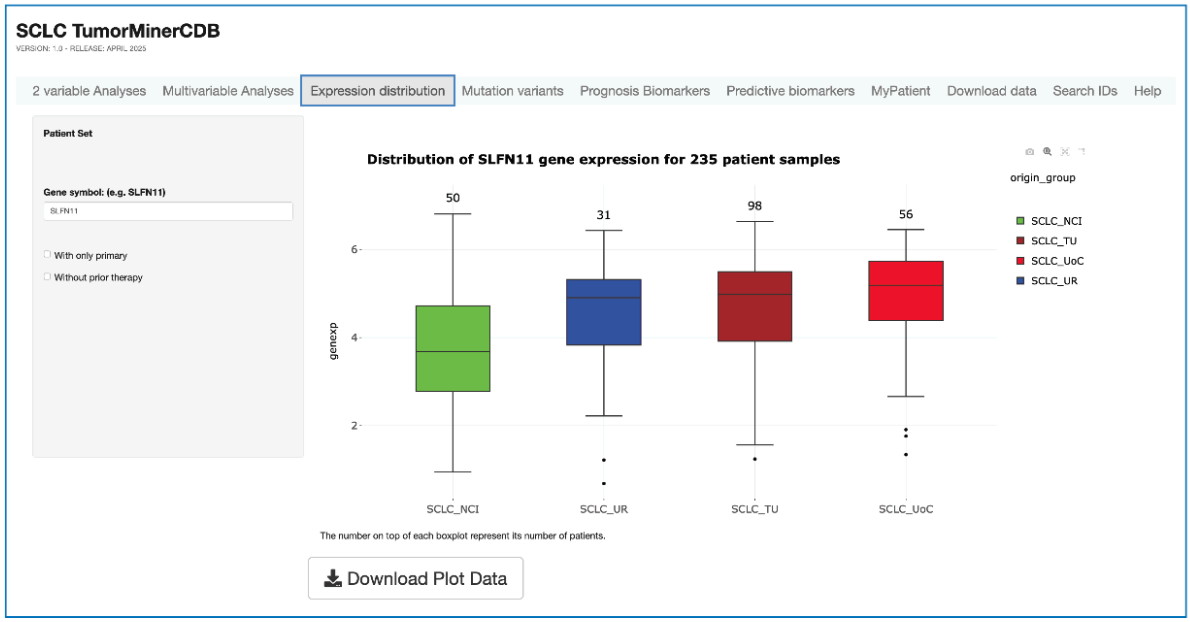


Figure S2 (related to Figs 1-2). **A**. Snapshot of the plot corresponding to the predictive value of the 8 variables (“Predictor Identifiers”) listed in Figure 2A under “Multivariable Analyses”. **B**. Snapshot of the “Technical Details” corresponding to the plot shown in panel A and the Heatmap shown in Figure 2A. **C**. Linear regression analyses for the Heatmaps shown in panels B-E of Figure 2. Pearson correlation coefficients and p-values were obtained from the “Plot” tab (as shown in panel A for Figure 2A). **D**. Expression of the Notch pathway canonical genes REST, NOTCH1-4 predict the expression of YAP1 (Pearson correlation (r) = 0.83; p-value = 2.4e-60).

A



B

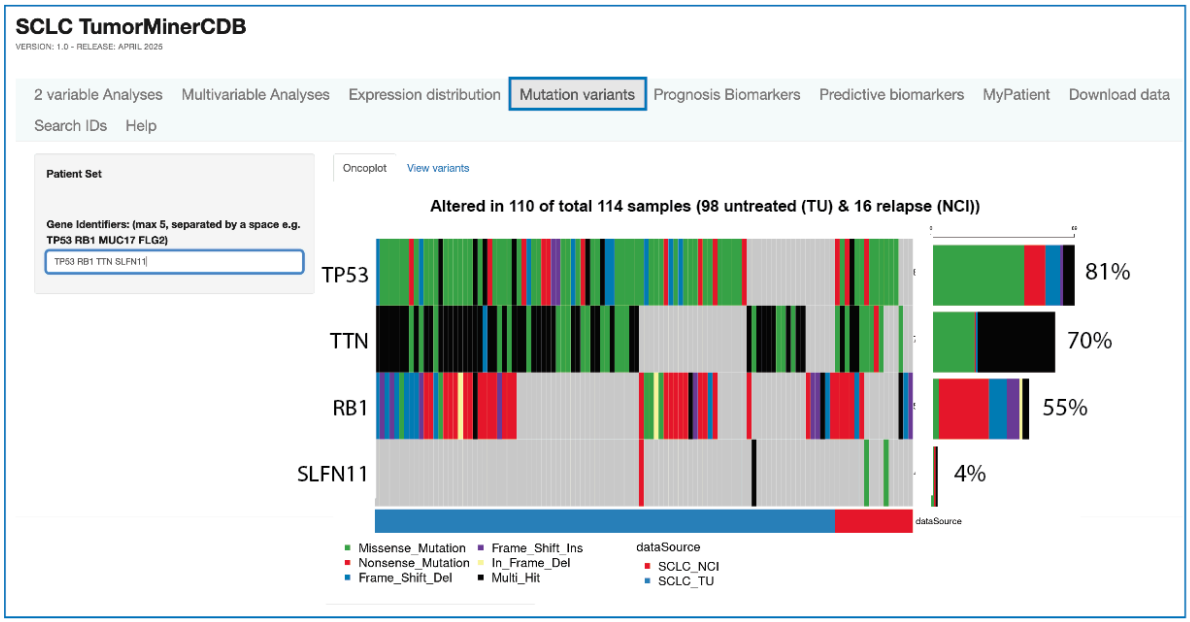


Figure S3 (related to Figs 1). **A**. Snapshot of the "Expression distribution" tab from the SCLC_TumorMinerCDB website in each of the 4 datasets of SCLC_TumorMinerCDB. Gene expression (genexp) is expressed as log2(FPKM+1). Statistical comparisons are performed by Student t-test. **** indicates $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. The "Download Plot Data" tab at the bottom left enables users to export the data. **B**. Snapshot of the "Mutation Variants" tab in the SCLC_TumorMinerCDB website. Number to the right indicate the percentage of samples with deleterious mutations.

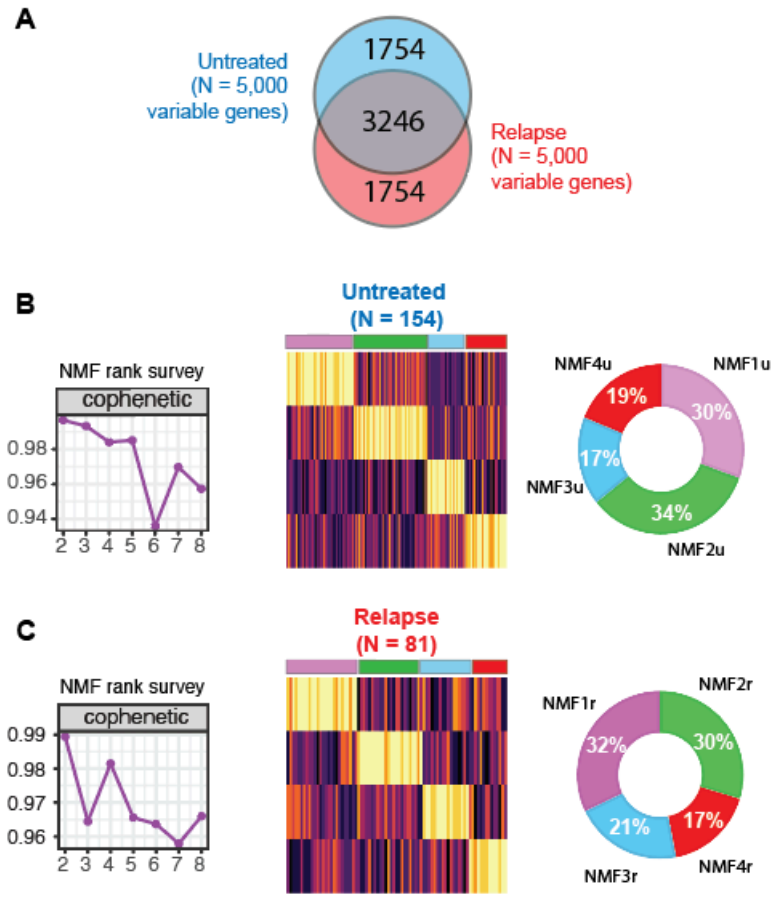


Figure S4 (related to Fig. 3). **A**. Venn-diagram showing the overlap of Untreated (top 5000) and Relapse (top 5000) most variable genes, selected using Median Absolute deviation (MAD) approach. **B** and **C** NMF rank summary for untreated and relapse SCLC samples, respectively. The cophenetic correlation coefficient is plotted for range of clusters (k=2:8), depicting k=4 as optimal number of clusters for both cohorts. Consensus matrix heatmaps at k=4 show well defined and stable clusters. Right: pie charts showing the distribution of samples in each NMF cluster.

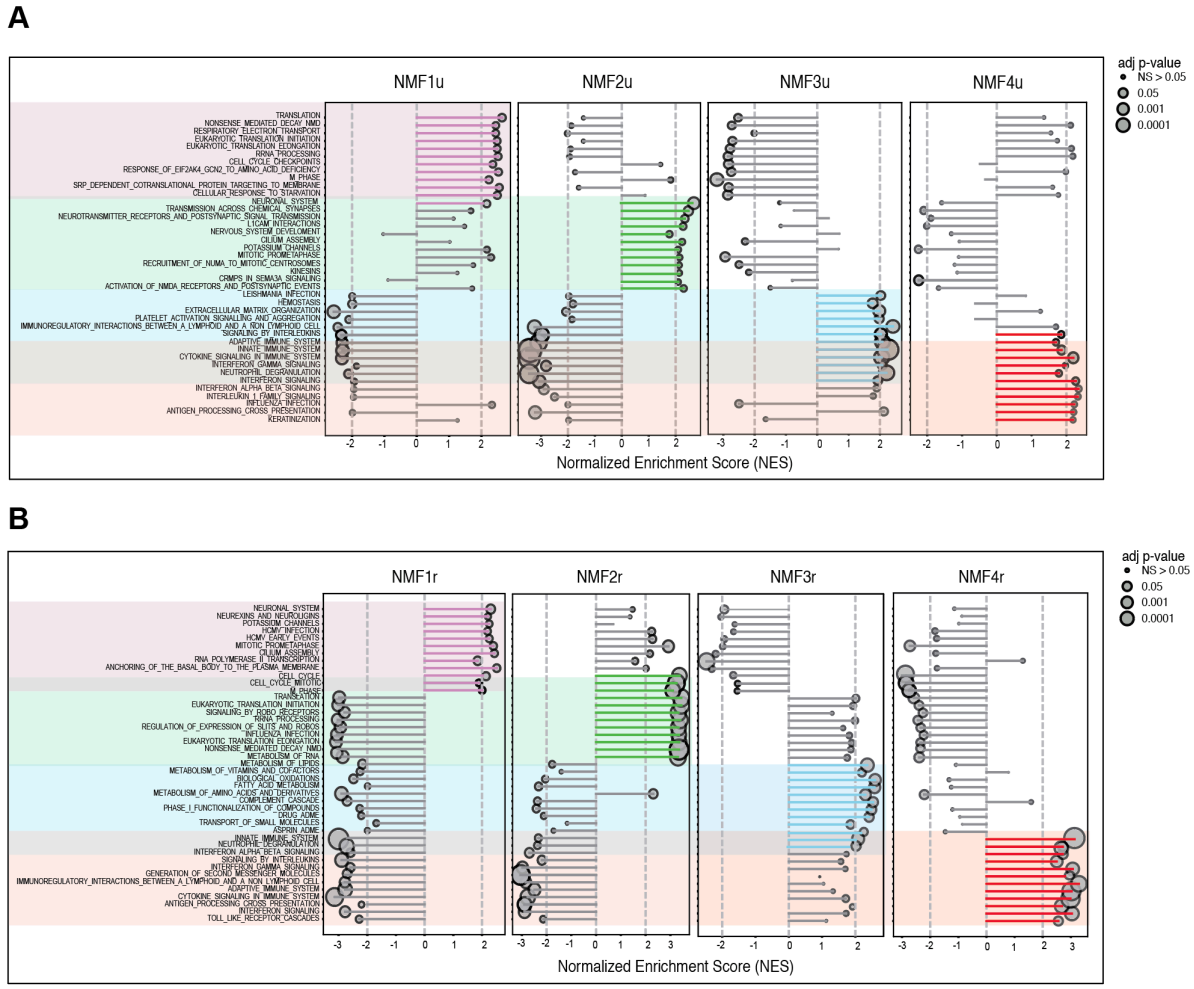


Figure S5 (matching Fig. 3). **A**. PrerankGSEA Reactome pathway analysis for the 4 NMF groups including the samples of untreated patients. **B**. PrerankGSEA Reactome pathway analysis for the 4 NMF groups including the samples of relapse patients.

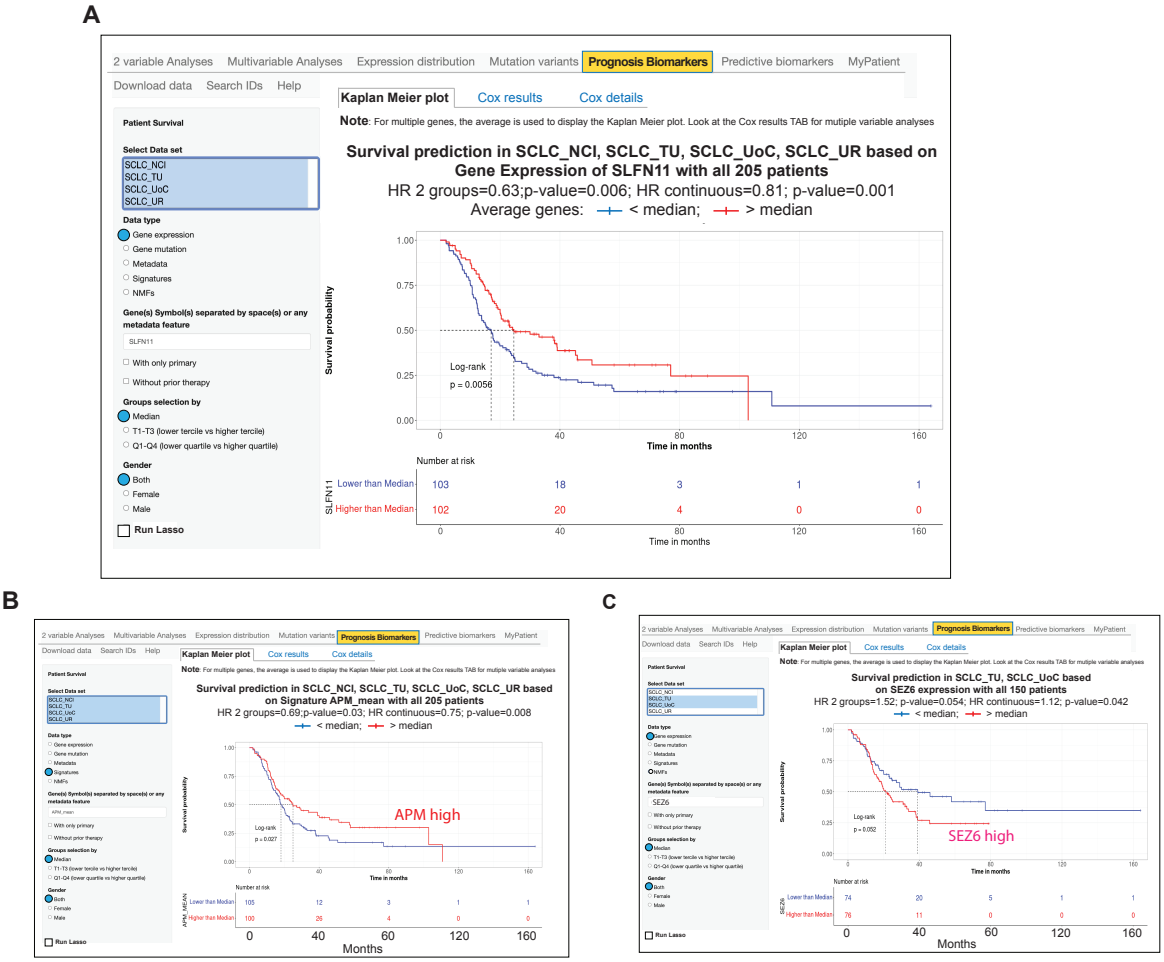


Figure S6 (related to Fig. 5). **A**. Good prognostic value of SLFN11 expression. **B**. Good prognostic value (risk factor) of the APM signature. **C**. Poor prognosis value (risk factor) on SEZ6 expression.

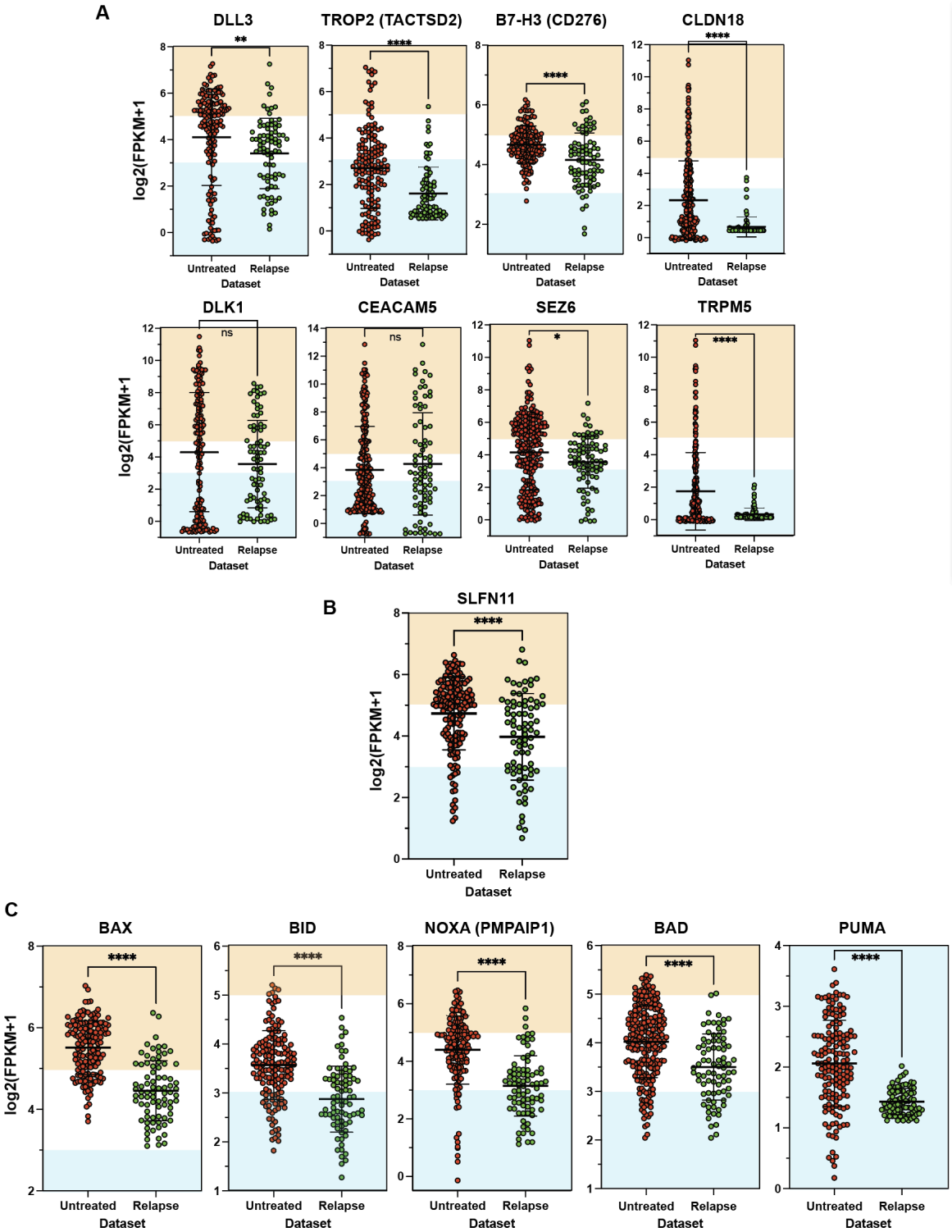


Figure S7 (related to Fig. 6). Distribution of predictive biomarkers across the 235 samples comparing samples from untreated (n = 154) and relapse (n = 81) patients. Colored areas defined samples with significant expression ((log2(FPKM+1) > 5.0 in pink) and low expression (log2(FPKM+1) <3 in blue). **A**. Surface biomarkers for ADCs. **B**. Broad range for SLFN11 expression across patient samples and downregulation at relapse. **C**. Broad expression range for proapoptotic genes across untreated patient samples and significant downregulation at relapse. Statistical comparisons were performed by Student t-test. **** indicates $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

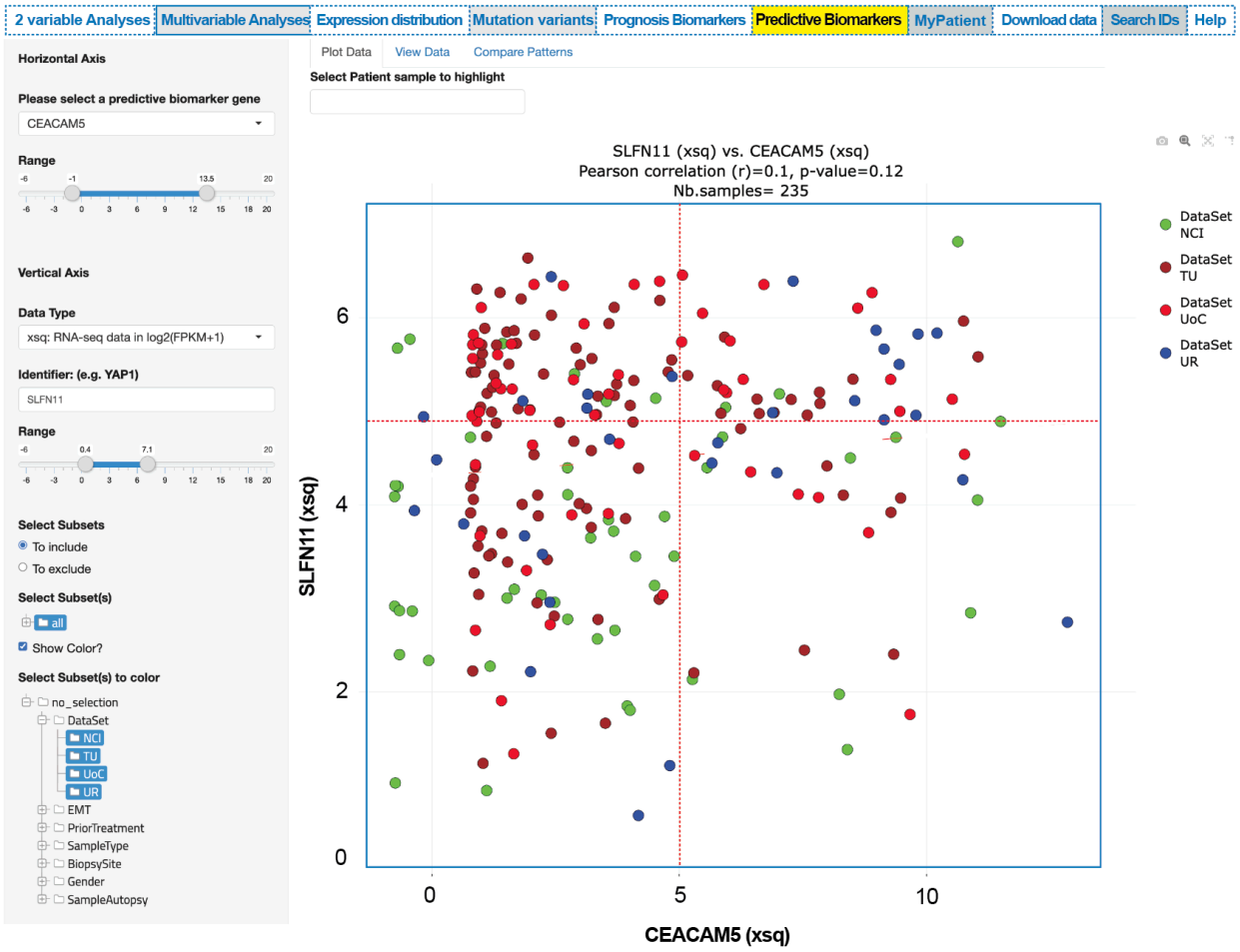
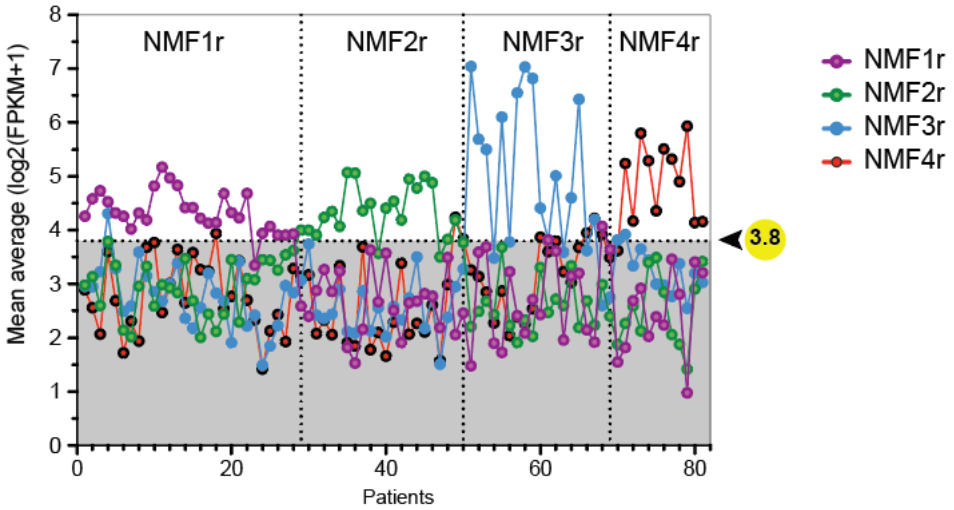


Figure S8 (related to Fig. 6). Expression of CEACAM5 and SLFN11 across the 235 samples of SCLC_TumorMiner. The figure is a snapshot of the website. The red dotted lines separate the samples with high gene expression (log2(FPKM+1) > 5.

937



938
939 **Figure S9 (related to Fig. 7)**. Mean average NMF expression for each and all 81 samples from patients at relapse. A
940 mean average threshold value log2(FPKM+1) > 3.8 is sufficient to assign each sample to one of the 4 NMF's. The
941 gray area defines non-significant values.
942

943 **REFERENCES**
944

945 1. Rudin, C.M., Poirier, J.T., Byers, L.A., Dive, C., Dowlati, A., George, J., Heymach, J.V.,
946 Johnson, J.E., Lehman, J.M., MacPherson, D., et al. (2019). Molecular subtypes of small
947 cell lung cancer: a synthesis of human and mouse model data. Nat Rev Cancer.
948 10.1038/s41568-019-0133-9.
949 2. Gay, C.M., Stewart, C.A., Park, E.M., Diao, L., Groves, S.M., Heeke, S., Nabet, B.Y.,
950 Fujimoto, J., Solis, L.M., Lu, W., et al. (2021). Patterns of transcription factor programs
951 and immune pathway activation define four major subtypes of SCLC with distinct
952 therapeutic vulnerabilities. Cancer Cell *39*, 346-360 e347. 10.1016/j.ccell.2020.12.014.
953 3. Qu, S., Fetsch, P., Thomas, A., Pommier, Y., Schrump, D.S., Miettinen, M.M., and Chen,
954 H. (2022). Molecular Subtypes of Primary SCLC Tumors and Their Associations With
955 Neuroendocrine and Therapeutic Markers. Journal of thoracic oncology : official
956 publication of the International Association for the Study of Lung Cancer *17*, 141-153.
957 10.1016/j.jtho.2021.08.763.
958 4. Pacheco, J., and Bunn, P.A. (2019). Advancements in Small-cell Lung Cancer: The
959 Changing Landscape Following IMpower-133. Clin Lung Cancer *20*, 148-160 e142.
960 10.1016/j.cllc.2018.12.019.
961 5. Paz-Ares, L., Dvorkin, M., Chen, Y., Reinmuth, N., Hotta, K., Trukhin, D., Statsenko, G.,
962 Hochmair, M.J., Ozguroglu, M., Ji, J.H., et al. (2019). Durvalumab plus platinum-
963 etoposide versus platinum-etoposide in first-line treatment of extensive-stage small-cell
964 lung cancer (CASPIAN): a randomised, controlled, open-label, phase 3 trial. Lancet *394*,
965 1929-1939. 10.1016/S0140-6736(19)32222-6.
966 6. Nabet, B.Y., Hamidi, H., Lee, M.C., Banchereau, R., Morris, S., Adler, L., Gayevskiy, V.,
967 Elhossiny, A.M., Srivastava, M.K., Patil, N.S., et al. (2024). Immune heterogeneity in
968 small-cell lung cancer and vulnerability to immune checkpoint blockade. Cancer Cell.
969 https://doi.org/10.1016/j.ccell.2024.01.010.
970 7. Tlemsani, C., Pongor, L., Elloumi, F., Girard, L., Huffman, K.E., Roper, N., Varma, S.,
971 Luna, A., Rajapakse, V.N., Sebastian, R., et al. (2020). SCLC-CellMiner: A Resource for
972 Small Cell Lung Cancer Cell Line Genomics and Pharmacology Based on Genomic
973 Signatures. Cell Rep *33*, 108296. 10.1016/j.celrep.2020.108296.
974 8. Ahn, M.-J., Cho, B.C., Felip, E., Korantzis, I., Ohashi, K., Majem, M., Juan-Vidal, O.,
975 Handzhiev, S., Izumi, H., Lee, J.-S., et al. (2023). Tarlatamab for Patients with Previously
976 Treated Small-Cell Lung Cancer. New England Journal of Medicine *389*, 2063-2075.
977 10.1056/NEJMoa2307980.
978 9. Su, P.L., Chakravarthy, K., Furuya, N., Brownstein, J., Yu, J., Long, M., Carbone, D., Li,
979 Z., and He, K. (2024). DLL3-guided therapies in small-cell lung cancer: from antibody-
980 drug conjugate to precision immunotherapy and radioimmunotherapy. Mol Cancer *23*, 97.
981 10.1186/s12943-024-02012-z.
982 10. Simpson, K.L., Rothwell, D.G., Blackhall, F., and Dive, C. (2025). Challenges of small
983 cell lung cancer heterogeneity and phenotypic plasticity. Nat Rev Cancer *25*, 447-462.
984 10.1038/s41568-025-00803-0.
985 11. Zhou, D., Byers, L.A., Sable, B., Smit, M.-A.D., Sadraei, N.H., Dutta, S., and Upreti, V.V.
986 (2024). Clinical Pharmacology Profile of AMG 119, the First Chimeric Antigen
987 Receptor T (CAR-T) Cell Therapy Targeting Delta-Like Ligand 3 (DLL3), in Patients with
988 Relapsed/Refractory Small Cell Lung Cancer (SCLC). The Journal of Clinical
989 Pharmacology *64*, 362-370. https://doi.org/10.1002/jcph.2346.

990 12. Rudin, C.M., Pietanza, M.C., Bauer, T.M., Ready, N., Morgensztern, D., Glisson, B.S.,
991 Byers, L.A., Johnson, M.L., Burris, H.A., 3rd, Robert, F., et al. (2017). Rovalpituzumab
992 tesirine, a DLL3-targeted antibody-drug conjugate, in recurrent small-cell lung cancer: a
993 first-in-human, first-in-class, open-label, phase 1 study. Lancet Oncol *18*, 42-51.
994 10.1016/S1470-2045(16)30565-4.
995 13. Thomas, A., and Pommier, Y. (2019). Targeting Topoisomerase I in the Era of Precision
996 Medicine. Clin Cancer Res *25*, 6581-6589. 10.1158/1078-0432.CCR-19-1089.
997 14. Zhang, B., Stewart, C.A., Wang, Q., Cardnell, R.J., Rocha, P., Fujimoto, J., Solis Soto,
998 L.M., Wang, R., Novegil, V., Ansell, P., et al. (2022). Dynamic expression of Schlafen 11
999 (SLFN11) in circulating tumour cells as a liquid biomarker in small cell lung cancer. Br J
1000 Cancer. 10.1038/s41416-022-01811-9.
1001 15. Jo, U., and Pommier, Y. (2022). Structural, molecular, and functional insights into Schlafen
1002 proteins. Experimental & Molecular Medicine. 10.1038/s12276-022-00794-0.
1003 16. Murai, J., Tang, S.-W., Leo, E., Baechler, S.A., Redon, C.E., Zhang, H., Al Abo, M.,
1004 Rajapakse, V.N., Nakamura, E., Jenkins, L.M.M., et al. (2018). SLFN11 Blocks Stressed
1005 Replication Forks Independently of ATR. Molecular Cell *69*, 371-384.e376.
1006 10.1016/j.molcel.2018.01.012.
1007 17. George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretic, L., Kong, G., Leenders, F., Lu, X.,
1008 Fernandez-Cuesta, L., Bosco, G., et al. (2015). Comprehensive genomic profiles of small
1009 cell lung cancer. Nature *524*, 47-53. 10.1038/nature14664.
1010 18. Liu, Q., Zhang, J., Guo, C., Wang, M., Wang, C., Yan, Y., Sun, L., Wang, D., Zhang, L.,
1011 Yu, H., et al. (2024). Proteogenomic characterization of small cell lung cancer identifies
1012 biological insights and subtype-specific therapeutic strategies. Cell *187*, 184-203 e128.
1013 10.1016/j.cell.2023.12.004.
1014 19. Lissa, D., Takahashi, N., Desai, P., Manukyan, I., Schultz, C.W., Rajapakse, V., Velez,
1015 M.J., Mulford, D., Roper, N., Nichols, S., et al. (2022). Heterogeneity of neuroendocrine
1016 transcriptional states in metastatic small cell lung cancers and patient-derived models. Nat
1017 Commun *13*, 2023. 10.1038/s41467-022-29517-9.
1018 20. Murai, J., Thomas, A., Miettinen, M., and Pommier, Y. (2019). Schlafen 11 (SLFN11), a
1019 restriction factor for replicative stress induced by DNA-targeting anti-cancer therapies.
1020 Pharmacol Ther *201*, 94-102. 10.1016/j.pharmthera.2019.05.009.
1021 21. Zoppoli, G., Regairaz, M., Leo, E., Reinhold, W.C., Varma, S., Ballestrero, A., Doroshow,
1022 J.H., and Pommier, Y. (2012). Putative DNA/RNA helicase Schlafen-11 (SLFN11)
1023 sensitizes cancer cells to DNA-damaging agents. Proc Natl Acad Sci U S A *109*, 15030-
1024 15035. 10.1073/pnas.1205943109.
1025 22. Pozo, K., Kollipara, R.K., Kelenis, D.P., Rodarte, K.E., Ullrich, M.S., Zhang, X., Minna,
1026 J.D., and Johnson, J.E. (2021). ASCL1, NKX2-1, and PROX1 co-regulate subtype-specific
1027 genes in small-cell lung cancer. iScience *24*, 102953.
1028 https://doi.org/10.1016/j.isci.2021.102953.
1029 23. Anderson, K.R., Torres, C.A., Solomon, K., Becker, T.C., Newgard, C.B., Wright, C.V.,
1030 Hagman, J., and Sussel, L. (2009). Cooperative transcriptional regulation of the essential
1031 pancreatic islet gene NeuroD1 (beta2) by Nkx2.2 and neurogenin 3. J Biol Chem *284*,
1032 31236-31248. 10.1074/jbc.M109.048694.
1033 24. Gosmain, Y., Marthinet, E., Cheyssac, C., Guerardel, A., Mamin, A., Katz, L.S., Bouzakri,
1034 K., and Philippe, J. (2010). Pax6 controls the expression of critical genes involved in

1035 pancreatic alpha cell differentiation and function. J Biol Chem *285*, 33381-33393.
1036 10.1074/jbc.M110.147215.
1037 25. Mastrantonio, R., You, H., and Tamagnone, L. (2021). Semaphorins as emerging clinical
1038 biomarkers and therapeutic targets in cancer. Theranostics *11*, 3262-3277.
1039 10.7150/thno.54023.
1040 26. Huang, Y.H., Klingbeil, O., He, X.Y., Wu, X.S., Arun, G., Lu, B., Somerville, T.D.D.,
1041 Milazzo, J.P., Wilkinson, J.E., Demerdash, O.E., et al. (2018). POU2F3 is a master
1042 regulator of a tuft cell-like variant of small cell lung cancer. Genes Dev *32*, 915-928.
1043 10.1101/gad.314815.118.
1044 27. Montoro, D.T., Haber, A.L., Biton, M., Vinarsky, V., Lin, B., Birket, S.E., Yuan, F., Chen,
1045 S., Leung, H.M., Villoria, J., et al. (2018). A revised airway epithelial hierarchy includes
1046 CFTR-expressing ionocytes. Nature *560*, 319-324. 10.1038/s41586-018-0393-7.
1047 28. Ireland, A.S., Hawgood, S.B., Xie, D.A., Barbier, M.W., Lucas-Randolph, S., Tyson, D.R.,
1048 Zuo, L.Y., Witt, B.L., Govindan, R., Dowlati, A., et al. (2024). Basal cell of origin resolves
1049 neuroendocrine–tuft lineage plasticity in cancer. bioRxiv, 2024.2011.2013.623500.
1050 10.1101/2024.11.13.623500.
1051 29. Wu, X.S., He, X.Y., Ipsaro, J.J., Huang, Y.H., Preall, J.B., Ng, D., Shue, Y.T., Sage, J.,
1052 Egeblad, M., Joshua-Tor, L., and Vakoc, C.R. (2022). OCA-T1 and OCA-T2 are
1053 coactivators of POU2F3 in the tuft cell lineage. Nature *607*, 169-175. 10.1038/s41586-022-
1054 04842-7.
1055 30. Abdel Wadood, N., Hollenhorst, M.I., Elhawy, M.I., Zhao, N., Englisch, C., Evers, S.B.,
1056 Sabachvili, M., Maxeiner, S., Wyatt, A., Herr, C., et al. (2025). Tracheal tuft cells release
1057 ATP and link innate to adaptive immunity in pneumonia. Nature Communications *16*, 584.
1058 10.1038/s41467-025-55936-5.
1059 31. Baine, M.K., Febres-Aldana, C.A., Chang, J.C., Jungbluth, A.A., Sethi, S., Antonescu,
1060 C.R., Travis, W.D., Hsieh, M.S., Roh, M.S., Homer, R.J., et al. (2022). POU2F3 in SCLC:
1061 Clinicopathologic and Genomic Analysis With a Focus on Its Diagnostic Utility in
1062 Neuroendocrine-Low SCLC. Journal of thoracic oncology : official publication of the
1063 International Association for the Study of Lung Cancer *17*, 1109-1121.
1064 10.1016/j.jtho.2022.06.004.
1065 32. Shue, Y.T., Drainas, A.P., Li, N.Y., Pearsall, S.M., Morgan, D., Sinnott-Armstrong, N.,
1066 Hipkins, S.Q., Coles, G.L., Lim, J.S., Oro, A.E., et al. (2022). A conserved
1067 YAP/Notch/REST network controls the neuroendocrine cell fate in the lungs. Nat Commun
1068 *13*, 2690. 10.1038/s41467-022-30416-2.
1069 33. Kim, J.W., Ko, J.H., and Sage, J. (2022). DLL3 regulates Notch signaling in small cell
1070 lung cancer. iScience *25*, 105603. 10.1016/j.isci.2022.105603.
1071 34. Ireland, A.S., Micinski, A.M., Kastner, D.W., Guo, B., Wait, S.J., Spainhower, K.B.,
1072 Conley, C.C., Chen, O.S., Guthrie, M.R., Soltero, D., et al. (2020). MYC Drives Temporal
1073 Evolution of Small Cell Lung Cancer Subtypes by Reprogramming Neuroendocrine Fate.
1074 Cancer Cell *38*, 60-78 e12. 10.1016/j.ccell.2020.05.001.
1075 35. Masuda, K., Yoshida, T., Motoi, N., Shinno, Y., Matsumoto, Y., Okuma, Y., Goto, Y.,
1076 Horinouchi, H., Yamamoto, N., Watanabe, S.-i., et al. (2025). Schlafen 11 Expression in
1077 Patients With Small Cell Lung Cancer and Its Association With Clinical Outcomes.
1078 Thoracic Cancer *16*, e15529. https://doi.org/10.1111/1759-7714.15529.
1079 36. Jiang, L., Huang, J., Higgs, B.W., Hu, Z., Xiao, Z., Yao, X., Conley, S., Zhong, H., Liu,
1080 Z., Brohawn, P., et al. (2016). Genomic Landscape Survey Identifies SRSF1 as a Key

1081 Oncodriver in Small Cell Lung Cancer. PLoS Genet *12*, e1005895.
1082 10.1371/journal.pgen.1005895.
1083 37. Meric-Bernstam, F., Yuca, E., Evans, K.W., Zhao, M., Maejima, T., Karibe, T., Raso,
1084 M.G., Tang, X., Zheng, X., Rizvi, Y.Q., et al. (2025). Antitumor Activity and Biomarker
1085 Analysis for TROP2 Antibody–Drug Conjugate Datopotamab Deruxtecan in Patient-
1086 Derived Breast Cancer Xenograft Models. Clinical Cancer Research *31*, 573-587.
1087 10.1158/1078-0432.Ccr-24-1948.
1088 38. Blackhall, F., Jao, K., Greillier, L., Cho, B.C., Penkov, K., Reguart, N., Majem, M.,
1089 Nackaerts, K., Syrigos, K., Hansen, K., et al. (2021). Efficacy and Safety of
1090 Rovalpituzumab Tesirine Compared With Topotecan as Second-Line Therapy in DLL3-
1091 High SCLC: Results From the Phase 3 TAHOE Study. Journal of thoracic oncology :
1092 official publication of the International Association for the Study of Lung Cancer *16*, 1547-
1093 1558. 10.1016/j.jtho.2021.02.009.
1094 39. Dowlati, A., Chiang, A.C., Cervantes, A., Babu, S., Hamilton, E., Wong, S.F., Tazbirkova,
1095 A., Sullivan, I.G., van Marcke, C., Italiano, A., et al. (2025). Phase 2 Open-Label Study of
1096 Sacituzumab Govitecan as Second-Line Therapy in Patients With Extensive-Stage SCLC:
1097 Results From TROPiCS-03. Journal of thoracic oncology : official publication of the
1098 International Association for the Study of Lung Cancer. 10.1016/j.jtho.2024.12.028.
1099 40. Sun, N.Y., Kumar, S., Kim, Y.S., Varghese, D., Mendoza, A., Nguyen, R., Okada, R.,
1100 Reilly, K., Widemann, B., Pommier, Y., et al. (2024). Identification of DLK1, a Notch
1101 ligand, as an immunotherapeutic target and regulator of tumor cell plasticity and
1102 chemoresistance in adrenocortical carcinoma. bioRxiv. 10.1101/2024.10.09.617077.
1103 41. Han, K., Chen, Y., Sun, X., Wen, L., Wu, Y., Chen, S., Wei, L., Yu, J., Zeng, T., Jiang, L.,
1104 and Tan, L. (2025). Combining serum CDK1 with tumor markers for the diagnosis of small
1105 cell lung cancer. Clin Transl Oncol *27*, 2005-2013. 10.1007/s12094-024-03722-y.
1106 42. Muley, T., Herth, F.J., Heussel, C.P., Kriegsmann, M., Thomas, M., Meister, M.,
1107 Schneider, M.A., Wehnl, B., Mang, A., and Holdenrieder, S. (2024). Prognostic value of
1108 tumor markers ProGRP, NSE and CYFRA 21-1 in patients with small cell lung cancer and
1109 chemotherapy-induced remission. Tumour Biol *46*, S219-S232. 10.3233/TUB-230016.
1110 43. Kudoh, S., Tenjin, Y., Kameyama, H., Ichimura, T., Yamada, T., Matsuo, A., Kudo, N.,
1111 Sato, Y., and Ito, T. (2020). Significance of achaete-scute complex homologue 1 (ASCL1)
1112 in pulmonary neuroendocrine carcinomas; RNA sequence analyses using small cell lung
1113 cancer cells and Ascl1-induced pulmonary neuroendocrine carcinoma cells. Histochem
1114 Cell Biol *153*, 443-456. 10.1007/s00418-020-01863-z.
1115 44. Ajay, A., Wang, H., Rezvani, A., Savari, O., Grubb, B.J., McColl, K.S., Yoon, S., Joseph,
1116 P.L., Kopp, S.R., Kresak, A.M., et al. (2025). Assessment of targets of antibody drug
1117 conjugates in SCLC. NPJ Precis Oncol *9*, 1. 10.1038/s41698-024-00784-7.
1118 45. Sun, F.D., and Das, M.S. (2024). Targeting the EZH2-SLFN11 pathway-a lesson in
1119 developing molecularly-informed treatments for recurrent small cell lung cancer.
1120 Translational Cancer Research *13*, 6608-6612. 10.21037/tcr-24-1755.
1121 46. Thomas, A., Tanaka, M., Trepel, J., Reinhold, W.C., Rajapakse, V.N., and Pommier, Y.
1122 (2017). Temozolomide in the Era of Precision Medicine. Cancer Res *77*, 823-826.
1123 10.1158/0008-5472.CAN-16-2983.
1124 47. Schultz, C.W., Zhang, Y., Elmeskini, R., Zimmermann, A., Fu, H., Murai, Y., Wangsa, D.,
1125 Kumar, S., Takahashi, N., Atkinson, D., et al. (2023). ATR inhibition augments the

efficacy of lurbinectedin in small-cell lung cancer. EMBO Molecular Medicine *15*, e17313. https://doi.org/10.15252/emmm.202217313.

48. Polley, E., Kunkel, M., Evans, D., Silvers, T., Delosh, R., Laudeman, J., Ogle, C., Reinhart, R., Selby, M., Connelly, J., et al. (2016). Small Cell Lung Cancer Screen of Oncology Drugs, Investigational Agents, and Gene and microRNA Expression. J Natl Cancer Inst *108*. 10.1093/jnci/djw122.
49. Chan, J.M., Quintanal-Villalonga, Á., Gao, V.R., Xie, Y., Allaj, V., Chaudhary, O., Masilionis, I., Egger, J., Chow, A., Walle, T., et al. (2021). Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. Cancer Cell *39*, 1479-1496.e1418. https://doi.org/10.1016/j.ccell.2021.09.008.
50. Stewart, C.A., Gay, C.M., Xi, Y., Sivajothi, S., Sivakamasundari, V., Fujimoto, J., Bolisetty, M., Hartsfield, P.M., Balasubramaniyan, V., Chalishazar, M.D., et al. (2020). Single-cell analyses reveal increased intratumoral heterogeneity after the onset of therapy resistance in small-cell lung cancer. Nature Cancer. 10.1038/s43018-019-0020-z.
51. Zhang, W., Girard, L., Zhang, Y.A., Haruki, T., Papari-Zareei, M., Stastny, V., Ghayee, H.K., Pacak, K., Oliver, T.G., Minna, J.D., and Gazdar, A.F. (2018). Small cell lung cancer tumors and preclinical models display heterogeneity of neuroendocrine phenotypes. Transl Lung Cancer Res *7*, 32-49. 10.21037/tlcr.2018.02.02.
52. Qi, J., Zhang, J., Liu, N., Zhao, L., and Xu, B. (2022). Prognostic Implications of Molecular Subtypes in Primary Small Cell Lung Cancer and Their Correlation With Cancer Immunity. Frontiers in oncology *12*, 779276. 10.3389/fonc.2022.779276.
53. Borromeo, M.D., Savage, T.K., Kollipara, R.K., He, M., Augustyn, A., Osborne, J.K., Girard, L., Minna, J.D., Gazdar, A.F., Cobb, M.H., and Johnson, J.E. (2016). ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. Cell Rep *16*, 1259-1272. 10.1016/j.celrep.2016.06.081.
54. Choudhury, N.J., Lai, W.V., Makhnin, A., Heller, G., Eng, J., Li, B., Preeshagul, I., Santini, F.C., Offin, M., Ng, K., et al. (2024). A Phase I/II Study of Valemetostat (DS-3201b), an EZH1/2 Inhibitor, in Combination with Irinotecan in Patients with Recurrent Small-Cell Lung Cancer. Clin Cancer Res *30*, 3697-3703. 10.1158/1078-0432.CCR-23-3383.
55. Tang, S.W., Thomas, A., Murai, J., Trepel, J.B., Bates, S.E., Rajapakse, V.N., and Pommier, Y. (2018). Overcoming Resistance to DNA-Targeted Agents by Epigenetic Activation of Schlafen 11 (SLFN11) Expression with Class I Histone Deacetylase Inhibitors. Clin Cancer Res *24*, 1944-1953. 10.1158/1078-0432.CCR-17-0443.
56. Rudin, C.M., Reck, M., Johnson, M.L., Blackhall, F., Hann, C.L., Yang, J.C., Bailis, J.M., Bebb, G., Goldrick, A., Umejiego, J., and Paz-Ares, L. (2023). Emerging therapies targeting the delta-like ligand 3 (DLL3) in small cell lung cancer. J Hematol Oncol *16*, 66. 10.1186/s13045-023-01464-y.
57. Paul, S., Konig, M.F., Pardoll, D.M., Bettegowda, C., Papadopoulos, N., Wright, K.M., Gabelli, S.B., Ho, M., van Elsas, A., and Zhou, S. (2024). Cancer therapy with antibodies. Nat Rev Cancer *24*, 399-426. 10.1038/s41568-024-00690-x.
58. Shen, L., Sun, X., Chen, Z., Guo, Y., Shen, Z., Song, Y., Xin, W., Ding, H., Ma, X., Xu, W., et al. (2024). ADCdb: the database of antibody-drug conjugates. Nucleic Acids Res *52*, D1097-D1109. 10.1093/nar/gkad831.
59. DeLucia, D.C., Cardillo, T.M., Ang, L., Labrecque, M.P., Zhang, A., Hopkins, J.E., De Sarkar, N., Coleman, I., da Costa, R.M.G., Corey, E., et al. (2021). Regulation of CEACAM5 and Therapeutic Efficacy of an Anti-CEACAM5-SN38 Antibody-drug

1172 Conjugate in Neuroendocrine Prostate Cancer. Clin Cancer Res *27*, 759-774.
1173 10.1158/1078-0432.CCR-20-3396.
1174 60. Du, Y., Wen, Y., and Huang, J. (2024). Analysis of variation of serum CEA, SCC,
1175 CYFRA21-1 in patients with lung cancer and their diagnostic value with EBUS-TBNA. J
1176 Med Biochem *43*, 363-371. 10.5937/jomb0-37083.
1177 61. Yu, M., and Wang, P. (2024). A comparative study of ProGRP and CEA as serological
1178 markers in small cell lung cancer treatment. Discov Oncol *15*, 485. 10.1007/s12672-024-
1179 01323-3.
1180 62. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools:
1181 efficient and comprehensive analysis of somatic variants in cancer. Genome Res *28*, 1747-
1182 1756. 10.1101/gr.239244.118.
1183