Predicting drug response of small cell lung cancer cell lines based on enrichment analysis of complex gene signatures

Kolos Nemes, Alexandra Benő, Petronella Topolcsányi, Éva Magó, Gabriella Mihalekné Fűr, Lőrinc S. Pongor *

Cancer Genomics and Epigenetics Core Group, Hungarian Center of Excellence for Molecular Medicine (HCEMM), Szeged, Hungary

ARTICLE INFO

ABSTRACT

Advances in the field of genomics and transcriptomics have enabled researchers to identify gene signatures related to development and treatment of Small Cell Lung Cancer. In most cases, complex gene expression patterns are identified, comprising of genes with differential behavior. Most tools use single-genes as predictors of drug response, with only limited options for multi-gene use. Here we examine the potential of predicting drug response using these complex gene expression signatures by employing clustering and signal enrichment in Small Cell Lung Cancer. Our results demonstrate clustering genes from complex expression patterns helps identify differential activity of gene groups with alternate function which can then be used to predict drug response.

1. Introduction

Small Cell Lung Cancer (SCLC) is a highly aggressive and rapidly progressing form of lung tumor. While historically its classified primarily based on its microscopic appearance, recent advances in genomics and molecular biology have revealed that SCLC is a complex disease comprised of distinct molecular subtypes (Govindan et al., 2006; Rudin et al., 2019). These subtypes are characterized by unique genetic alterations (George et al., 2015; Thomas et al., 2020; Tlemsani et al., 2021), transcriptional profiles (George et al., 2015; Lissa et al., 2022; Wooten et al., 2019), and phenotypic features (Borromeo et al., 2016; Huang et al., 2018; Pongor et al., 2023), which have significant implications for diagnosis, prognosis, and treatment strategies (Gay et al., 2021; Roper et al., 2021; Schultz et al., 2023; Takahashi et al., 2020; Thomas et al., 2020, 2021; Tlemsani et al., 2021).

SCLC tumors are predominantly neuroendocrine (NE), characterized by using immunohistochemistry targeting markers such as chromogranin-A (CHGA) or synaptophysin (SYP). A small subset of SCLC tumors present low levels of NE markers, defined as non-neuroendocrine (non-NE) tumors (Gazdar et al., 2017; McColl et al., 2017). A recently proposed classification scores and differentiates tumors based on the expression signature of 50 genes, comprised of 25 NE and 25 non-NE specific genes (Zhang et al., 2018). The SCLC subtypes can be further classified based on the expression of four markers: achaete-scute homolog 1 (ASCL1), neurogenic differentiation factor 1

(NEUROD1), POU class 2 homeodomain box 3 (POU2F3) and Yes-associated protein 1 (YAP1), where the first two are enriched in NE tumors, while the latter two are enriched in non-NE tumors (Rudin et al., 2019; Tlemsani et al., 2020; Wooten et al., 2019). Recently, a new 'inflamed' subtype has been identified, enriched predominantly with non-NE tumors, which showed better response to chemo-immunotherapy treatment (Gay et al., 2021; Roper et al., 2021).

Cell line models are a cost-effective way to test tens- to hundreds of drugs. To this end, there several databases are available, where users can compare genomic features (gene expression, copy-number status, mutational status and methylation) with response to drugs or metastatic potential (Barretina et al., 2012; Garnett et al., 2012; Iorio et al., 2016; Jin et al., 2020; Polley et al., 2016; Pongor et al., 2022; Rajapakse et al., 2018; Tlemsani et al., 2020). In addition, there have been efforts to identify additional drivers by employing CRISPR and gene silencing screens, where growth rates were used as readouts (Tsherniak et al., 2017). With these tools, we can easily test if a candidate gene is essential, or if it correlates with other genetic features or drug response.

In many instances it becomes necessary to condense complex datasets into a singular vector of values. Such examples are correlation analyses, or comparing signal distribution between sample sets, used to identify therapeutic targets and drug combinations. Single-sample Gene Set Enrichment Analysis (ssGSEA) (Barbie et al., 2009), an extension of GSEA (Mootha et al., 2003; Subramanian et al., 2005), offers an elegant and straightforward method for assessing and comparing the activity of

---

molecular pathways. One challenge is utilization of gene expression patterns, where genes exhibit diverse behaviors, such as upregulation or downregulation, which can be further categorized based on their temporal dynamics. To address this, we employ gene clustering before performing ssGSEA analysis, which can be used to predict drug response. This preprocessing step aids in the identification of gene sets within these expression signatures, which, in turn, can be harnessed for predictive modeling of drug responses. We demonstrate the usefulness by interrogating commonly used gene expression signatures used for classification of SCLC.

2. Materials and methods

2.1. Data preparation

Comprehensive gene expression datasets, including those from the NCI, Broad Institute/MIT, and Sanger/Massachusetts General Hospital (MGH) (Barretina et al., 2012; Garnett et al., 2012; Iorio et al., 2016; Polley et al., 2016; Tlemsani et al., 2020), and SCLC patient data (Cerami et al., 2012; George et al., 2015), were acquired. Along with the gene expression data, associated annotations and drug response data were also downloaded. The datasets were downloaded from the Cell-MinerCDB and SCLC-CellMinerCDB websites (https://discover.nci.nih.gov/rsconnect/cellminercdb/, https://discover.nci.nih.gov/rsconnect/SclcCellMinerCDB/) (Rajapakse et al., 2018; Tlemsani et al., 2020). Gene signatures used for clustering and ssGSEA analysis in this study were: (1) the NE score gene set (Zhang et al., 2018), (2) Epithelial-mesenchymal Transition (EMT) genes (Kohn et al., 2014) and (3) inflamed signature (Gay et al., 2021).

2.2. Data analysis

Data manipulation and analysis were conducted using custom R (version 4.3.2). In the initial stages, data cleaning processes such as removing missing values (using the 'dropna' function) or subsetting datasets to include only common samples during cross comparison were carried out. We also subset datasets by gene signatures. In case of data visualization through heatmaps, a z-scoring procedure was applied to standardize the data, ensuring a uniform scale. The ssGSEA was performed utilizing the *GSVA* (version 1.48.3) package (Hänzelmann et al., 2013), with the 'ssGSEA' method specified. Additionally, k-means clustering was implemented via the *stats* package (version 4.3.2). For data visualization, we employed the *ComplexHeatmap* (version 2.16.0) package (Gu et al., 2016) and the *ggplot2* package (version 3.4.3). The EMT score was calculated using the EMTscore function from https://github.com/korkutlab/imogimap (Bozorgui et al., 2023), the NEscore was calculated based on (Zhang et al., 2018). Gene ontology analysis was calculated using the following packages: *clusterProfiler* (version 4.8.3) (Yu et al., 2012), *org.Hs.eg.db* (version 3.17.0), *enrichplot* (version 1.20.3), and DOSE (version 3.26.1) (Yu et al., 2015).

For the final phase of our analyses, we focused on determining the correlation between variables. This was achieved by employing the Spearman correlation method (ggplot2 package), a non-parametric measure. This approach allowed us to assess the strength and direction of the association between the datasets under study.

3. Results and discussion

3.1. Clustering genes prior to ssGSEA models

To demonstrate the advantages of clustering prior to ssGSEA analysis, we tested the clustering-based enrichment analysis using three distinct scenarios. Since ssGSEA compares the expression of the signature genes to the entire gene set, we simulated random expression matrices of 50 „samples" and 2k „genes", of which 50 genes were the interrogated sets.

In the first scenario, we investigated the impact of augmenting the signal within two gene sets that exhibited contrasting expression patterns between 'A samples' and 'B samples.' For each case, we introduced incremental changes of 0.01, 0.05, 0.1, 0.25, 0.5, and 1, equivalent to signal increases of 1%, 5%, 10%, 25%, 50%, and 100%, respectively (illustrative examples are presented in Fig. 1A). Notably, at the 0.1 increment level, discernible differences in gene expression between the two sample groups become visually evident, leading to noticeable variations in ssGSEA enrichments (Fig. 1B). On the other hand, when conducting the analysis using all genes without prior clustering, the opposing expression patterns of the two gene sets nullify the enrichment effect. This observation underscores the critical significance of the clustering step, which enables the capture of even a moderate difference in signal.

In the second scenario, we investigated the influence of negatively correlated genes on the enrichment score. To explore this, we initially constructed an expression matrix wherein the overall gene expression levels increased across the samples (referred to as 'Original'). Subsequently, we systematically reversed the expression patterns of 5, 25, 45, and eventually, all genes (see Figure S1A). As indicated by the ssGSEA enrichment score above the heatmaps, the reversal of only 5 genes had a relatively minor impact, with the enrichment score showing little change. However, when half of the genes (25 genes) exhibited reversed behavior, the enrichment effect was entirely nullified. Conversely, when genes were subjected to clustering, it notably emphasized the contrasting enrichment scores between the two modeled groups.

The third scenario focused on examining the impact of introducing noise to a gene signature, a situation often encountered in real-world scenarios where an identified gene set is tested on different datasets or data sources. The original matrix represented an expression dataset where the overall gene expression of the signature genes increased across the samples. In each modeling iteration, we randomly shuffled the order of either 5, 25, 45, or all genes (as illustrated in Fig. 1C). As an increasing number of genes were subjected to shuffling, the range of enrichment scores for all genes experienced a substantial reduction (as shown in Fig. 1D). Furthermore, the correlation between the enrichment scores of the shuffled signal and the original unshuffled signal significantly decreased (representative examples in Fig. 1D left, Figure S1B left for all cases). Conversely, when we employed gene clustering to isolate the actual signature genes, it helped preserve the correlation with the unshuffled signal (representative examples in Fig. 1D right, Figure S1B right for all cases).

3.2. Neuroendocrine score

A recently developed neuroendocrine score (NE score) (Zhang et al., 2018) relies on correlating the expression of samples to a gene signature comprising of 25 genes highly expressed in NE cells and 25 genes highly expressed in non-NE cells (Fig. 2A). Zhang and colleagues provide the mean expression of the signature for a set of neuroendocrine samples as well as non-neuroendocrine samples that serve as basis for the comparison. In short, the NE score of a sample is determined through a three-step process: 1) calculating the sample's gene expression correlation with both NE and non-NE gene sets, 2) deducting the correlation coefficient of the non-NE gene set from that of the NE gene set, and 3) dividing the value by two.

Since the gene signature is comprised of two gene categories, this allows the assessment of enrichment in each category through ssGSEA (TableS1). This results in a negative correlation between the two gene groups, as depicted in Fig. 2B. To verify the consistency of this method, we compared enrichment scores obtained from the NCI-SCLC dataset with those from the CCLE and GDSC databases, illustrated in Fig. 2C and Figure S2A. Our analysis across different datasets revealed correlation coefficients between 0.81 and 0.95, underscoring the effectiveness of ssGSEA. However, minor variations in results might be attributed to the heterogeneous cell morphologies in these samples (Krohn et al., 2014;
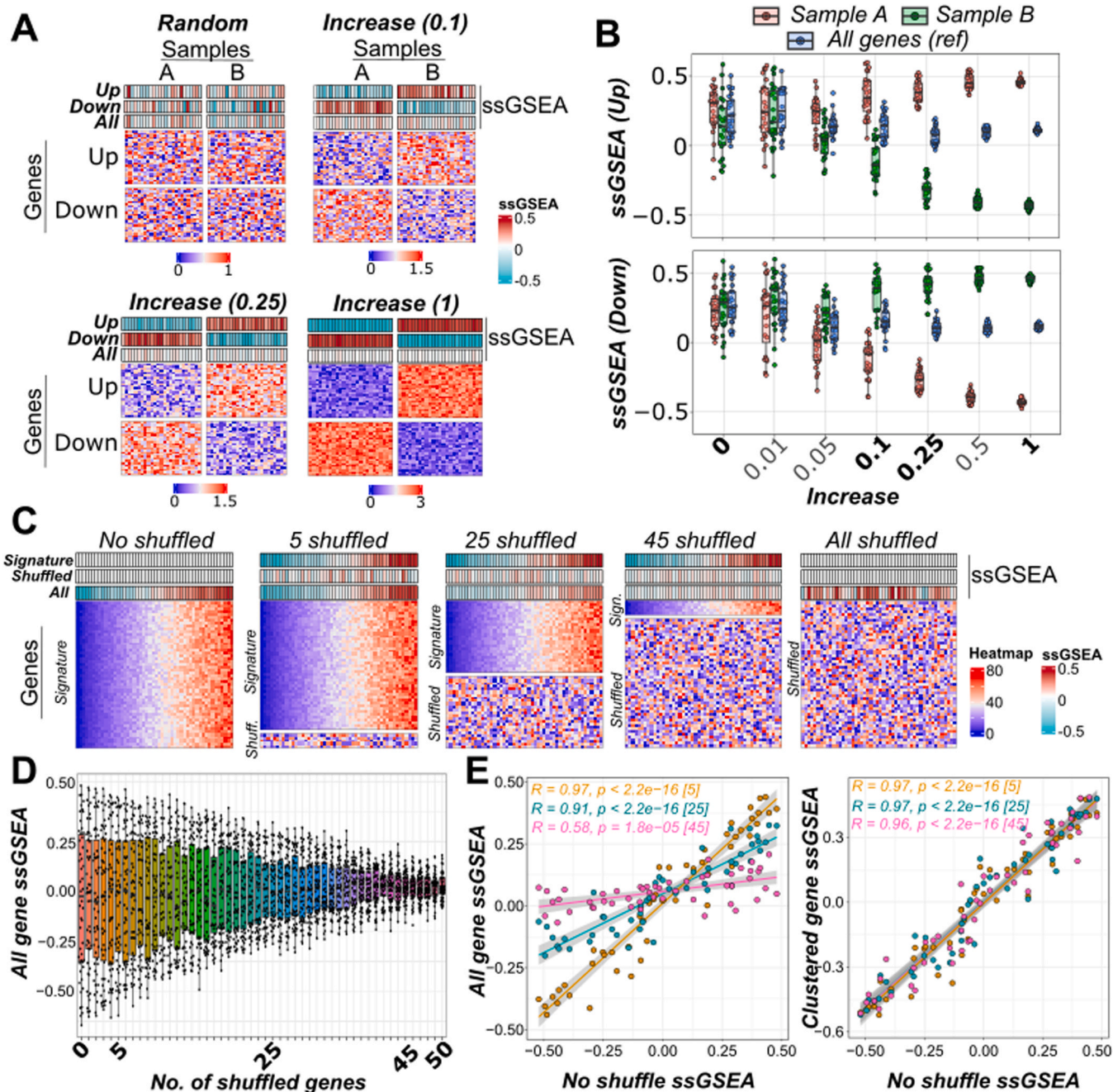
Fig. 1. Clustering and ssGSEA models. A) Heatmaps of random genes, where signal is increased in the two sample groups. B) ssGSEA score distribution in 'Upregulated' and 'Downregulated' genes. Boxplots are grouped based on columns: condition 'A' in red, condition 'B' in green, and for reference ssGSEA score was also calculated with all genes shown in blue. C) Heatmaps where genes are shuffled in increasing numbers. Top annotation shows the ssGSEA scores for all genes, shuffled genes only, or 'signature' genes only. D) ssGSEA score distribution comparison as more genes are shuffled. All genes are used in each case to calculate ssGSEA scores. E) Correlation comparison of ssGSEA signal for All genes (left) or clustered genes (right) when 5, 25 or 45 genes are shuffled with the ssGSEA signal of the completely unshuffled samples.

Pongor et al., 2023).

Comparing the proposed NE score with the enrichment scores, we found a strong positive correlation of 0.86 (p<e-16) when using the NE gene set, and a strong negative correlation of −0.89 (p<e-16) with the non-NE gene set enrichment score, as shown in Figure S2B. As anticipated, the ASCL1+ and NEUROD1+ cell lines displayed higher NE enrichment scores, consistent with their neuroendocrine characteristics (see Fig. 2D and Figure S2C). We also evaluated these scores in relation to the effectiveness of the mTOR inhibitor INK-128, previously shown to have higher activity on non-NE cells (Tlemsani et al., 2020). In line with expectations, the NE score and NE ssGSEA enrichment score showed a

positive correlation (Fig. 2E-F), while the non-NE enrichment score was inversely correlated with the response to INK-128 (Fig. 2G). These findings highlight that the gene signature clustering can effectively be used for calculating enrichment without the necessity for creating specialized scoring methods.

3.3. Epithelial-mesenchymal transition

Epithelial-mesenchymal transition (EMT) is an important process where epithelial cells acquire motile and invasive characteristics seen in mesenchymal cells. To assess EMT status, a gene set can be used that
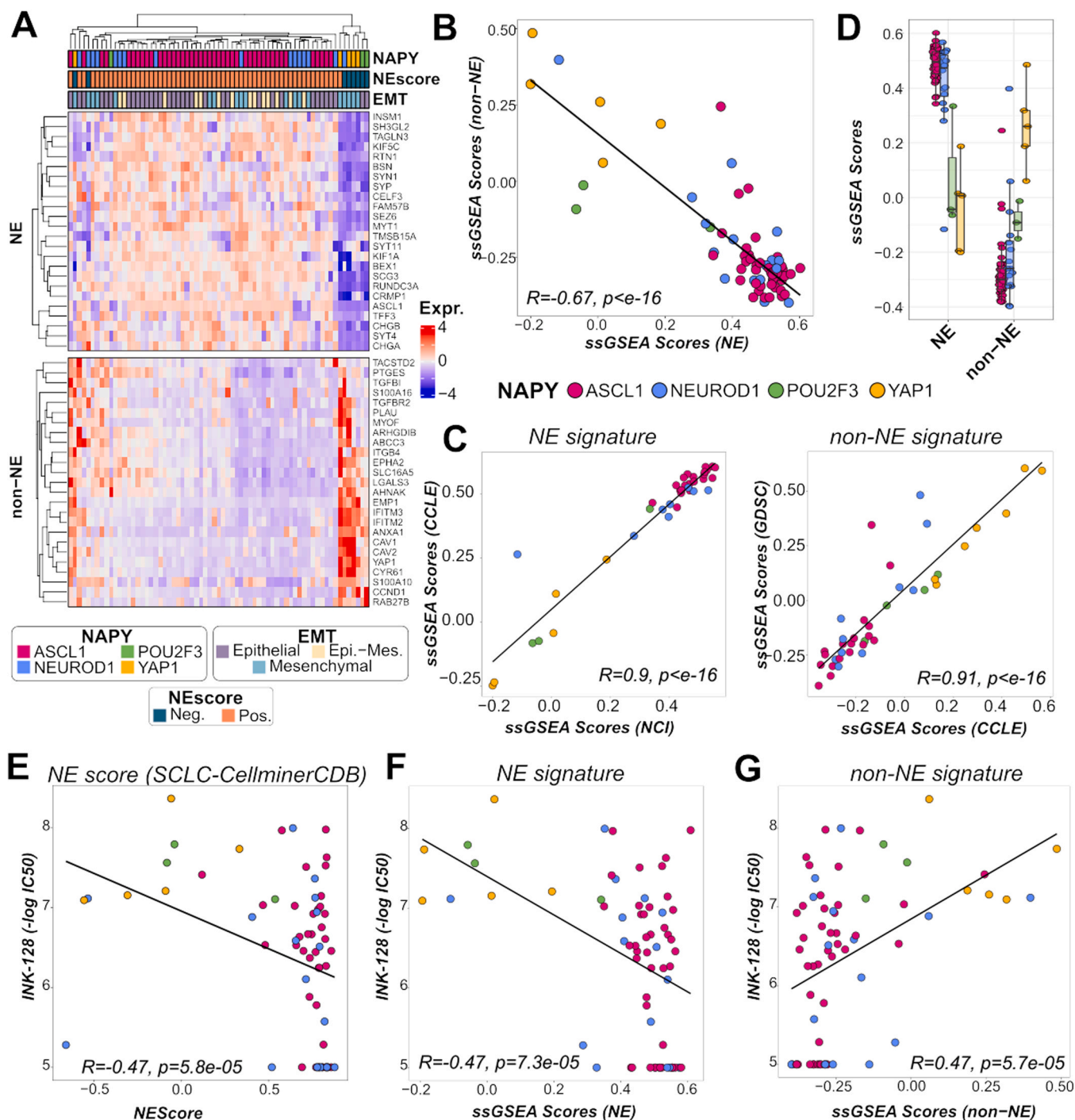
Fig. 2. Neuroendocrine score analysis. A) Heatmap of the NE signature genes. B) Correlation of the ssGSEA scores calculated for NE genes and non-NE genes. C) Distribution of ssGSEA scores in the four SCLC subtypes. D) Cross-correlation of ssGSEA scores between data sources. E-G) Correlation of INK-128 with the E) NE score, F) NE ssGSEA score and G) non-NE ssGSEA score.

includes genes predominantly expressed in either epithelial or mesenchymal states, as identified in studies (such as Byers et al., 2013; Kohn et al., 2014; Liberzon et al., 2015). The gene signature established by Kohn and colleagues is presented in Fig. 3A (Table S2). There was a negative correlation between the enrichment scores of these two gene sets, as shown in Fig. 3B. Notably, cell lines positive for ASCL1 and POU2F3 exhibited higher epithelial enrichment (Fig. 3C and Figure S3A). When correlating these scores across NCI, CCLE, and GDSC datasets, there was a strong correlation observed (with coefficients ranging from 0.82 to 0.96, p<e-16, as seen in Figure S3B). Additionally,

cell lines with lower epithelial enrichment showed improved responses to the drugs docetaxel and elesclomol, which targets mitochondrial metabolism (Fig. 3D).

We observed tight correlation with the epithelial score enrichment and the EMT signature (Bozorgui et al., 2023), while the mesenchymal enrichment score had significantly worse negative correlation (Fig. 3E). To test if this is an artifact of the calculation, we compared the EMT score across 1035 cell lines of the CCLE database. Based on the expression pattern (Figure S3C), some cancer types have elevated epithelial gene expression (such as SCLC, Bowel, Breast), while other cancer types
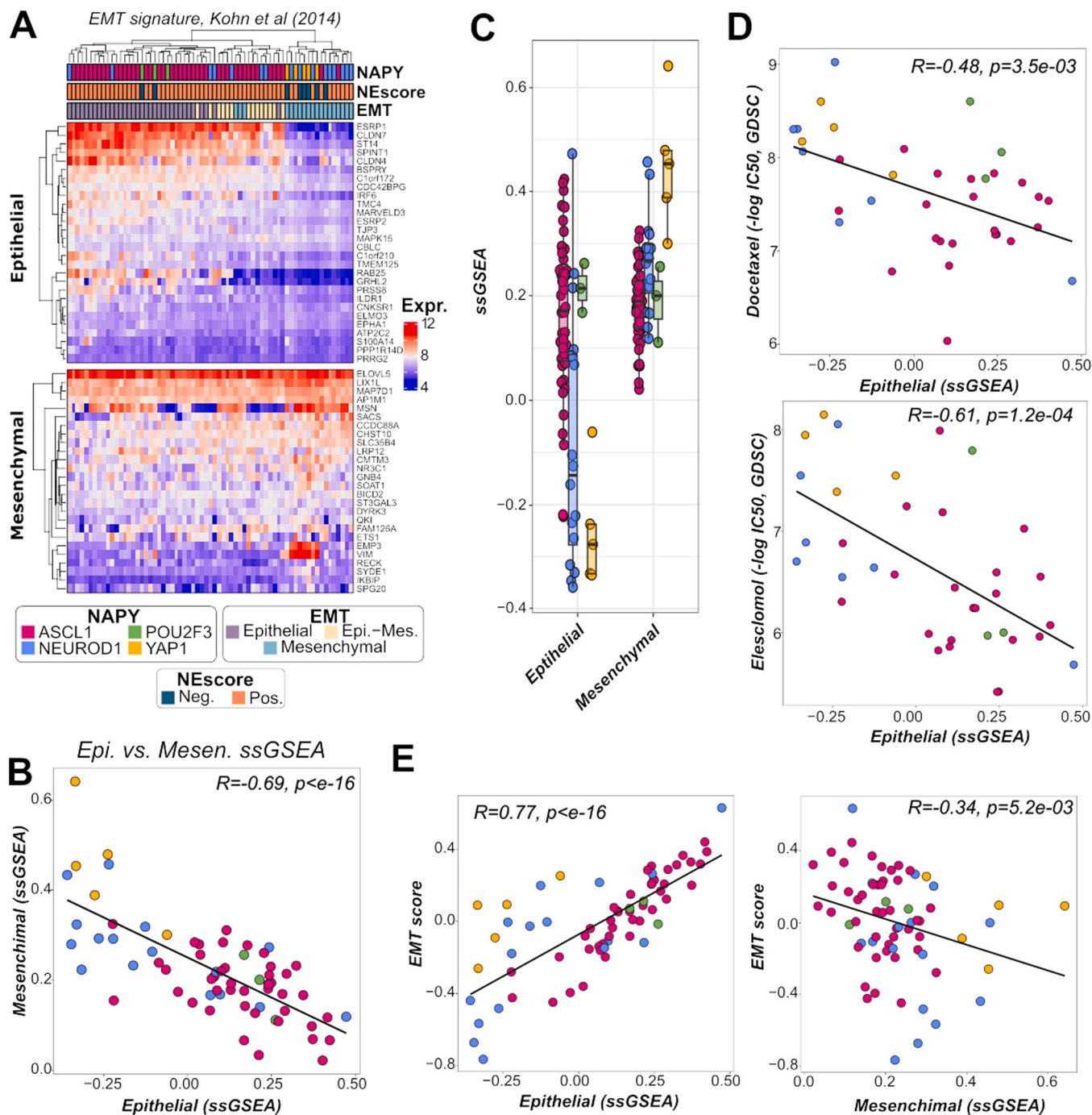
Fig. 3. Analysis of the EMT scores. A) Expression heatmap of EMT genes developed by Kohn and colleagues. B) Correlation of the ssGSEA scores of the epithelial and mesenchymal genes. C) Epithelial and mesenchymal ssGSEA score distribution in the SCLC subtypes. D) Correlation of the epithelial ssGSEA score against docetaxel (top) and elesclomol (bottom). E) Correlation of EMT score and the epithelial (left) or mesenchymal (right) ssGSEA scores.

display elevated mesenchymal expression levels (such as Bone, Brain and Skin). Both the epithelial and mesenchymal enrichment scores correlated very tightly with the EMT score (Figure S3D), where the epithelial score displayed negative correlation.

3.4. Inflamed subtype

Recently a new classification has emerged, that classified samples into different subtypes using a gene expression signature of approximately 1300 genes. The four identified subtypes were the ASCL1-driven, NEUROD1-driven, POU2F3-driven and Inflamed subtype, where the

Inflamed subtype benefitted significantly from treatment with carboplatin/etoposide (EP) plus the PD-L1 antibody atezolizumab.

Cluster analysis of the gene signature using the primary SCLC tumor cohort was able to separate five gene groups: 1) neuroendocrine specific genes (NE-high), 2) ASCL1 enriched genes (ASCL1-high), 3) NEUROD1 specific genes (NEUROD1-high), 4) POU2F3 enriched genes (POU2F3-high) and 5) inflamed high genes (Inflamed) (Table S3). Clustering cell lines based on the grouping from tumor samples produced similar expression patterns, with one marked difference, where the inflamed specific genes can be further separated into two groups: one with variable expression among cell lines (termed Inflamed V1), and one with low

variability (termed Inflamed V2) (Z-scored heatmap in Fig. 4A, log2 transformed expression in Figure S4A). Contrary to the cell lines, the Inflamed V2 genes had variable expression among the tumor samples, potentially expressed by non-cancer cells such as immune cells. Despite the strong differential expression patterns seen in the two inflamed groups, no significant differential pathway enrichment was observed using gene ontology analysis (Table S4).

The SCLC subtypes further highlighted the differential enrichment scores of the clustered gene sets (Fig. 4B). The ASCL1+and NEUROD1+

subtypes both had strong enrichment of the NE-high genes. The ASCL1-high genes were more enriched in ASCL1+ cell lines, while the NEUROD1-high genes were more enriched in the NEUROD1+ cell lines. As expected, the POU2F3-high genes were elevated in the POU2F3+ cell lines. The Inflamed V1 signature was enriched in the non-NE POU2F3+ and YAP1+ cell lines. The inflamed V2 signature presented with very low enrichment in all subtypes, resulting from the lower expression level of these genes. Similar results were seen when using data from both the CCLE and GDSC databases (Figure S4B). Pairwise comparison across the
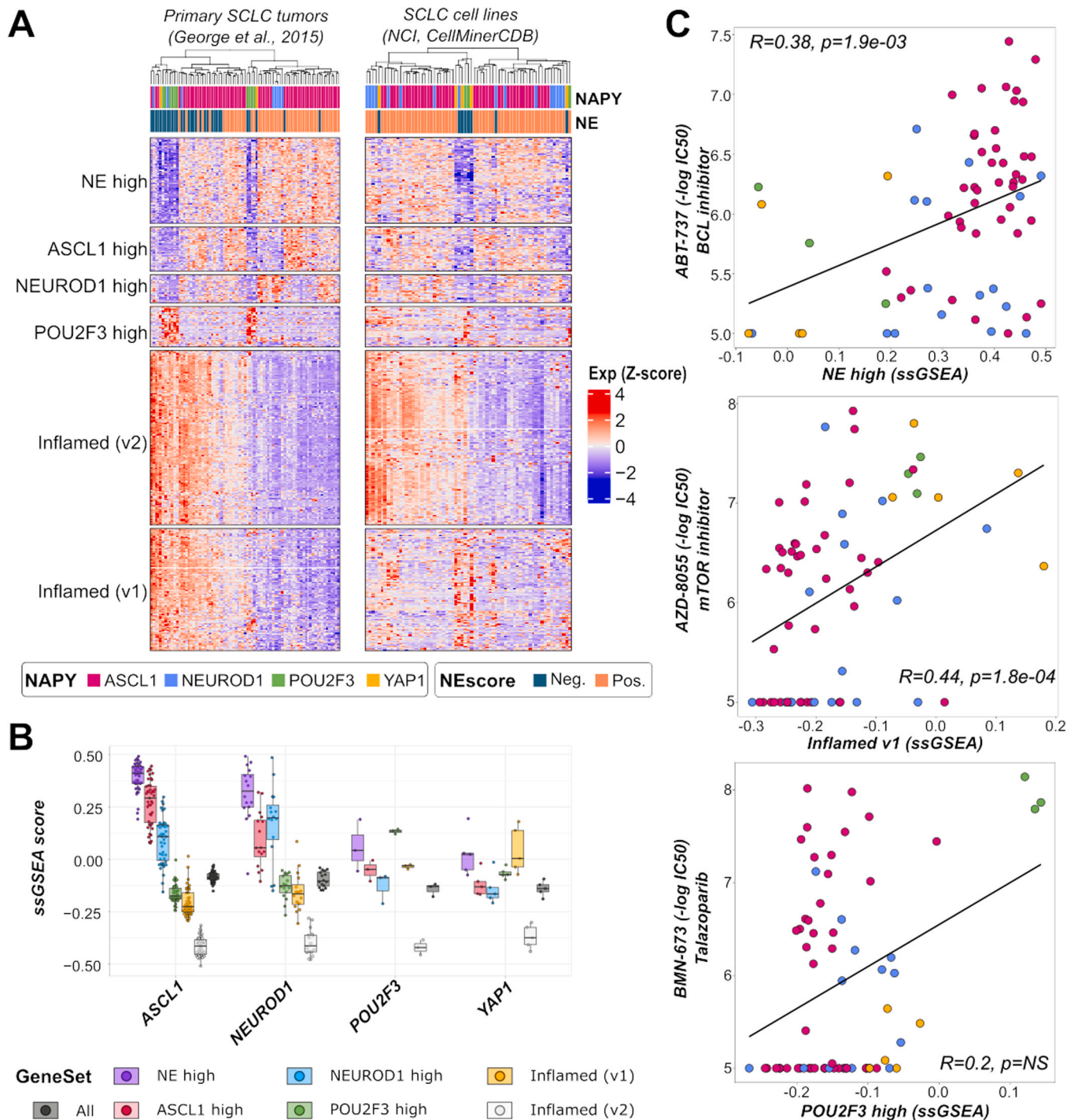
Fig. 4. Inflamed subtype signature. A) Expression heatmap (Z-scored) of primary SCLC tumors (left) and cell lines (right). The order of the gene groups (slices) is the same in the two heatmaps, but the gene order within the slices is not the same. B) Comparison of enrichment score distribution of gene clusters in the four SCLC subtypes. C) Correlation of drug response with ssGSEA scores. (Top) ABT-737 response vs 'NE high' ssGSEA scores; (Center) AZD-8055 response vs 'Inflamed v1' ssGSEA scores; BMN-673 response vs 'POU2F3 high' ssGSEA scores.

different data sources generated correlation coefficients above 0.89 in all cases (range [0.89–0.97]) excluding the Inflamed V2 gene set comparisons (Figure S5).

The NE-high genes correlated positively with response to the ABT-737 BCL2 inhibitor, while the inflamed-V1 enrichment correlated positively with the AZD-8055 mTOR inhibitor. Although there was no significant correlation with the PARP inhibitor and POU2F3-high enrichment score, the POU2F3+ cell lines had strong enrichment scores and response sensitivity.

Overall, our results demonstrate that clustering gene signatures can help condense diverse gene expression patterns, which can then be used to identify drug candidates. The enrichment scores are highly reproducible between databases which employed different expression quantification methods (microarray and RNA-seq), as long as gene groups are correctly identified. Using this method on complex gene expression signatures can help identify gene groups and pathways that can further differentiate SCLC subtypes, potentially helping in finding responder patients and refining treatment strategies to improve patient care.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the Lőrinc S. Pongor used ChatGPT in order to rephrase text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Funding

CRediT authorship contribution statement

Topolcsányi Petronella: Visualization, Writing – original draft, Writing – review & editing. **Benő Alexandra:** Data curation, Visualization, Writing – original draft, Writing – review & editing. **Mihalekné Fűr Gabriella:** Visualization, Writing – original draft, Writing – review & editing. **Magó Éva:** Visualization, Writing – original draft, Writing – review & editing. **Nemes Kolos:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Pongor Lőrinc Sándor:** Conceptualization, Funding acquisition, Investigation, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jbiotec.2024.01.010.

References

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., Fröhling, S., Chan, E.M., Sos, M.L., Michel, K., Mermel, C., Silver, S.J., Weir, B.A., Reiling, J.H., Sheng, Q., Hahn, W.C., 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462* (7269), 108–112. https://doi.org/10.1038/nature08460.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M.F., Monahan, J.E., Morais, P., Meltzer, J., Korejwa, A., Jané-Valbuena, J., Garraway, L.A., 2012. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483* (7391), 603–607. https://doi.org/10.1038/nature11003.

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., Johnson, J.E., 2016. ASCL1 and NEUROD1 reveal heterogeneity in pulmonary neuroendocrine tumors and regulate distinct genetic programs. Cell Rep. *16* (5), 1259–1272. https://doi.org/10.1016/j.celrep.2016.06.081.

Bozorgui, B., Kong, E.K., Luna, A., Korkut, A., 2023. Mapping the functional interactions at the tumor-immune checkpoint interface. Commun. Biol. *6* (1), 462 https://doi.org/10.1038/s42003-023-04777-3.

Byers, L.A., Diao, L., Wang, J., Saintigny, P., Girard, L., Peyton, M., Shen, L., Fan, Y., Giri, U., Tumula, P.K., Nilsson, M.B., Gudikote, J., Tran, H., Cardnell, R.J.G., Bearss, D.J., Warner, S.L., Foulks, J.M., Kanner, S.B., Gandhi, V., Heymach, J.V., 2013. An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. Clin. Cancer Res.: Off. J. Am. Assoc. Cancer Res. *19* (1), 279–290. https://doi.org/10.1158/1078-0432.CCR-12-1558.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C., Schultz, N., 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2* (5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095.

Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., Liu, Q., Iorio, F., Surdez, D., Chen, L., Milano, R.J., Bignell, G.R., Tam, A.T., Davies, H., Stevenson, J.A., Benes, C. H., 2012. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483* (7391), 570–575. https://doi.org/10.1038/nature11005.

Gay, C.M., Stewart, C.A., Park, E.M., Diao, L., Groves, S.M., Heeke, S., Nabet, B.Y., Fujimoto, J., Solis, L.M., Lu, W., Xi, Y., Cardnell, R.J., Wang, Q., Fabbri, G., Cargill, K.R., Vokes, N.I., Ramkumar, K., Zhang, B., Della Corte, C.M., Byers, L.A., 2021. Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. Cancer Cell *39* (3), 346–360. https://doi.org/10.1016/j.ccell.2020.12.014.

Gazdar, A.F., Bunn, P.A., Minna, J.D., 2017. Small-cell lung cancer: what we know, what we need to know and the path forward. Nat. Rev. Cancer *17* (12), 725–737. https://doi.org/10.1038/nrc.2017.87.

George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., Müller, C., Dahmen, I., Jahchan, N.S., Park, K.-S., Yang, D., Karnezis, A.N., Vaka, D., Torres, A., Wang, M.S., Thomas, R.K., 2015. Comprehensive genomic profiles of small cell lung cancer. Nature *524* (7563), 47–53. https://doi.org/10.1038/nature14664.

Govindan, R., Page, N., Morgensztern, D., Read, W., Tierney, R., Vlahiotis, A., Spitznagel, E.L., Piccirillo, J., 2006. Changing epidemiology of small-cell lung cancer in the United States over the last 30 years: analysis of the surveillance, epidemiologic, and end results database. J. Clin. Oncol.: Off. J. Am. Soc. Clin. Oncol. *24* (28), 4539–4544. https://doi.org/10.1200/JCO.2005.04.4859.

Gu, Z., Eils, R., Schlesner, M., 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32* (18), 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

Hänzelmann, S., Castelo, R., Guinney, J., 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinforma. *14*, 7 https://doi.org/10.1186/1471-2105-14-7.

Huang, Y.-H., Klingbeil, O., He, X.-Y., Wu, X.S., Arun, G., Lu, B., Somerville, T.D.D., Milazzo, J.P., Wilkinson, J.E., Demerdash, O.E., Spector, D.L., Egeblad, M., Shi, J., Vakoc, C.R., 2018. POU2F3 is a master regulator of a tuft cell-like variant of small cell lung cancer. Genes Dev. *32* (13–14), 915–928. https://doi.org/10.1101/gad.314815.118.

Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., Cokelaer, T., Greninger, P., van Dyk, E., Chang, H., de Silva, H., Heyn, H., Deng, X., Egan, R.K., Liu, Q., Garnett, M.J., 2016. A landscape of pharmacogenomic interactions in cancer. Cell *166* (3), 740–754. https://doi.org/10.1016/j.cell.2016.06.017.

Jin, X., Demere, Z., Nair, K., Ali, A., Ferraro, G.B., Natoli, T., Deik, A., Petronio, L., Tang, A.A., Zhu, C., Wang, L., Rosenberg, D., Mangena, V., Roth, J., Chung, K., Jain, R.K., Clish, C.B., Vander Heiden, M.G., Golub, T.R., 2020. A metastasis map of human cancer cell lines. Nature *588* (7837), 331–336. https://doi.org/10.1038/s41586-020-2969-2.

Kohn, K.W., Zeeberg, B.M., Reinhold, W.C., Pommier, Y., 2014. Gene expression correlations in human cancer cell lines define molecular interaction networks for epithelial phenotype. PLOS One 9 (6), e99269. https://doi.org/10.1371/journal. pone.0099269.

Krohn, A., Ahrens, T., Yalcin, A., Plönes, T., Wehrle, J., Taromi, S., Wollner, S., Follo, M., Brabletz, T., Mani, S.A., Claus, R., Hackanson, B., Burger, M., 2014. Tumor cell heterogeneity in small cell lung cancer (SCLC): phenotypical and functional differences associated with epithelial-mesenchymal transition (EMT) and DNA methylation changes. PLOS One 9 (6), e100249. https://doi.org/10.1371/journal. pone.0100249.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., Tamayo, P., 2015. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst. 1 (6), 417–425. https://doi.org/10.1016/j.cels.2015.12.004.

Lissa, D., Takahashi, N., Desai, P., Manukyan, I., Schultz, C.W., Rajapakse, V., Velez, M. J., Mulford, D., Roper, N., Nichols, S., Vilimas, R., Sciuto, L., Chen, Y., Guha, U., Rajan, A., Atkinson, D., El Meskini, R., Weaver Ohler, Z., Thomas, A., 2022. Heterogeneity of neuroendocrine transcriptional states in metastatic small cell lung cancers and patient-derived models. Nat. Commun. 13 (1), 2023 https://doi.org/ 10.1038/s41467-022-29517-9.

McColl, K., Wildey, G., Sakre, N., Lipka, M.B., Behtaj, M., Kresak, A., Chen, Y., Yang, M., Velcheti, V., Fu, P., Dowlati, A., 2017. Reciprocal expression of INSM1 and YAP1 defines subgroups in small cell lung cancer. Oncotarget 8 (43), 73745–73756. https://doi.org/10.18632/oncotarget.20572.

Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M.J., Patterson, N., Mesirov, J.P., Golub, T.R., Tamayo, P., Spiegelman, B., Lander, E.S., Hirschhorn, J.N., Groop, L.C., 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. 34 (3), 267–273. https://doi.org/10.1038/ng1180.

Polley, E., Kunkel, M., Evans, D., Silvers, T., Delosh, R., Laudeman, J., Ogle, C., Reinhart, R., Selby, M., Connelly, J., Harris, E., Fer, N., Sonkin, D., Kaur, G., Monks, A., Malik, S., Morris, J., Teicher, B.A., 2016. Small cell lung cancer screen of oncology drugs, investigational agents, and gene and microRNA expression. J. Natl. Cancer Inst. 108 (10) https://doi.org/10.1093/jnci/djw122.

Pongor, L.S., Tlemsani, C., Elloumi, F., Arakawa, Y., Jo, U., Gross, J.M., Mosavarpour, S., Varma, S., Kollipara, R.K., Roper, N., Teicher, B.A., Aladjem, M.I., Reinhold, W., Thomas, A., Minna, J.D., Johnson, J.E., Pommier, Y., 2022. Integrative epigenomic analyses of small cell lung cancer cells demonstrates the clinical translational relevance of gene body methylation. IScience 25 (11), 105338. https://doi.org/ 10.1016/j.isci.2022.105338.

Pongor, L.S., Schultz, C.W., Rinaldi, L., Wangsa, D., Redon, C.E., Takahashi, N., Fialkoff, G., Desai, P., Zhang, Y., Burkett, S., Hermoni, N., Vilk, N., Gutin, J., Gergely, R., Zhao, Y., Nichols, S., Vilimas, R., Sciuto, L., Graham, C., Thomas, A., 2023. Extrachromosomal DNA amplification contributes to small cell lung cancer heterogeneity and is associated with worse outcomes. Cancer Discov. 13 (4), 928–949. https://doi.org/10.1158/2159-8290.CD-22-0796.

Rajapakse, V.N., Luna, A., Yamade, M., Loman, L., Varma, S., Sunshine, M., Iorio, F., Sousa, F.G., Elloumi, F., Aladjem, M.I., Thomas, A., Sander, C., Kohn, K.W., Benes, C. H., Garnett, M., Reinhold, W.C., Pommier, Y., 2018. CellMinerCDB for integrative cross-database genomics and pharmacogenomics analyses of cancer cell lines. IScience 10, 247–264. https://doi.org/10.1016/j.isci.2018.11.029.

Roper, N., Velez, M.J., Chiappori, A., Kim, Y.S., Wei, J.S., Sindiri, S., Takahashi, N., Mulford, D., Kumar, S., Ylaya, K., Trindade, C., Manukyan, I., Brown, A.-L., Trepel, J. B., Lee, J.-M., Hewitt, S., Khan, J., Thomas, A., 2021. Notch signaling and efficacy of PD-1/PD-L1 blockade in relapsed small cell lung cancer. Nat. Commun. 12 (1), 3880 https://doi.org/10.1038/s41467-021-24164-y.

Rudin, C.M., Poirier, J.T., Byers, L.A., Dive, C., Dowlati, A., George, J., Heymach, J.V., Johnson, J.E., Lehman, J.M., MacPherson, D., Massion, P.P., Minna, J.D., Oliver, T. G., Quaranta, V., Sage, J., Thomas, R.K., Vakoc, C.R., Gazdar, A.F., 2019. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nat. Rev. Cancer 19 (5), 289–297. https://doi.org/10.1038/s41568-019-0133-9.

Schultz, C.W., Zhang, Y., Elmeskini, R., Zimmermann, A., Fu, H., Murai, Y., Wangsa, D., Kumar, S., Takahashi, N., Atkinson, D., Saha, L.K., Lee, C.-F., Elenbaas, B., Desai, P., Sebastian, R., Sharma, A.K., Abel, M., Schroeder, B., Krishnamurthy, M., Thomas, A., 2023. ATR inhibition augments the efficacy of lurbinectedin in small-cell lung cancer. EMBO Mol. Med. 15 (8), e17313 https://doi.org/10.15252/ emmm.202217313.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102 (43), 15545–15550. https://doi. org/10.1073/pnas.0506580102.

Takahashi, N., Rajapakse, V.N., Pongor, L., Kumar, S., Tlemsani, C., Erwin-Cohen, R., Young, H.A., Hewitt, S., Wei, J.S., Khan, J., Villarino, A.V., Trepel, J.B., Thomas, A., 2020. Dynamics of genomic and immune responses during primary immunotherapy resistance in mismatch repair-deficient tumors. Cold Spring Harb. Mol. Case Stud. 6 (5) https://doi.org/10.1101/mcs.a005678.

Thomas, A., Mian, I., Tlemsani, C., Pongor, L., Takahashi, N., Maignan, K., Snider, J., Li, G., Frampton, G., Ali, S., Kim, S., Nichols, S., Rajapakse, V., Guha, U., Sharon, E., Fujimoto, J., Moran, C.A., Wistuba, I.I., Wei, J.S., Carson, K.R., 2020. Clinical and genomic characteristics of small cell lung cancer in never smokers: results from a retrospective multicenter cohort study. Chest 158 (4), 1723–1733. https://doi.org/ 10.1016/j.chest.2020.04.068.

Thomas, A., Takahashi, N., Rajapakse, V.N., Zhang, X., Sun, Y., Ceribelli, M., Wilson, K. M., Zhang, Y., Beck, E., Sciuto, L., Nichols, S., Elenbaas, B., Puc, J., Dahmen, H., Zimmermann, A., Varonin, J., Schultz, C.W., Kim, S., Shimellis, H., Thomas, C.J., 2021. Therapeutic targeting of ATR yields durable regressions in small cell lung cancers with high replication stress. Cancer Cell 39 (4), 566–579.e7. https://doi.org/ 10.1016/j.ccell.2021.02.014.

Tlemsani, C., Pongor, L., Elloumi, F., Girard, L., Huffman, K.E., Roper, N., Varma, S., Luna, A., Rajapakse, V.N., Sebastian, R., Kohn, K.W., Krushkal, J., Aladjem, M.I., Teicher, B.A., Meltzer, P.S., Reinhold, W.C., Minna, J.D., Pommier, Y., 2020. SCLC-CellMiner: a resource for small cell lung cancer cell line genomics and pharmacology based on genomic signatures. Cell Rep. 33 (3), 108296 https://doi. org/10.1016/j.celrep.2020.108296.

Tlemsani, C., Takahashi, N., Pongor, L., Rajapakse, V.N., Tyagi, M., Wen, X., Fasaye, G.-A., Schmidt, K.T., Desai, P., Kim, C., Rajan, A., Swift, S., Sciuto, L., Vilimas, R., Webb, S., Nichols, S., Figg, W.D., Pommier, Y., Calzone, K., Thomas, A., 2021. Whole-exome sequencing reveals germline-mutated small cell lung cancer subtype with favorable response to DNA repair-targeted therapies. Sci. Transl. Med. 13 (578) https://doi.org/10.1126/scitranslmed.abc7488.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., Meyers, R.M., Ali, L., Goodale, A., Lee, Y., Jiang, G., Hsiao, J., Gerath, W.F.J., Howell, S., Merkel, E., Hahn, W.C., 2017. Defining a cancer dependency map. Cell 170 (3), 564–576.e16. https://doi.org/10.1016/j.cell.2017.06.010.

Wooten, D.J., Groves, S.M., Tyson, D.R., Liu, Q., Lim, J.S., Albert, R., Lopez, C.F., Sage, J., Quaranta, V., 2019. Systems-level network modeling of small cell lung cancer subtypes identifies master regulators and destabilizers. PLOS Comput. Biol. 15 (10), e1007343 https://doi.org/10.1371/journal.pcbi.1007343.

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: A J. Integr. Biol. 16 (5), 284–287. https://doi.org/10.1089/omi.2011.0118.

Yu, G., Wang, L.-G., Yan, G.-R., He, Q.-Y., 2015. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinforma. (Oxf., Engl.) 31 (4), 608–609. https://doi.org/10.1093/bioinformatics/btu684.

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., Gazdar, A.F., 2018. Small cell lung cancer tumors and preclinical models display heterogeneity of neuroendocrine phenotypes. Transl. Lung Cancer Res. 7 (1), 32–49. https://doi.org/10.21037/tlcr.2018.02.02.