Title

Analysis of transcriptome datasets reveals involvement of epithelial-mesenchymal transition genes in small cell lung cancer phenotypic plasticity

Authors

Sarah M. Groves[1], Nicholas Panchy[2], Darren R. Tyson[1,3], Leonard A. Harris[4,5,6], Vito Quaranta[1,3]*, Tian Hong[2, 7]*

1. Department of Biochemistry, Vanderbilt University, Nashville, TN 37235, USA.
2. Department of Biochemistry and Cellular & Molecular Biology, The University of Tennessee, Knoxville, Knoxville TN 37996, USA.
3. Department of Pharmacology, Vanderbilt University, Nashville, TN 37235, USA.
4. Department of Biomedical Engineering, University of Arkansas, Fayetteville, AR 72701, USA.
5. Interdisciplinary Graduate Program in Cell and Molecular Biology, University of Arkansas, Fayetteville, AR 72701, USA.
6. Cancer Biology Program, Winthrop P. Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA.
7. National Institute for Mathematical and Biological Synthesis, Knoxville, TN 37996, USA.

* To whom correspondence should be addressed

vito.quaranta@vanderbilt.edu or hongtian@utk.edu

Author ORCID
Sarah M. Groves: 0000-0002-3122-2605
Darren R. Tyson: 0000-0002-3272-4308
Leonard A. Harris: 0000-0003-2112-6940
Vito Quaranta: 0000-0001-7491-8672
Tian Hong: 0000-0002-8212-7050

Abstract

Small cell lung cancer (SCLC) is an aggressive cancer recalcitrant to treatment. SCLC arises in lung epithelium, and the phenotypic heterogeneity of SCLC cells plays critical roles in disease progression and drug resistance. Particularly, the transition from neuroendocrine (NE) cells to non-neuroendocrine (non-NE) cells, as well as the cooperations between these cell types, are observed in SCLC progression. At least five subtypes of SCLC tumors have recently been defined by signature gene expression and these subtypes have been extended to individual SCLC tumor cells. However, the molecular and cellular programs that distinguish the cell types or allow transitions among them are unclear. Here, we systematically analyzed the relationship between SCLC and epithelial to mesenchymal transition (EMT)—a cellular process well known to contribute to invasiveness and drug resistance of cancer cells—using multiple transcriptome datasets from a mouse tumor model, human cancer cell lines, and human tumor samples. Our analysis of these datasets consistently showed that an extreme epithelial state corresponds to the SCLC-A2 subtype. Furthermore, two NE subtypes have significant expression of a subset of mesenchymal genes, with surprisingly limited overlap with the mesenchymal genes highly expressed in non-NE cells. Our work reveals a previously unknown connection between an EMT program and a specific SCLC subtype and an underappreciated divergence of EMT trajectories contributing to the SCLC phenotypic heterogeneity.

Introduction

Small cell lung cancer (SCLC) is one of the most aggressive human cancers with characteristic early metastasis and acquisition of therapeutic resistance (Gazdar et al., 2017). SCLC tumors are neuroendocrine (NE) tumors that arise in the lung epithelium, and pulmonary NE cells (PNEC) are considered their normal counterparts (Borges et al., 1997; Song et al., 2012). However, multiple reports have demonstrated intratumoral heterogeneity in most SCLC tumors (Calbo et al., 2011; Lim et al., 2017; Shue et al., 2018). In fact, there is accumulating evidence that SCLC NE cells can transition to non-NE states (Ireland et al., 2020; Lim *et al.*, 2017; Stewart et al., 2020). Furthermore, prior studies suggest cooperative interactions occur between NE and non-NE cells during SCLC tumor progression (Calbo *et al.*, 2011; Ko et al., 2021; Kwon et al., 2015). An open question is whether transitions from NE to non-NE subtypes, and possibly vice versa, are driven by intrinsic factors (e.g., transcription factor network dynamics (Wooten et al., 2019)), extrinsic factors (i.e., microenvironmental influences, such as hypoxia or therapeutic agents), or a combination of both.

A standard to classify SCLC cells based on expression levels of four key transcription factors (TFs), *ASCL1*, *NEUROD1*, *YAP1* or *POU2F3*, was recently adopted (Rudin et al., 2019; Yazawa, 2015). Based on the most abundantly expressed TF, or an underlying signature, three NE (A, A2 and N) and two non-NE (Y and P) subtypes have been proposed (Groves et al., 2022; Wooten et al., 2019). An additional subtype, I, that is enriched in inflammatory genes has also been described (Gay et al., 2021). However, the gene regulatory mechanisms that generate and maintain the molecular and phenotypic identities of SCLC subtypes remain unclear. Of particular interest are mechanisms underlying NE to non-NE phenotypic transitions since they may directly contribute to tumor progression. Because PNECs and SCLC cells are of epithelial origin (endodermally derived) (Noguchi et al., 2020) and SCLC tumors are highly metastatic, we hypothesized that similarities may exist between NE/non-NE transitions and the well-studied epithelial-mesenchymal transition (EMT). EMT is a cellular process in which epithelial cells lose tight cell-junction and gain the ability to migrate, characteristic features of metastatic spread (Kalluri and Weinberg, 2009). EMT also contributes to tumor drug resistance (Fischer et al., 2015; Krebs et al., 2017; Mani et al., 2008), and an EMT transition has been associated with drug resistance in human SCLC tumors (Stewart *et al.*, 2020; Sutherland et al., 2022). However, theoretical and experimental studies revealed that EMT is a multi-stage process and the transcriptional programs (e.g. subsets of mesenchymal signature genes) activated in different contexts can be diverse (Cook and Vanderhyden, 2020; Hong et al., 2015; Lu et al., 2013; Watanabe et al., 2019; Ye et al., 2015; Zhang et al., 2014). Thus, while multiple reports have suggested a role for EMT in SCLC phenotypic plasticity, drug resistance, and metastasis, to the best of our knowledge, a systematic analysis of the relationships between NE and non-NE SCLC subtypes and EMT has not previously been reported.

In this paper, we adopted accepted scoring metrics for epithelial (E) and mesenchymal (M) programs to position NE and non-NE SCLC subtypes in the EMT spectrum. We found that the A2 subtype is strongly epithelial-like (scores high along the E axis) and weakly mesenchymal-like (scores low along the M axis), supporting it as the most E-like subtype of SCLC. In contrast, the

non-NE subtypes (P and Y) score higher on the M axis, while being lowest along the E axis. Surprisingly, the NE subtypes A and N consistently score significantly lower than A2 in the E axis. Moreover, transcripts of some mesenchymal genes, such as ZEB1, are more abundant in N than non-NE subtypes, suggesting the existence of multiple mesenchymal states in SCLC. These findings support involvement of EMT programs in SCLC phenotypic heterogeneity and provide potential mechanistic insights into NE/non-NE transitions.

Results

The SCLC A2 subtype is highly enriched in epithelial gene expression

We first used a time course of single-cell RNA-sequencing (scRNA-seq) dataset from an SCLC genetically engineered mouse model (GEMM) with overactive Myc oncogene (*Rb1*$^{fl/fl}$;*Trp53*$^{fl/fl}$; Lox-Stop-Lox [LSL]-*Myc*^{T58A}, RPM) (Ireland *et al.*, 2020). The 15138 single cells were annotated with the SCLC subtypes using archetype analysis, as described in a recent study (Groves *et al.*, 2022). An SCLC signature comprised of 105 genes was used for subtyping individual cells into four classes (A/N, A2, P/Y and Y) representing extreme expression patterns of NE and non-NE subtypes (see Methods) (Groves *et al.*, 2022). In parallel, using normalized gene expression, we computed the epithelial (E) score and mesenchymal (M) score for this dataset. The scores were computed with a list of 232 epithelial (E) associated genes, a list of 193 mesenchymal (M) associated genes and the single-sample gene set enrichment analysis (ssGSEA) (see Methods) (Deshmukh et al., 2021; Krug et al., 2019; Tan et al., 2014). In the EMT spectrum depicted by the E and M scores, we observed a striking difference between the position of A2 NE subtype and other NE or non-NE subtypes (Figure 1A). The A2 cells had a significantly higher E score and a significantly lower M score compared to other subtypes ($p<10^{-5}$ for both comparisons). Consistent with the E/M scores, we found that A2 cells had significantly higher levels of *Cdh1* (coding E-cadherin) ($p<10^{-5}$), a widely used E marker gene crucial for cell adhesion, compared to other cells (Figure 1A inset). These results indicate that gene expression in the A2 subtype is most restricted to an epithelial lineage, whereas, in all other SCLC subtypes (NE and non-NE), M gene expression appears to be permitted to various extents.

We next asked whether the association between the A2 subtype and the extreme epithelial lineage among RPM mouse SCLC tumor cells extends to human SCLC cells. We first analyzed the bulk RNA-seq data from 120 human SCLC cell lines assigned with subtype identities (see Methods). Several A2 cell lines (e.g., DMS53) were located in the extreme E position in the EMT spectrum (Figure 1B), in agreement with A2 cells from RPM tumors (Figure 1A). As a group, A2 cell lines had the highest mean E score and the lowest M score among SCLC subtypes, although these scores were closer to A and Y cells than in the RPM tumor cell data (Figure 1B). This partial overlap might be due to the cell heterogeneity within cell lines that is not resolved in bulk RNA-seq data. We therefore visualized 13945 single-cell transcriptomes from eight human SCLC cell lines (Groves *et al.*, 2022) with the same E/M projection method. In this dataset, A2 cells are more distinct from other subtypes than in the bulk RNA-seq data, and the separation is more similar to

the RPM tumor cell data (Figure 1C). Interestingly, in both bulk and single-cell RNA-seq datasets for SCLC cell lines, NE subtypes A has an intermediate EMT expression profile (Figure 1B and C).

The strong association between SCLC subtypes and EMT progression raises the question of whether this connection is trivially due to a sharing of genes between the 105 SCLC subtype signature gene set and the 425 EMT signature gene set. However, there are only 20 genes appearing in both gene sets, indicating limited overlap (Jaccard index 0.039). Furthermore, excluding the 20 overlapping SCLC signature genes from the EMT gene set had minimal effect on the distributions of SCLC cells in the EMT spectrum in each of the three datasets (Figure S1). Taken together, our analyses indicate a strong and nontrivial association between SCLC subtypes and EMT progression. In particular, a specific SCLC NE subtype, A2, exhibits phenotypes associated with highly restricted epithelial lineages in mammals.

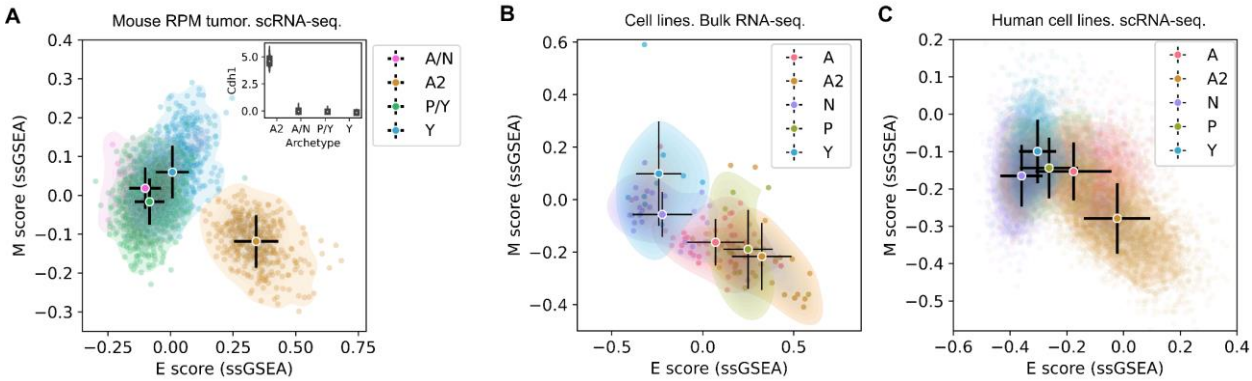


Figure 1. Distributions of SCLC cells across the epithelial and mesenchymal spectrum. **A**. E and M scores of 15138 mouse *Myc*-overexpressing SCLC GEMM tumor cells in an *ex vivo* model. Scores were computed with ssGSEA and 425 previously identified EMT genes. Circles and black bars show means and standard deviations of SCLC subtypes. Inset shows normalized expression of Cdh1 in four types of specialist cells. **B**. E and M scores of 120 SCLC cell lines computed with the same method as in A. **C**. E and M scores of 13945 cells from 8 SCLC cell lines computed with the same method as in A.

Mesenchymal scoring is divergent across non-A2 subtypes

The ssGSEA-based E/M scoring did not show any dramatic difference among A, N, P and Y (i.e. non-A2) subtypes (Figure 1). However, expression of individual M genes, such as *Vim*, differed significantly among non-A2 subtypes (Figure 2A and B). This suggests the possibility that the M genes have divergent expression patterns across the SCLC subtypes which may mask the ssGSEA-based scoring. Recent studies in tumors and cell lines also suggested diversity of EMT programs (Cook and Vanderhyden, 2020; Pastushenko et al., 2018; Watanabe *et al.*, 2019). In particular, mesenchymal genes (upregulated in EMT) were found to be regulated differently in

terms of their connectivity to EMT signals (Watanabe *et al.*, 2019). To dissect potentially divergent M programs in SCLC cells, we used an alternative scoring method based on nonnegative principal component analysis (nnPCA). For each gene list, nnPCA generates leading PCs that describe the variance in gene expression with biologically meaningful directionality. The projections of data points onto these PCs were used as gene set (E/M) scores. We selected the primary PCs based on the highest explained variance of the mean scores of the specialist groups (see Methods). We asked whether there is a heterogeneous pattern of EMT genes among the non-A2 SCLC subtypes. With the single-cell data from both the mouse RPM model and the SCLC cell lines, we noticed that the first PC for M scores explained less variance compared to the first PC for E scores (Figure 2C and D top). Furthermore, the first two M PCs produced comparable variances explained. These results suggest diversity of M gene expression, warranting further investigations. Interestingly, M scores obtained from the first PC ranked the SCLC subtypes differently from the second: A/N subtypes had *higher* scores than Y subtype cells based on the first M PC (nnPC1), but they had lower scores than Y subtype with the second M PC (nnPC2) (Figure 2C and D scatter plots). In fact, two representative M genes—*Vim* (the gene encoding vimentin, an intermediate filament protein component of mesenchymal cell cytoskeletons) and *Zeb1* (a widely studied EMT transcription factor)—had an anticorrelated pattern between A/N and Y subtypes in the RPM dataset (compare Figure 2E to Figure 2A) (Mendez et al., 2010; Wellner et al., 2009). Although this anticorrelation was less prominent in the cell line data (compare Figure 2F to Figure 2B), distinct *Vim* and *Zeb1* expressions of SCLC subtypes were observed. Overall, like the case of *Vim* and *Zeb1*, the mean differences in expression of M genes between A/N subtypes and Y subtype were diverse, and these differences were positively correlated between the RPM data and the cell line data (Pearson correlation coefficient 0.29, $p<10^{-5}$). Some M genes, such as *Zeb1*, had higher expression in the A/N subtypes than in the Y subtype, while some others, such as *Vim*, had the opposite pattern (Figure 2G and H). We defined these two groups of M genes as M1 and M2 respectively (FDR<0.05 for N-Y differences). We performed gene ontology analysis and found that, while 20 enriched biological processes were shared between the two groups, M1 genes were uniquely enriched in more than 200 processes including ‘negative regulation of cell adhesion involved in substrate-bound cell migration’ (fold enrichment >100, FDR=0.005), whereas M2 were uniquely enriched in 11 processes including ‘extracellular matrix organization’ (fold enrichment =14.08, FDR=0.02) (Tables S1 and S2). This result supports the functional divergence of the M genes that were differentially expressed in NE and non-NE SCLC subtypes.

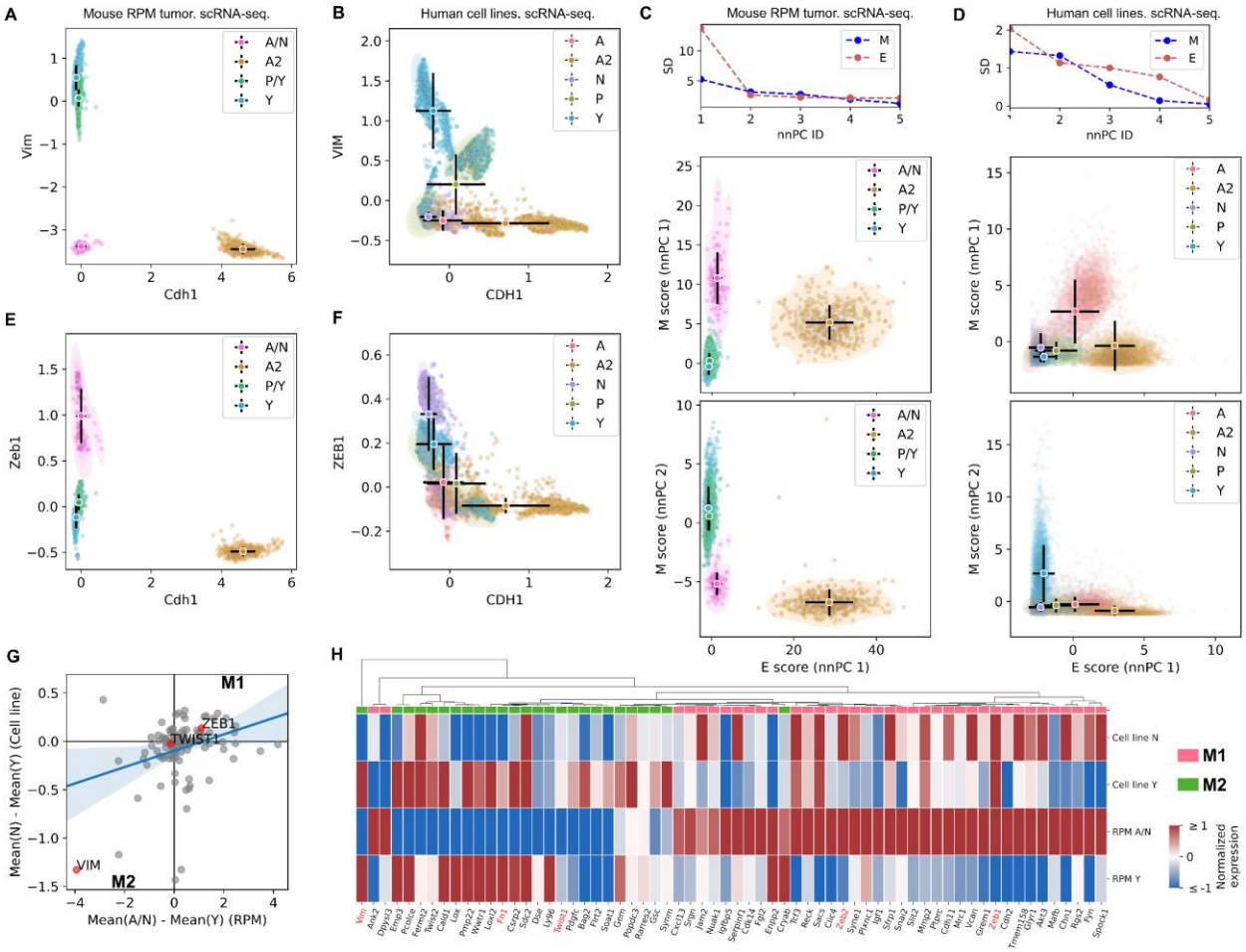


Figure 2. **Divergence of mesenchymal gene expressions in SCLC subtypes**. **A, B**. Normalized expression levels of Vim (M gene) and Cdh1 (E gene) in 15138 mouse *Myc*-driven tumor cells (A) and in 13945 cells from 8 SCLC cell lines (B). Imputed data were plotted in B for visual aid. Circles and black bars show means and standard deviations of SCLC subtypes. **C, D**. Line plots: nnPCA was performed. Plots show standard deviations of SCLC subtype means for mouse and human SCLC cells from the top 5 PCs. Scatter plots: nnPCA-based E and M scores. **E, F**. Normalized expression levels of Zeb1 (M gene) and Cdh1 (E gene) in 15138 mouse *Myc*-driven tumor cells (E) and in 13945 cells from 8 SCLC cell lines (F). Imputed data were plotted in F for visual aid. **G**. Mean differences between N cells (A/N cells for mouse) and Y cells in individual M gene's expression in mouse and human SCLC cells. **H**. A heatmap showing the diversity of M gene expression across N (A/N) and Y subtypes.

The strongly epithelial A2 subtype is detectable in a subset of SCLC patient samples

The distinction of A2 subtype from A subtype has generally not been considered in previous studies by other investigators. For example, Chan et al. profiled 54,523 individual SCLC cells from 19 human samples and classified them into A, N and P subtypes (note the absence of the Y subtype) (Chan et al., 2021). We therefore asked whether our analytical framework can be used to discover previously unknown, extremely epithelial-like cells in this human SCLC tumor dataset.

We first performed E and M scoring for this dataset (Figure 3A) and found the A cells (as subtyped in Chan et al., 2021) fell into two distinct groups, one of which was highly E-like (Figure 3B, red arrow). We hypothesized that the cells labeled as A subtype by Chan et al. include both extreme E-like (A2) and EMT-intermediate-like (A) subtypes. We re-subtyped the SCLC cells from Chan et al. using the method for analyzing other datasets in this study, and we found a significant fraction of the previously labeled A cells had high scores for A2 subtype and low scores for A subtype. In particular, the distinct group of E-like cells corresponds to the group of cells that had high A2 scores and low A scores (Figure 3C and D). Interestingly, the cell cluster with high A2 scores correspond to a unique tumor sample (Figure 3A, pink) with a unique treatment type (Figure S2).

Divergent M gene expression patterns are detectable in human SCLC tumor samples

Systematic analysis for divergent M programs could not be performed due to the absence of any Y subtype cells in the dataset. Nonetheless, heterogeneous expressions of M genes *ZEB1* and *VIM* were observed within the N subtype (Figure 3E). Furthermore, three M-like cell clusters with similar E and M score rankings (Figure 3A, red, dark green and brown) had dramatically different profiles of *VIM* and *ZEB1* expression (Figure 3E) even though all were classified as N (Figure 3F). These results further support the association between the SCLC and EMT program and reveal a previously underappreciated connection of EMT to heterogeneity of SCLC tumor cells.

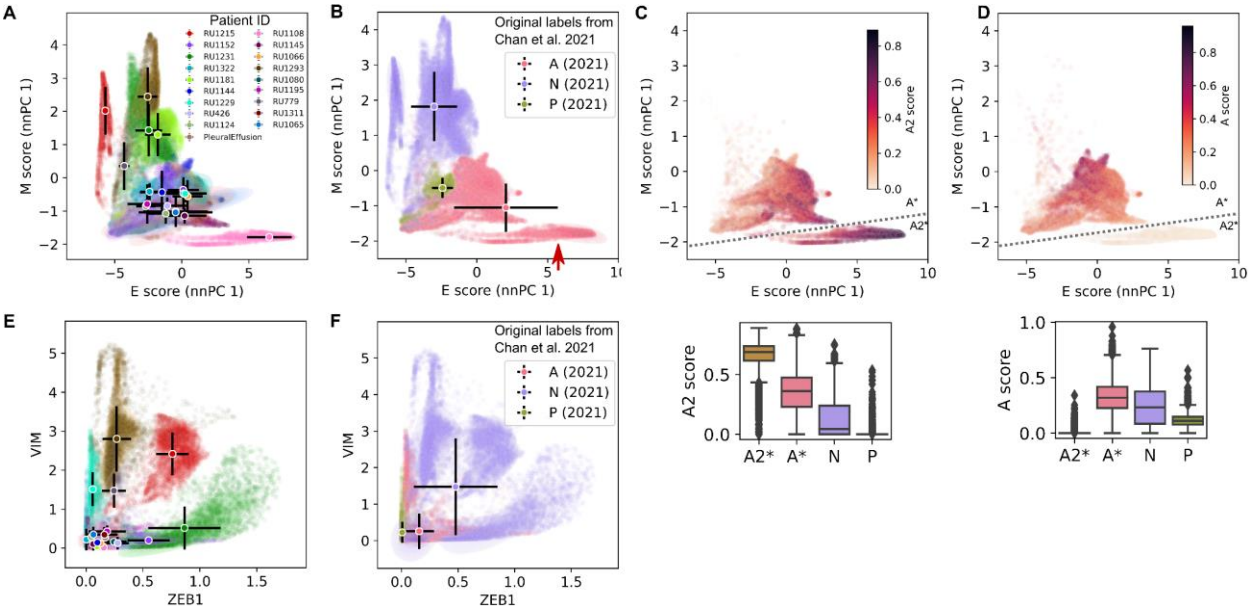


Figure 3. Predictions of epithelial-like A2 SCLC subtypes in human tumor cells. **A**. Scatter plot shows nnPCA-based E and M scores for 54,523 SCLC cells (Chan *et al.*, 2021). Color code represents each of the 19 patients. Circles and black bars show means and standard deviations

of cell scores for each patient. **B.** Same data as in A with SCLC subtype labels obtained from Chan et al. (2021) study. Circles and black bars show means and standard deviations of scores across the previously determined SCLC subtypes. **C**, **D**. Scatter plots show new A2 and A subtype scores in the previously determined SCLC A subtype cells from the data in B. Boxplots show subtype scores in this study across newly defined A* and A2* subtypes based on the threshold indicated in above) as well as N and P subtypes defined by Chan et al. (A and B). **E**. Normalized expression of *ZEB1* and *VIM* in the same dataset as in A. Color code represents each of the 19 patients as in A. **F**. The same data as in E but color labels defined by SCLC subtype as shown in B.

Discussion

Understanding the molecular basis for genetic and non-genetic tumor heterogeneity is crucial for developing next-generation therapeutic strategies (Altschuler and Wu, 2010; Jordan et al., 2016; Sáez-Ayala et al., 2013; Shaffer et al., 2017). Recent advances in defining subtypes of SCLC tumors provided such potentials (Gay *et al.*, 2021), but cellular and molecular mechanisms contributing to the phenotypic heterogeneity is still unclear. Several lines of evidence suggested that tumor cells may hijack EMT, canonically defined as a developmental program (Thiery et al., 2009), to increase their invasiveness and drug resistance via cell state transitions (Fischer *et al.*, 2015; Mani *et al.*, 2008). Interestingly, SCLC tumors have characteristics of both NE cells and epithelial cells (Stovold et al., 2013). While it was shown that some genes promoting EMT were activated in NE/non-NE transitions of SCLC (Groves *et al.*, 2022; Ireland *et al.*, 2020; Ito et al., 2017), connections between SCLC subtypes to the EMT spectrum remained elusive. It was also unclear whether SCLC involves multiple (unique) partial EMT states. Through analysis of transcriptome datasets from multiple sources including a mouse tumor model, cell lines, and human tumors, our work reveals a strong correspondence between an EMT spectrum, containing an epithelial state and several divergent partial EMT states, and recently defined SCLC subtypes. Instead of a previously implicated simple relation of NE-to-non-NE transition to EMT, our study indicated the partial EMT status of the NE subtypes A and N, whereas the NE subtype A2 is fully epithelial-like. In addition, we showed divergence of mesenchymal gene expression in N (an NE) and Y (a non-NE) subtypes. This work builds on our prior studies that had suggested the partial EMT status of N and an enrichment of epithelial genes in A2 (Groves *et al.*, 2022) by revealing the nontrivial relationships between NE/non-NE SCLC subtypes and EMT progressions and help guide future work to improve the understanding SCLC tumor heterogeneity.

Our analysis showed the divergent expressions of mesenchymal factors. This is consistent with recent observations on context-specific EMT (Cook and Vanderhyden, 2020). Interestingly, the EMT transcription factor ZEB1 and EMT effector gene VIM have an anti-correlated expression pattern between N and Y subtypes. Although the contribution of ZEB1 to EMT was demonstrated *in vivo* and *in vitro* (Celià-Terrassa et al., 2018; Cieply et al., 2013; Han et al., 2022; Watanabe *et al.*, 2019), its high expression may not be required for some partial EMT states. Our previous work

showed that expressions of some mesenchymal genes do not require ZEB1 activation, and that the high expression of a group of EMT genes positively controlled by ZEB1, but not TGF-β, is correlated with better prognosis of breast cancer patients (Watanabe *et al.*, 2019). It is possible that ZEB1 is transiently required for maintaining a group of NE cells during SCLC progression, and the transitions to more drug-resistant subtypes may require the down-regulation of ZEB1.

The heterogeneity and coordination of NE and non-NE SCLC subtypes have been observed previously. However, the factors driving transitions between NE and non-NE subtypes are not well characterized. It has long been known that EMT is reversible in both embryonic and postnatal development, and this reversible process can also be triggered in cancer cell lines using either dynamic extracellular signals (e.g. TGF-β) or forced expression of intracellular factors (e.g. TFs or microRNAs) (Cursons et al., 2018; Grande et al., 2015; Hong *et al.*, 2015; Watanabe *et al.*, 2019). Our findings suggest that the machinery for reversible EMT may also be responsible for the transitions between NE and non-NE cells during the progression of SCLC. Future work is warranted to determine the temporal sequence of activation among EMT TFs, signaling molecules, and subtype-defining factors for SCLC, such that the intrinsic and extrinsic factors contributing to the NE/non-NE transition can be dissected.

Methods

RNA-sequencing data of SCLC cells

We obtained and batch-corrected bulk RNA sequencing expression data on SCLC cell lines from the Cancer Cell Line Encyclopedia (50) and cBioPortal (70) (Cerami et al., 2012). Single-cell RNA sequencing data was downloaded from Gene Expression Omnibus (GEO) at GSE193959 (human cell lines) and GSE149180 (RPM mouse tumor time course) (Groves *et al.*, 2022; Ireland *et al.*, 2020). Human tumor data was downloaded from the Human Tumor Atlas Network deposited by Chan et al. (Chan *et al.*, 2021) at Synapse ID syn23630203. Human cell line and RPM mouse tumor datasets were preprocessed as described in Groves et al. (Groves *et al.*, 2022), including filtering and normalization of total counts using the Python package *Scanpy* (Wolf et al., 2018), log-transformation using the *log1p* function from the *Numpy* package, and scaling using *Scanpy*. Log-normalized expression levels were used for comparisons (Mann-Whitney U test) between subtypes.

Subtyping of SCLC cells

Single cell subtype labeling was done as described in Groves et al. (Groves *et al.*, 2022). Briefly, after archetype analysis was applied to the MAGIC-imputed single cell datasets and vertices were identified (Hart et al., 2015; Van Dijk et al., 2018), we labeled the cells closest to each archetype as that subtype if the archetype score was > 0.95 (each cell receives weighted archetype scores that sum to 1). SCLC archetypal signatures were generated from bulk human cell line and tumor transcriptomics data, giving a matrix of 105 genes by 5 archetypes (SCLC-A, SCLC-A2, SCLC-N, SCLC-P, and SCLC-Y). In order to align the single-cell archetypes with our bulk archetype signatures, we consider the scores for each cell described in the STAR Methods section "Bulk gene signature scoring of single cells using archetype signature matrix" from Groves et al. (Groves *et al.*, 2022). For each bulk signature x and for each single-cell archetype a, we ran the following significance test:

(1) Find the mean bulk score x for a specialist, m.
(2) Choose a random sample of size n_a, where n_a is the number of specialists, with replacement from the remaining cells (i.e. cells that are not specialists, including generalists and other specialist cells). Find the mean bulk score for this sample. N.B. Because some time points have very few cells, we sample evenly from each time point to ensure adequate representation across the time points.
(3) Repeat this random selection 1000 times.
(4) Generate a *p*-value, which is equal to the percentage of means from this random distribution above m.
(5) Using statsmodels.states.multitest, correct *p*-values for multiple tests. We used the Bonferroni-Holm method to control the family-wise error rate. Consider $q < 0.1$ significant.

Therefore, each archetype was labeled with an SCLC subtype if enriched in that subtype's signature. For this work, we considered only the cells labeled with such a subtype (i.e. specialists). For the RPM tumor cells, six archetypes were originally found in Groves et al., 2022 (A/N, A2, P/Y, two Y groups, and one archetype not enriched in any signatures). Here, we consider those enriched in SCLC subtype signatures, combining the two Y groups resulting in A/N, A2, P/Y and Y.

Single sample gene set enrichment analysis (ssGSEA)

ssGSEA was performed to compute the enrichment scores for individual cells (for scRNA-seq data) or individual cell lines (for bulk RNA-seq data) (Deshmukh *et al.*, 2021; Krug *et al.*, 2019). A list of 232 epithelial signature genes, and a list of 193 mesenchymal signature genes were used to compute the E enrichment score (E score) and M enrichment score (M score) respectively (Cursons *et al.*, 2018; Panchy et al., 2022; Tan *et al.*, 2014). The combined list of 425 EMT genes has 20 genes that also appear in the 105 signature genes used for SCLC subtyping. To ensure that the inferred relationship between EMT scoring and SCLC subtyping was not simply due to the shared list of genes, we excluded the intersection of the two gene sets and performed additional ssGSEA scoring for each dataset. The patterns of the scoring were not altered by the exclusion.

Non-negative principal component analysis (nnPCA)

To quantify the divergent progression of EMT across the subtypes of SCLC cells, we used a second approach to compute the E and M scores. We performed nnPCA using the gene sets mentioned above (Panchy et al., 2021; Panchy *et al.*, 2022). nnPCA determines the approximately orthogonal axes with non-negative coefficients (loadings) for features (genes). Variances of projections of data points (cell or cell lines) onto these axes are maximized via an optimization method (Sigg and Buhmann, 2008). To select the principal components (PCs) that best represent the EMT programs across SCLC subtypes, we used two criteria to rank the PCs in a semi-supervised manner. We first selected the top five PCs that have the highest variances explained for individual samples (cells or cell lines). Among the five PCs, we re-rank them based on the variances of means of individual SCLC subtypes. Similar to ssGSEA, we excluded the EMT gene set and the SCLC gene set and performed additional nnPCA scoring for each dataset. The ranks of the PCs and the patterns of the nnPCA were not altered by the exclusion.

Conflicts of Interest

V.Q. is an Academic co-Founder and equity holder for Parthenon Therapeutics, Inc. and Duet BioSystems, Inc. D.R.T. is on the Scientific Advisory Board of VRise Therapeutics. The other authors declare no conflict of interest.

Acknowledgement and Funding

This work was supported by funds from National Institutes of Health R01GM140462 (T.H.), R50CA24378 (D.R.T.), U54CA217450 (V.Q and S.M.G), and from NSF DGE-1445197 (S.M.G).

Supplementary Figures

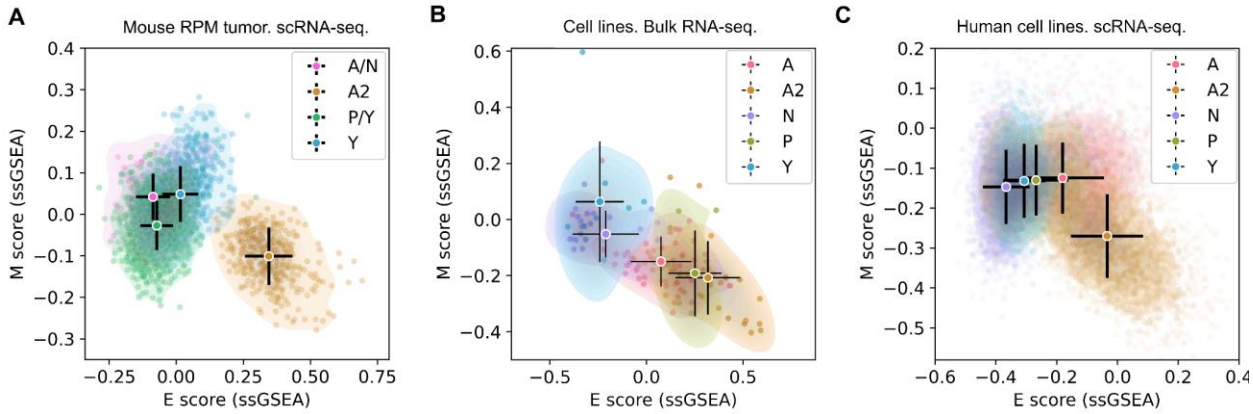


Figure S1. **Distributions of SCLC cells across the epithelial and mesenchymal spectrum using non-overlapping gene sets**. **A**. E and M scores of 15138 mouse *Myc*-driven tumor cells in an *ex vivo* model. Archetype analysis was performed to determine generalist cells and specialist cells of four types (SCLC subtypes). Scores were computed with ssGSEA and 405 previously identified EMT genes that were not used for SCLC subtyping. **B**. E and M scores of 120 SCLC cell lines computed with the same method as in A. **C**. E and M scores of 13945 cells from 8 SCLC cell lines computed with the same method as in A.

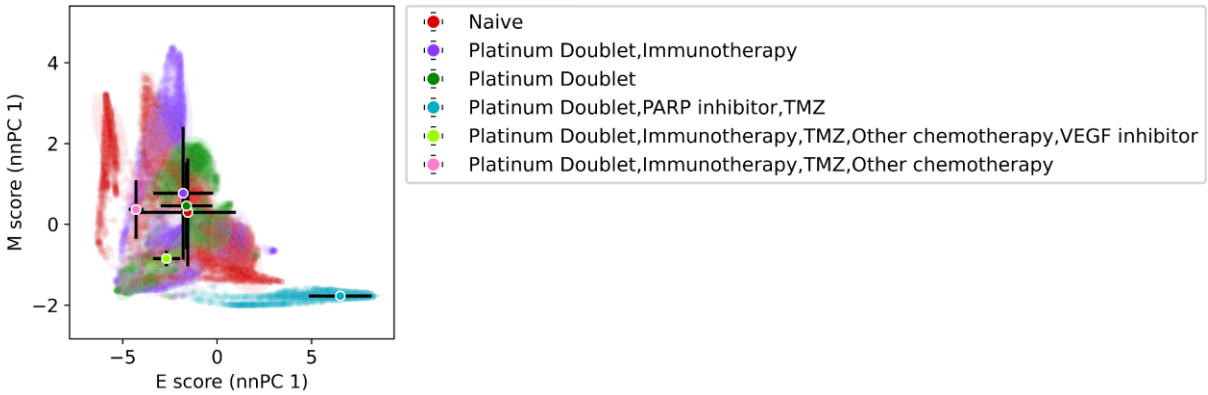


Figure S2. Treatments of human tumor cells visualized in EMT space. Scatter plot shows nnPCA-based E and M scores for 54,523 SCLC cells (Chan *et al.*, 2021). Color code represents each of the treatment types. Circles and black bars show means and standard deviations of cell scores for each treatment type.

References

Altschuler, S.J., and Wu, L.F. (2010). Cellular heterogeneity: do differences make a difference? Cell *141*, 559-563.

Borges, M., Linnoila, R.I., Van De Velde, H.J.K., Chen, H., Nelkin, B.D., Mabry, M., Baylin, S.B., and Ball, D.W. (1997). An achaete-scute homologue essential for neuroendocrine differentiation in the lung. Nature *386*, 852-855.

Calbo, J., van Montfort, E., Proost, N., van Drunen, E., Beverloo, H.B., Meuwissen, R., and Berns, A. (2011). A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. Cancer Cell *19*, 244-256.

Celià-Terrassa, T., Bastian, C., Liu, D., Ell, B., Aiello, N.M., Wei, Y., Zamalloa, J., Blanco, A.M., Hang, X., Kunisky, D., et al. (2018). Hysteresis control of epithelial-mesenchymal transition dynamics conveys a distinct program with enhanced metastatic ability. Nat. Commun. *9*, 5005. 10.1038/s41467-018-07538-7.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., and Larsson, E. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401-404.

Chan, J.M., Quintanal-Villalonga, Á., Gao, V.R., Xie, Y., Allaj, V., Chaudhary, O., Masilionis, I., Egger, J., Chow, A., and Walle, T. (2021). Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. Cancer Cell *39*, 1479-1496.

Cieply, B., Farris, J., Denvir, J., Ford, H.L., and Frisch, S.M. (2013). Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. Cancer Res. *73*, 6299-6309. 10.1158/0008-5472.can-12-4082.

Cook, D.P., and Vanderhyden, B.C. (2020). Context specificity of the EMT transcriptional response. Nat. Commun. *11*, 1-9.

Cursons, J., Pillman, K.A., Scheer, K.G., Gregory, P.A., Foroutan, M., Hediyeh-Zadeh, S., Toubia, J., Crampin, E.J., Goodall, G.J., Bracken, C.P., and Davis, M.J. (2018). Combinatorial Targeting by MicroRNAs Co-ordinates Post-transcriptional Control of EMT. Cell Systems *7*, 77-91.e77. 10.1016/j.cels.2018.05.019.

Deshmukh, A.P., Vasaikar, S.V., Tomczak, K., Tripathi, S., Den Hollander, P., Arslan, E., Chakraborty, P., Soundararajan, R., Jolly, M.K., and Rai, K. (2021). Identification of EMT signaling cross-talk and gene regulatory networks by single-cell RNA sequencing. Proc. Natl. Acad. Sci. USA *118*.

Fischer, K.R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S.T., Choi, H., El Rayes, T., Ryu, S., Troeger, J., et al. (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature *527*, 472-476. 10.1038/nature15748.

Gay, C.M., Stewart, C.A., Park, E.M., Diao, L., Groves, S.M., Heeke, S., Nabet, B.Y., Fujimoto, J., Solis, L.M., and Lu, W. (2021). Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. Cancer Cell *39*, 346-360.

Gazdar, A.F., Bunn, P.A., and Minna, J.D. (2017). Small-cell lung cancer: what we know, what we need to know and the path forward. Nature Reviews Cancer *17*, 725-737.

Grande, M.T., Sanchez-Laorden, B., Lopez-Blau, C., De Frutos, C.A., Boutet, A., Arevalo, M., Rowe, R.G., Weiss, S.J., Lopez-Novoa, J.M., and Nieto, M.A. (2015). Snail1-induced partial epithelial-to-mesenchymal transition drives renal fibrosis in mice and can be targeted to reverse established disease. Nat. Med. *21*, 989-997. 10.1038/nm.3901.

Groves, S.M., Ildefonso, G.V., McAtee, C.O., Ozawa, P.M.M., Ireland, A.S., Stauffer, P.E., Wasdin, P.T., Huang, X., Qiao, Y., and Lim, J.S. (2022). Archetype tasks link intratumoral heterogeneity to plasticity and cancer hallmarks in small cell lung cancer. Cell Systems.

Han, Y., Villarreal-Ponce, A., Gutierrez Jr, G., Nguyen, Q., Sun, P., Wu, T., Sui, B., Berx, G., Brabletz, T., and Kessenbrock, K. (2022). Coordinate control of basal epithelial cell fate and stem cell maintenance by core EMT transcription factor Zeb1. Cell Reports *38*, 110240.

Hart, Y., Sheftel, H., Hausser, J., Szekely, P., Ben-Moshe, N.B., Korem, Y., Tendler, A., Mayo, A.E., and Alon, U. (2015). Inferring biological tasks using Pareto analysis of high-dimensional data. Nat. Methods *12*, 233-235.

Hong, T., Watanabe, K., Ta, C.H., Villarreal-Ponce, A., Nie, Q., and Dai, X. (2015). An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States. PLoS Comput. Biol. *11*, e1004569. 10.1371/journal.pcbi.1004569.

Ireland, A.S., Micinski, A.M., Kastner, D.W., Guo, B., Wait, S.J., Spainhower, K.B., Conley, C.C., Chen, O.S., Guthrie, M.R., and Soltero, D. (2020). MYC drives temporal evolution of small cell lung cancer subtypes by reprogramming neuroendocrine fate. Cancer Cell *38*, 60-78.

Ito, T., Kudoh, S., Ichimura, T., Fujino, K., Hassan, W.A.M.A., and Udaka, N. (2017). Small cell lung cancer, an epithelial to mesenchymal transition (EMT)-like cancer: significance of inactive Notch signaling and expression of achaete-scute complex homologue 1. Hum. Cell *30*, 1-10.

Jordan, N.V., Bardia, A., Wittner, B.S., Benes, C., Ligorio, M., Zheng, Y., Yu, M., Sundaresan, T.K., Licausi, J.A., and Desai, R. (2016). HER2 expression identifies dynamic functional states within circulating breast cancer cells. Nature *537*, 102-106.

Kalluri, R., and Weinberg, R.A. (2009). The basics of epithelial-mesenchymal transition. J. Clin. Invest. *119*, 1420-1428. 10.1172/JCI39104.

Ko, J., Winslow, M.M., and Sage, J. (2021). Mechanisms of small cell lung cancer metastasis. EMBO Mol. Med. *13*, e13122.

Krebs, A.M., Mitschke, J., Lasierra Losada, M., Schmalhofer, O., Boerries, M., Busch, H., Boettcher, M., Mougiakakos, D., Reichardt, W., Bronsert, P., et al. (2017). The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. Nat. Cell Biol. *19*, 518.

Krug, K., Mertins, P., Zhang, B., Hornbeck, P., Raju, R., Ahmad, R., Szucs, M., Mundt, F., Forestier, D., and Jane-Valbuena, J. (2019). A Curated Resource for Phosphosite-specific Signature Analysis*[S]. Mol. Cell. Proteomics *18*, 576-593.

Kwon, M.-c., Proost, N., Song, J.-Y., Sutherland, K.D., Zevenhoven, J., and Berns, A. (2015). Paracrine signaling between tumor subclones of mouse SCLC: a critical role of ETS transcription factor Pea3 in facilitating metastasis. Genes Dev. *29*, 1587-1592.

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

Lu, M., Jolly, M.K., Levine, H., Onuchic, J.N., and Ben-Jacob, E. (2013). MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. Proc. Natl. Acad. Sci. U. S. A. *110*, 18144-18149.

Mani, S.a., Guo, W., Liao, M.-J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704-715. 10.1016/j.cell.2008.03.027.

Mendez, M.G., Kojima, S.-I., and Goldman, R.D. (2010). Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. The FASEB Journal *24*, 1838-1851.

Noguchi, M., Furukawa, K.T., and Morimoto, M. (2020). Pulmonary neuroendocrine cells: physiology, tissue homeostasis and disease. Dis. Model. Mech. *13*, dmm046920.

Panchy, N., Watanabe, K., and Hong, T. (2021). Interpretable, Scalable, and Transferrable Functional Projection of Large-Scale Transcriptome Data Using Constrained Matrix Decomposition. Frontiers in Genetics *12*, 1555.

Panchy, N., Watanabe, K., Takahashi, M., Willems, A., and Hong, T. (2022). Comparative single-cell transcriptomes of dose and time dependent epithelial-mesenchymal spectrums. NAR Genomics and Bioinformatics *In press*. 10.1101/2022.05.06.490972.

Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., and Lemaire, S. (2018). Identification of the tumour transition states occurring during EMT. Nature *556*, 463-468.

Rudin, C.M., Poirier, J.T., Byers, L.A., Dive, C., Dowlati, A., George, J., Heymach, J.V., Johnson, J.E., Lehman, J.M., and MacPherson, D. (2019). Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nature Reviews Cancer *19*, 289-297.

Sáez-Ayala, M., Montenegro, M.F., Sánchez-del-Campo, L., Fernández-Pérez, M.P., Chazarra, S., Freter, R., Middleton, M., Piñero-Madrona, A., Cabezas-Herrera, J., and Goding, C.R. (2013). Directed phenotype switching as an effective antimelanoma strategy. Cancer Cell *24*, 105-119.

Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., and Xiao, M. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature *546*, 431.

Shue, Y.T., Lim, J.S., and Sage, J. (2018). Tumor heterogeneity in small cell lung cancer defined and investigated in pre-clinical mouse models. Translational lung cancer research *7*, 21.

Sigg, C.D., and Buhmann, J.M. (2008). Expectation-maximization for sparse and non-negative PCA. 2008. pp. 960-967.

Song, H., Yao, E., Lin, C., Gacayan, R., Chen, M.-H., and Chuang, P.-T. (2012). Functional characterization of pulmonary neuroendocrine cells in lung development, injury, and tumorigenesis. Proc. Natl. Acad. Sci. USA *109*, 17531-17536.

Stewart, C.A., Gay, C.M., Xi, Y., Sivajothi, S., Sivakamasundari, V., Fujimoto, J., Bolisetty, M., Hartsfield, P.M., Balasubramaniyan, V., and Chalishazar, M.D. (2020). Single-cell analyses reveal increased intratumoral heterogeneity after the onset of therapy resistance in small-cell lung cancer. Nature Cancer *1*, 423-436.

Stovold, R., Meredith, S.L., Bryant, J.L., Babur, M., Williams, K.J., Dean, E.J., Dive, C., Blackhall, F.H., and White, A. (2013). Neuroendocrine and epithelial phenotypes in small-cell lung cancer: implications for metastasis and survival in patients. Br. J. Cancer *108*, 1704-1711.

Sutherland, K.D., Ireland, A.S., and Oliver, T.G. (2022). Killing SCLC: insights into how to target a shapeshifting tumor. Genes Dev. *36*, 241-258.

Tan, T.Z., Miow, Q.H., Miki, Y., Noda, T., Mori, S., Huang, R.Y.J., and Thiery, J.P. (2014). Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. EMBO Mol. Med. *6*, 1279-1293.

Thiery, J.P., Acloque, H., Huang, R.Y.J., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871-890. 10.1016/j.cell.2009.11.007.

Van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., and Pattabiraman, D. (2018). Recovering gene interactions from single-cell data using data diffusion. Cell *174*, 716-729.

Watanabe, K., Panchy, N., Noguchi, S., Suzuki, H., and Hong, T. (2019). Combinatorial perturbation analysis reveals divergent regulations of mesenchymal genes during epithelial-to-mesenchymal transition. npj Syst Biol Appl *5*, 21.

Wellner, U., Schubert, J., Burk, U.C., Schmalhofer, O., Zhu, F., Sonntag, A., Waldvogel, B., Vannier, C., Darling, D., and Zur Hausen, A. (2009). The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. Nat. Cell Biol. *11*, 1487.

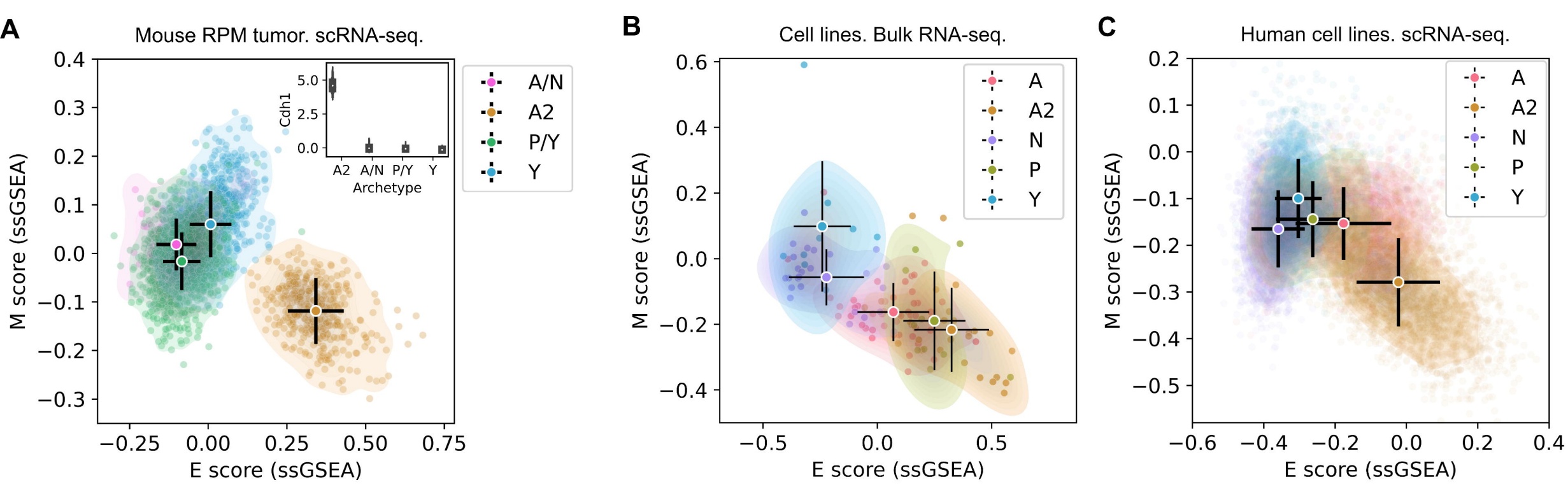
Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 1-5.

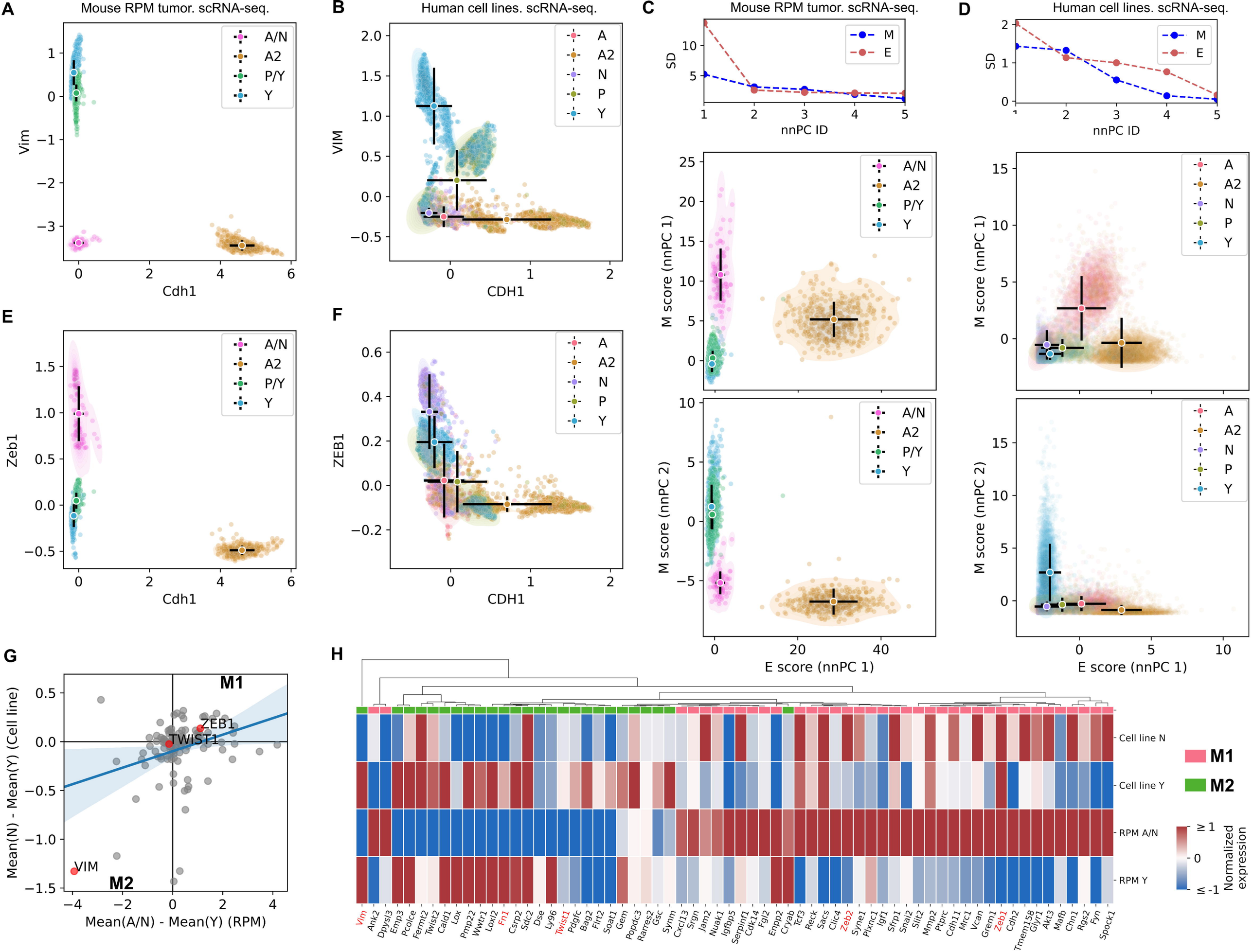
Wooten, D.J., Groves, S.M., Tyson, D.R., Liu, Q., Lim, J.S., Albert, R., Lopez, C.F., Sage, J., and Quaranta, V. (2019). Systems-level network modeling of Small Cell Lung Cancer subtypes identifies master regulators and destabilizers. PLoS computational biology *15*, e1007343.

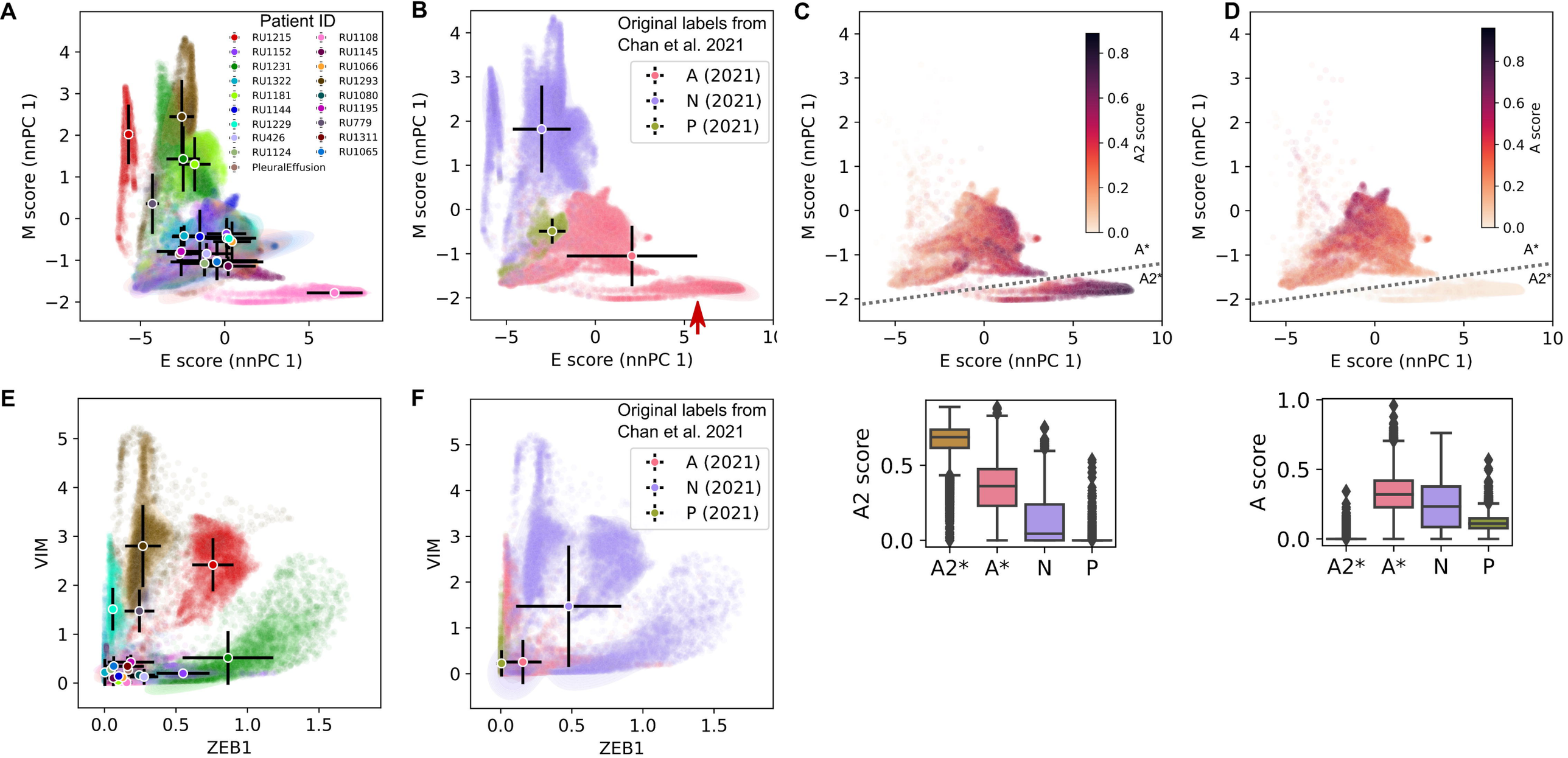
Yazawa, T. (2015). Recent advances in histogenesis research of lung neuroendocrine cancers: Evidence obtained from functional analyses of primitive neural/neuroendocrine cell-specific transcription factors. Pathol. Int. *65*, 277-285.

Ye, X., Tam, W.L., Shibue, T., Kaygusuz, Y., Reinhardt, F., Ng Eaton, E., and Weinberg, R.A. (2015). Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. Nature *525*, 256-260. 10.1038/nature14897.

Zhang, J., Tian, X.J., Zhang, H., Teng, Y., Li, R., Bai, F., Elankumaran, S., and Xing, J. (2014). TGF-β -induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. Sci. Signal. *7*, ra91-ra91. 10.1126/scisignal.2005304.


A
Mouse RPM tumor. scRNA-seq.
M score (ssGSEA)
E score (ssGSEA)
Cdh1
Archetype
A2
A/N
P/Y
Y
B
Cell lines. Bulk RNA-seq.
M score (ssGSEA)
E score (ssGSEA)
A
A2
N
P
Y
C
Human cell lines. scRNA-seq.
M score (ssGSEA)
E score (ssGSEA)
A
A2
N
P
Y


A
Mouse RPM tumor. scRNA-seq.
Vim
Cdh1
A/N
A2
P/Y
Y
B
Human cell lines. scRNA-seq.
VIM
CDH1
A
A2
N
P
Y
C
Mouse RPM tumor. scRNA-seq.
SD
nnPC ID
M
E
D
Human cell lines. scRNA-seq.
M score (nnPC 1)
M score (nnPC 2)
E score (nnPC 1)
E
Zeb1
F
ZEB1
G
Mean(N) - Mean(Y) (Cell line)
Mean(A/N) - Mean(Y) (RPM)
M1
M2
ZEB1
TWIST1
VIM
H
Cell line N
Cell line Y
RPM A/N
RPM Y
Normalized expression
Vim
Ank2
Dpysl3
Emp3
Pcolce
Fermt2
Twist2
Cald1
Lox
Pmp22
Wwtr1
Loxl2
Fn1
Csrp2
Sdc2
Dse
Ly96
Twist1
Pdgfc
Bag2
Flrt2
Soat1
Gem
Popdc3
Rarres2
Gsc
Synm
Cxcl13
Srgn
Jam2
Nuak1
Igfbp5
Serpinf1
Cdk14
Fgl2
Enpp2
Cryab
Tcf3
Reck
Sacs
Clic4
Zeb2
Syne1
Plxnc1
Igf1
Sfrp1
Snai2
Slit2
Mmp2
Ptprc
Cdh11
Mrc1
Vcan
Grem1
Zeb1
Cdh2
Tmem158
Glyr1
Akt3
Mafb
Chn1
Rgs2
Fyn
Spock1


A
Patient ID
RU1215
RU1152
RU1231
RU1322
RU1181
RU1144
RU1229
RU426
RU1124
PleuralEffusion
RU1108
RU1145
RU1066
RU1293
RU1080
RU1195
RU779
RU1311
RU1065
M score (nnPC 1)
E score (nnPC 1)
B
Original labels from
Chan et al. 2021
A (2021)
N (2021)
P (2021)
M score (nnPC 1)
E score (nnPC 1)
C
A2 score
A*
A2*
M score (nnPC 1)
E score (nnPC 1)
D
A score
A*
A2*
M score (nnPC 1)
E score (nnPC 1)
E
VIM
ZEB1
F
Original labels from
Chan et al. 2021
A (2021)
N (2021)
P (2021)
VIM
ZEB1
A2 score
A2*
A*
N
P
A score
A2*
A*
N
P