

Novel Hybrid Phenotype Revealed in Small Cell Lung Cancer by a Transcription Factor Network Model That Can Explain Tumor Heterogeneity

Akshata R. Udyavar, David J. Wooten, Megan Hoeksema, Mukesh Bansal, Andrea Califano, Lourdes Estrada, Santiago Schnell, Jonathan M. Irish, Pierre P. Massion, and Vito Quaranta

Abstract

Small cell lung cancer (SCLC) is a devastating disease due to its propensity for early invasion and refractory relapse after initial treatment response. Although these aggressive traits have been associated with phenotypic heterogeneity, our understanding of this association remains incomplete. To fill this knowledge gap, we inferred a set of 33 transcription factors (TF) associated with gene signatures of the known neuroendocrine/epithelial (NE) and non-neuroendocrine/mesenchymal-like (ML) SCLC phenotypes. The topology of this SCLC TF network was derived from prior knowledge and was simulated using Boolean modeling. These simulations predicted that the network settles into attractors, or TF expression patterns, that correlate with NE or ML phenotypes, suggesting that TF network dynamics underlie the emergence of heterogeneous SCLC phe-

notypes. However, several cell lines and patient tumor specimens failed to correlate with either the NE or ML attractors. By flow cytometry, single cells within these cell lines simultaneously expressed surface markers of both NE and ML differentiation, confirming the existence of a "hybrid" phenotype. Upon exposure to standard-of-care cytotoxic drugs or epigenetic modifiers, NE and ML cell populations converged toward the hybrid state, suggesting possible escape from treatment. Our findings indicate that SCLC phenotypic heterogeneity can be specified dynamically by attractor states of a master regulatory TF network. Thus, SCLC heterogeneity may be best understood as states within an epigenetic landscape. Understanding phenotypic transitions within this landscape may provide insights to clinical applications. *Cancer Res*; 77(5); 1063–74. ©2016 AACR.

Introduction

Small cell lung cancer (SCLC), accounting for ~13% of lung cancers (1), is exceptionally aggressive. Patients with extensive disease die ~1 year from diagnosis, and patients with limited disease experience a dismal 20% cure rate (2–4). Standard of care (2), confined to chemotherapy and radiotherapy for half a century, is largely ineffective as SCLC patients exhibit high initial response rates rapidly followed by treatment-refractory relapse.

Expression-based subtyping, impactful in other cancers (5), may be effective in SCLC because of phenotypic variability (2, 6) with respect to neuroendocrine features of its cell of origin (7, 8). A recent study (9) identified two transcriptional SCLC subtypes distinguishable by Notch pathway activity and aggressiveness, but without mutational differences. In genetic mouse models of SCLC, Calbo and colleagues showed that spontaneously occur-

ring neuroendocrine and non-neuroendocrine cell phenotypes coexist and cooperate to promote metastasis (8).

These reports indicate that a deeper understanding of cellular phenotypes could produce insights into biology and evolution of SCLC. A limitation of previous studies (8, 9) is that analyses were based on population averages, whereas variability in tumors should be considered at the single-cell level (5, 10, 11). It also remains unclear why this heterogeneity emerges.

To fill these knowledge gaps, we investigate SCLC phenotypic heterogeneity at the single-cell level using an integrative computational and experimental approach. Consistent with previous reports, we found two transcriptional subtypes at the population level in cell lines and patient specimens, characterized by gene coexpression modules enriched in neuroendocrine/epithelial (NE) and mesenchymal-like (ML) features. To understand how these phenotypes may arise in the absence of driving mutations (9), we hypothesized that they are attractors of a regulatory transcription factor (TF) network. This approach is grounded in the mathematical interpretation of Waddington's epigenetic landscape (12, 13), whereby attractors correspond to biological differentiation states or stable phenotypes. Based on this view, it has previously been proposed that malignant phenotypes in cancer correspond to attractors (14, 15), and some have suggested "differentiation therapy" from malignant to benign attractors as a possible treatment strategy (14–17).

To this end, we construct an SCLC master regulatory network of TFs from NE and ML gene-expression signatures. We then adopt a discrete Boolean modeling approach to simulate

Vanderbilt University School of Medicine, Nashville, Tennessee.

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Corrected online February 5, 2019.

A.R. Udyavar, D.J. Wooten, P.P. Massion, and Vito Quaranta contributed equally to this article.

Corresponding Author: Vito Quaranta, Vanderbilt University School of Medicine, 446B PRB, 2220 Pierce Avenue, Nashville, TN 37232. Phone: 615-936-2868; E-mail: vito.quaranta@vanderbilt.edu

doi: 10.1158/0008-5472.CAN-16-1467

©2016 American Association for Cancer Research.

the behavior of this TF network and evaluate its ability to dynamically control NE and ML phenotypes. Discrete models are well suited to provide insight into complex TF networks by identifying steady state TF patterns of expression, termed attractors. Mathematically, attractors represent the stable configurations of the dynamic TF network. Biologically, attractors correspond to transcriptional steady states of an epigenetic landscape formed by active (ON) and silent (OFF) TFs regulating each other. While discrete models are coarse approximations, they are nevertheless informative and circumvent the obstacle of unfeasible parameter acquisition (18, 19).

Simulations of our SCLC TF network predict attractors corresponding to the NE and ML SCLC subtypes. Furthermore, by distilling the NE and ML states to their core driving TFs, the model highlighted a shortcoming of the two-subtype classification, as several cell lines and patient samples did not match any attractors. Western blots revealed that these samples expressed similar levels of both NE and ML markers. Flow cytometry revealed that this "double-positive" phenotype reflected the character of individual single cells, confirming the existence of a previously unreported "hybrid" single-cell phenotype in SCLC. Exposure to cytotoxic and epigenetic drugs caused NE and ML cells to transition toward the hybrid state, implicating it as a refuge for survival of treated SCLC tumors.

Materials and Methods

Data normalization

The Cancer Cell Line Encyclopedia (CCLE) dataset (20) was downloaded from the Broad Institute as CEL files. Data were normalized and median centered using quantile RMA normalization using Affy Bioconductor package (21) in R. Probe-level data for all the datasets were converted to gene-level data by probe merging using collapseRows (22). Probes with no known gene symbols were removed.

Consensus clustering

Consensus clustering was performed using ConsensusClusterPlus v1.24.0 package in R v3.2.3 (23) on both the 53 SCLC cell line dataset from CCLE (20) and 28 SCLC patients from the Clinical Lung Cancer Genome Project (CLCGP; ref. 24), with 80% subsampling of both genes and samples, 1,000 repetitions, 1 - Pearson correlation, and k-means. Both the CCLE and CLCGP datasets were subsetted to only include genes measured in the GSE6044 (25) dataset, in order to maximize overlap with our previous work (26).

Weighted gene coexpression network analysis

Coexpression network analysis was performed in R v3.2.3 using the weighted gene coexpression network analysis (WGCNA) package v1.49 (27). As with the consensus clustering analysis, the CCLE SCLC dataset was subsetted to only include genes measured in GSE6044 (25). We used 1 - Pearson correlation to build a coexpression based dissimilarity matrix. Modules were generated using unsupervised average-linked hierarchical clustering with a static height of 0.95. We required that each module contain at least 100 genes.

Comparative Pathway enrichment analysis of the Blue and Turquoise modules was performed using BINGO and Enrichment map (28) and visualized in Cytoscape. Pathway enrichment for a cell line was computed by transforming expression values to

Z-scores and computing the average expression of all genes within a given pathway.

Transcriptional regulatory network construction

To generate an SCLC-specific transcriptional network, we applied the bootstrap version of ARACNE (29) on gene expression profiles from the 53 CCLE SCLC cell lines using the following parameters: $P = 10^{-7}$, $dpi = 0$ and 100 bootstraps, resulting in 27,224 interactions among 8,706 nodes. To evaluate if genes in Blue and Turquoise module are enriched for targets of any specific TF, we used Fisher exact test (30) to compute TF enrichment with the Blue and Turquoise module genes. We selected all TFs to be candidate master regulators if the Fisher exact test P value was ≤ 0.05 , leaving 96 and 207 TFs for Blue and Turquoise modules, respectively. Of these, 23 TFs were common to both modules (Supplementary Fig. S7D).

Modeling gene regulatory networks that were inferred entirely from the available data can suffer from circular reasoning: a dataset generates a network, which then predicts the dataset. To avoid this fallacy, we built the network topology strictly using sources external to the datasets of interest. Thus, we first filtered ARACNE TF predictions based on gold-standard TF-target binding site databases CHEA, ENCODE, TRANSFAC, JASPAR using EnrichR (31–35), and literature databases such as Pubmed and Glad4U (36). These filtration steps produced a list of 76 likely TF regulators of NE and/or ML differentiation. We took only heterogeneously expressed TFs (median absolute deviation above the 50th percentile) yielding a list of 38 TFs that we used to build a Boolean network for SCLC. Next, we extracted directed interactions between these TFs using only information from the above gold-standard references and literature. Where possible, interactions were classified as activating or inhibiting by manually searching the literature. Interactions that we could not find in the literature were classified as activating if the TFs are positively correlated across the CCLE dataset, or inhibitory given negative correlation (Supplementary Fig. S7B). Five TFs did not have outgoing edges and were eliminated, leaving 33 TFs.

Boolean network simulation and analysis

The TF network was simulated as a Boolean network where each node was either ON (active) or OFF (silent). Nodes were updated using the random order asynchronous method (18). We considered two distinct approaches updating TFs: (i) threshold updating, in which the total number of ON inputs are compared such that the target node is switched ON when there are more activators, and OFF when there are more inhibitors, and (ii) inhibitory dominant, in which having any inhibitor ON is sufficient to turn the target node OFF. The threshold update rule is referred to as a Hopfield neural network in some literature. Because the network's state space is so large, we only simulated a random subsample of the states. For both update rules, a state transition network was seeded with 8,000 randomly generated initial states, in which each TF had a 50% chance of being ON or OFF. When TFs are updated in asynchronous random order, it is possible that the state at time t may have several possible outgoing trajectories to time $t+1$. To account for this, each state that we observed was initialized and updated 30 times to sample distinct trajectories to the next state that may be influenced by the update order. Each newly observed state was queued to be updated in this fashion, until there were no new states identified. Attractors were identified by applying the attracting components algorithm from NetworkX

(<https://networkx.github.io/>) to the state transition graph for each update rule. Using threshold updates, we found 57 fixed-point attractors, and no oscillating ones. We observed the same set of attractors using synchronous updates with 2^{17} and 2^{18} initial states. Using the inhibitory dominant update rule, we found 6 fixed-point attractors, and 5 two-state oscillating attractors, finding the same set of attractors using synchronous updates from 2^{12} to 2^{18} initial states.

To score the correlation between samples and attractors, CCLE and CLCGP expression data were independently scaled between 0 and 1, and Pearson r was calculated pairwise between attractors and cell lines/patients. Statistical significance was determined by considering the highest correlation between each attractor and cell line, and comparing against a null distribution, obtained by generating 10,000 random TF vectors. The Mann-Whitney U test was used to compare the score distributions. The model attractors had higher correlation with cell lines than the random attractors (threshold: $P = 9.5e-34$, inhibitory-dominant: $P = 7.6e-9$), and higher correlation with patients than with the random attractors (threshold: $P = 1.7e-23$, inhibitory-dominant: $P = 1.7e-6$). Individual attractors were assigned a P value for their highest correlation with each sample by direct comparison with the null distribution (Supplementary Table S5).

Robustness of NE and ML attractor states

Cells need to be able to robustly guide their differentiation choices depending on driving signals, and therefore we would expect trajectories toward cell attractors to be robust. Structural coherence (37) is a topological metric that measures how reliably an initial condition evolves toward its appropriate attractor given a perturbation. This metric requires an estimate of the total size of the basin of attraction, so we were only able to apply it to attractors for which we could reliably estimate this size (a few basins were too small to reliably estimate their size). Values of random, maximum, observed, and structural coherence are reported in Supplementary Table S4. We also calculated Derrida curves (Supplementary Fig. S9; ref. 38), as the average growth or decay of an initial perturbation after one step, for perturbation sizes ranging from 1 to 32 TFs.

Antibodies and reagents

Antibodies used for Western blotting include EphA2, LEF1, c-MYC, NFkB1, ZEB1, MTF, CD56 and CD44 (Cellsignal), OVOL2, SOX2, ASCL1, GAPDH, and β -actin (Sigma), FOXA2 and SMAD3 (BD Biosciences). Fluorophore-conjugated primary antibodies were used for flow cytometry—CD56 BV605, CD151 PE, CD24 BUV395 (BD Biosciences), CD44 Pacific blue, CADM1 A647 (MBL), EPHA2 A488 (R&D Systems), CD133 PE-Cy7 (Biolegend).

Cell culture

Authenticated cell lines were obtained from ATCC from 2012 to 2015, authenticated by DNA STR profiling, morphology, and mycoplasma detection. Cell lines were grown in company recommended media. New cell lines were obtained as needed every 1 to 2 years. Cell lines were expanded in culture for less than 2 months, then frozen in aliquots for subsequent use. Cell lines are passaged no more than 30 times before being discarded (approximately 4 months). Any contaminated cell lines were discarded, and new

aliquots were thawed and cultured. *Mycoplasma* test was performed on all cell lines in culture every 2 weeks.

Flow cytometry data generation and analysis

One to two million cells were plated in T75 or T150 flasks the previous day and collected for flow experiment the next day as described below. For drug treatment experiments, cells were plated same as above the previous day, followed by drug addition the next day. Cells were incubated with drugs for 3 days at 37 degrees then collected for flow experiments.

Cells were dissociated using TrypLE (GIBCO) for 10 to 15 minutes followed by staining with Alexa 700 dye (Molecular Probes) for 5 minutes at 37 degrees. Cells were then washed and fixed with 2% paraformaldehyde (10 minutes at room temperature), followed by surface marker staining or permeabilization with ice-cold 100% methanol at -20 for 30 minutes. Cells were then stained with fluorescent conjugated antibodies for 30 minutes in dark at room temperature. Samples were washed with PBS and run on BD 5-laser instrument at the Vanderbilt Flow cytometry core. Fluorescent channels were compensated using anti-mouse IgK beads (BD Biosciences) that were tagged with fluorescent antibodies. First intact cells were gated on the Forward Scatter Area (FSC-A) and Side-Scatter Area (SSC-A) plot, where debris has a low FSC-SSC ratio. This was followed by gating for Alexa700-negative viable cells as described previously (Box 1 in ref. 39), where A700-positive populations are dead/dying cells. Finally, A700-negative viable cells were further gated to include only singlet populations. At least 20,000 to 30,000 viable singlet cells were collected per sample. All subsequent gates on fluorescent markers are made on viable singlet cells. The raw cytometer intensity readouts for fluorescent channels were first converted to log scale by using the $\text{asinh}()$ function with a cofactor of 150. Gating was conducted in Cytobank (40).

Western blotting

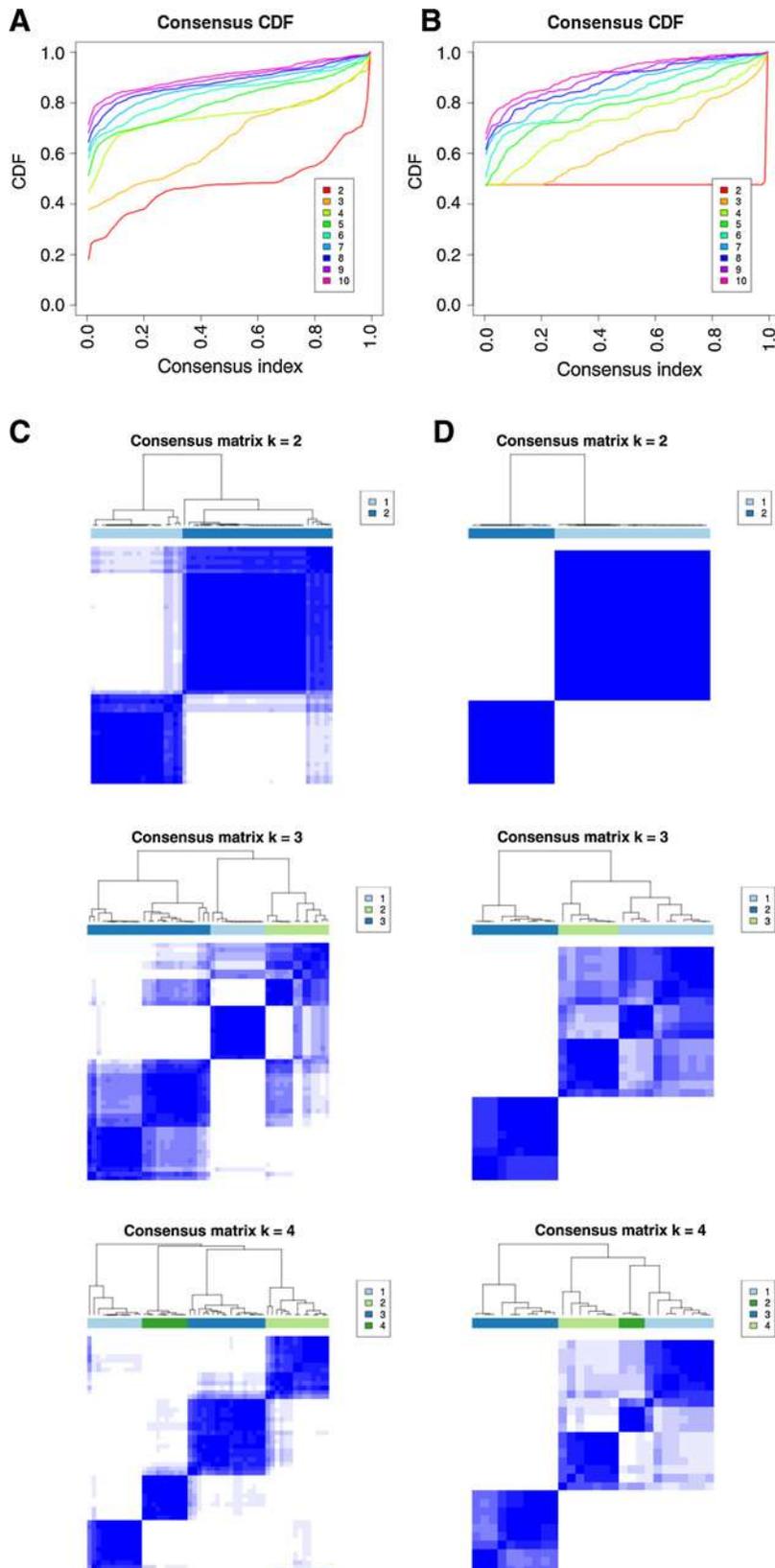
Cell lines were plated for 2 days in complete medium to equilibrate. Lysates were prepared by spinning cells at 4°C , aspirating the media, and adding M-PER lysis buffer (Pierce) containing 1X phosphatase inhibitors 2 and 3 and protease inhibitor (Sigma-Aldrich). Lysates were incubated for five minutes at room temperature, vortexed for 30 seconds and centrifuged at 15,000 rpm for 15 minutes (at 4°C). The protein concentration was quantified using BCA assay (Pierce). Lysates were boiled for 10 minutes at 100 degrees with 1X NuPage sample buffer (Molecular Probes) and run on 8% or 4% to 12% Tris-glycine gels (Molecular Probes). Semi-dry transfer was followed by blocking with 1X Casein-TBS. Blots were imaged by chemiluminescence.

Results

Characterization of SCLC NE and ML phenotypes

We applied consensus clustering (23) to the 53 SCLC cell lines in the CCLE database (20), and to 28 tumor samples from the CLCGP (24), with the expectation to detect previously observed neuroendocrine and nonneuroendocrine phenotypes (8, 9). Consensus clustering analysis, which provides a rationale for determining the number of robustly separated subtypes, indicated that the CCLE cell lines and the CLCGP patient specimens are most consistently separated into two distinct clusters (termed clusters 1 and 2; Fig. 1 and Supplementary Fig. S1, Supplementary Table S1).

Udyavar et al.

**Figure 1.**

Identification of two phenotypes in SCLC cell lines and patients. Consensus clustering was used to identify robust clusters of CCLC cell lines (**A** and **C**) and CLCGP tumor specimens (**B** and **D**). The consensus score is the frequency that a given pair of samples was placed in the same cluster over 1,000 iterations. The cumulative distribution of SCLC cell line (**A**) and patient (**B**) consensus scores is shown for values of k from 2 to 10. These results most strongly support two transcriptional subtypes in SCLC cell lines and patients, though in cell lines, four subtypes may also be a good fit. Corresponding consensus score heatmaps for SCLC cell lines (**C**) and patients (**D**) are shown for $k = 2$ to $k = 4$. Each point represents a pair of samples, colored by consensus score from white (0, never co-cluster) to blue (1, always co-cluster). $k = 5$ to $k = 8$ are shown in Supplementary Fig. S1.

To characterize these clusters, we applied WGCNA (27) to the 53 CCLE cell lines. WGCNA identified 13 gene coexpression modules (Supplementary Fig. S2A and Supplementary Table S2). To summarize the overall expression of each module, eigengenes (first principal component of all genes within a module) were computed (Fig. 2). The Blue module eigengene was upregulated (Bonferroni adjusted $P < 0.001$) within cluster 2, while the Turquoise module eigengene was upregulated (Bonferroni adjusted $P < 0.001$) within cluster 1 (Fig. 2 and Supplementary Fig. S2B). No other module eigengenes showed statistically significant differences between the consensus clusters (Supplementary Fig. S2B).

WGCNA modules are derived agnostically and reveal coexpressed genes participating in similar biology. Blue and Turquoise module genes therefore provide information about the overall character of the consensus clusters. Visual inspection reveals that hubs (highly connected genes) in the Blue and Turquoise modules are well-known NE or mesenchymal biomarkers, respectively (Supplementary Figs. S3 and S4). Statistical analyses by Gene Ontology and Enrichment Map (28) confirm enrichment of the Blue module for neuroendocrine and epithelial differentiation processes, while the Turquoise module is enriched for pathways involved in mesenchymal phenotype and epithelial-to-mesenchymal transition (EMT; Supplementary Figs. S5 and S6A). Additionally, we performed gene-set enrichment analysis (GSEA; ref. 41) with published signatures of proneural, mesenchymal, and proliferative glioblastoma subtypes (30) on the 12 cell lines with highest Blue module eigengene expression, and 9 cell lines with highest Turquoise module eigengene expression. The proneural (Supplementary Fig. S6B, bottom) and mesenchymal (Supplementary Fig. S6B, top) signatures were enriched in the cell lines with high Blue module and Turquoise module expression, respectively. As a control, the proliferative signature was significantly enriched in neither ($P > 0.1$, data not shown).

Thus, the Blue module genes are a signature of the canonical neuroendocrine state of SCLC, while the Turquoise module genes are a signature of nonneuroendocrine/mesenchymal cells. We therefore refer to cluster 2 (in which Blue module genes are high) as the SCLC NE subtype, and to the cluster 1 (in which Turquoise module genes are high) as the SCLC ML subtype (Fig. 1B and C).

Boolean simulations of TF network predict attractors corresponding to NE and ML states

Because NE and ML subtypes are not associated with driver mutations (9), we hypothesized they may be regulated epigenetically, as during normal cell differentiation. Cell identity in differentiation is largely controlled by regulatory networks of TFs that coordinate expression of each other and of target genes (42, 43). To understand specification of NE or ML cell identity, we derived a network of TFs that regulate expression of genes within the Blue and Turquoise modules (see Materials and Methods; Supplementary Fig. S7). By pruning nodes that had no outgoing edges, we reduced this network to a core set of 33 TFs and 361 interactions (Fig. 3A and Supplementary Table S3). We used random asynchronous order Boolean simulation with both threshold and inhibitory dominant update rules to identify attractors. Using the threshold update rule, we found 57 stable fixed-point attractors (Fig. 3B) and no oscillating attractors. With the inhibitory dominant rule, we found only 6 fixed-point attractors, but 5 two-state oscillating attractors as well. Both the threshold and inhibitory dominant attractors revealed similar features

in all subsequent analyses, suggesting that these results are robust to the precise nature of the regulatory interactions (Supplementary Fig. S8). We also inferred network robustness to perturbations using structural coherence (Supplementary Table S4; ref. 37) and Derrida curves (Supplementary Fig. S9; ref. 38). These results indicate that the basins of attraction are more robust than those of a random network, and that the network dynamics are ordered, rather than chaotic. Together, these observations suggest that the internal structure of interactions imposed to the TF network (Fig. 3A) is non-random and naturally leads to well-regulated stationary states, and that our coarse modeling approach is acceptable for this system.

Each attractor is a 33-dimensional vector of TFs, differing by the overall TF ON-OFF expression pattern (columns in Fig. 3B). Many of the attractors differ only by one or a few TFs. Hierarchical clustering segregated the threshold attractors into four distinct clusters (Fig. 3B). Results from the inhibitory dominant network are qualitatively similar and are shown in Supplementary Fig. S8. Based on the TF ON-OFF patterning in each attractor within the clusters, attractors 49 to 25 (along the dendrogram, Fig. 3B) have active TFs known to be either neuroendocrine (INSM1, POU3F2, SOX2, SOX11) or epithelial (FOXA2, OVOL2). In contrast, attractors 35 to 34 (along the dendrogram, Fig. 3B) contain active TFs (MYC, NFKB1, SMAD3) involved in mesenchymal differentiation or EMT.

We computed a correlation score between attractors and both cell lines and patients by scaling the CCLE (Fig. 3C) and CLCGP (not shown) expression data from 0 to 1 and calculating Pearson correlation coefficient pairwise between each attractor and sample. Several attractors exhibited high correlation with cell lines within the NE consensus cluster, while others exhibited high correlation with cell lines within the ML consensus cluster (Fig. 3D). Similar results were observed with patients (Fig. 3E), and in both cases the model's attractors were found to be significantly more correlated to the samples than random (Supplementary Fig. S10 and Supplementary Table S5).

These results confirm that the simulated dynamics of the 33 TF network agree well with the NE and ML nature of the cell lines, suggesting that the gene expression signatures of these phenotypes is driven by these underlying NE and ML TF attractors. Nevertheless, a few attractors showed no significant correlation with any cell line or tumor. These attractors were not further pursued, as we do not observe them biologically. More significantly, however, several cell lines and patient samples did not have significant correlation with any attractor. These samples may represent mixed populations of NE and ML cells, leading to poor correlation with either subtype at the population level, or alternatively may represent populations of "hybrid" cell types.

Experimental validation of network attractors reveals a hybrid single-cell SCLC phenotype

We probed the expression of cell lines for 10 of the 33 TFs. In general, NE TFs were expressed at a higher level in NE than ML cell lines, and vice versa (Fig. 4A). Consistent with the finding that some cell lines did not correlate with either the NE or ML attractors (Fig. 3D and Supplementary Fig. S8B), we observed cell lines that simultaneously expressed both NE and ML TFs. Similarly, CD56 (NE marker) and CD44 (ML marker) were found coexpressed in 3 out of 10 patient samples (pt1112-1, pt216-1, and pt460-1, Fig. 4B). The other tumor samples had mutually exclusive expression of either CD56 or CD44.

Udyavar et al.

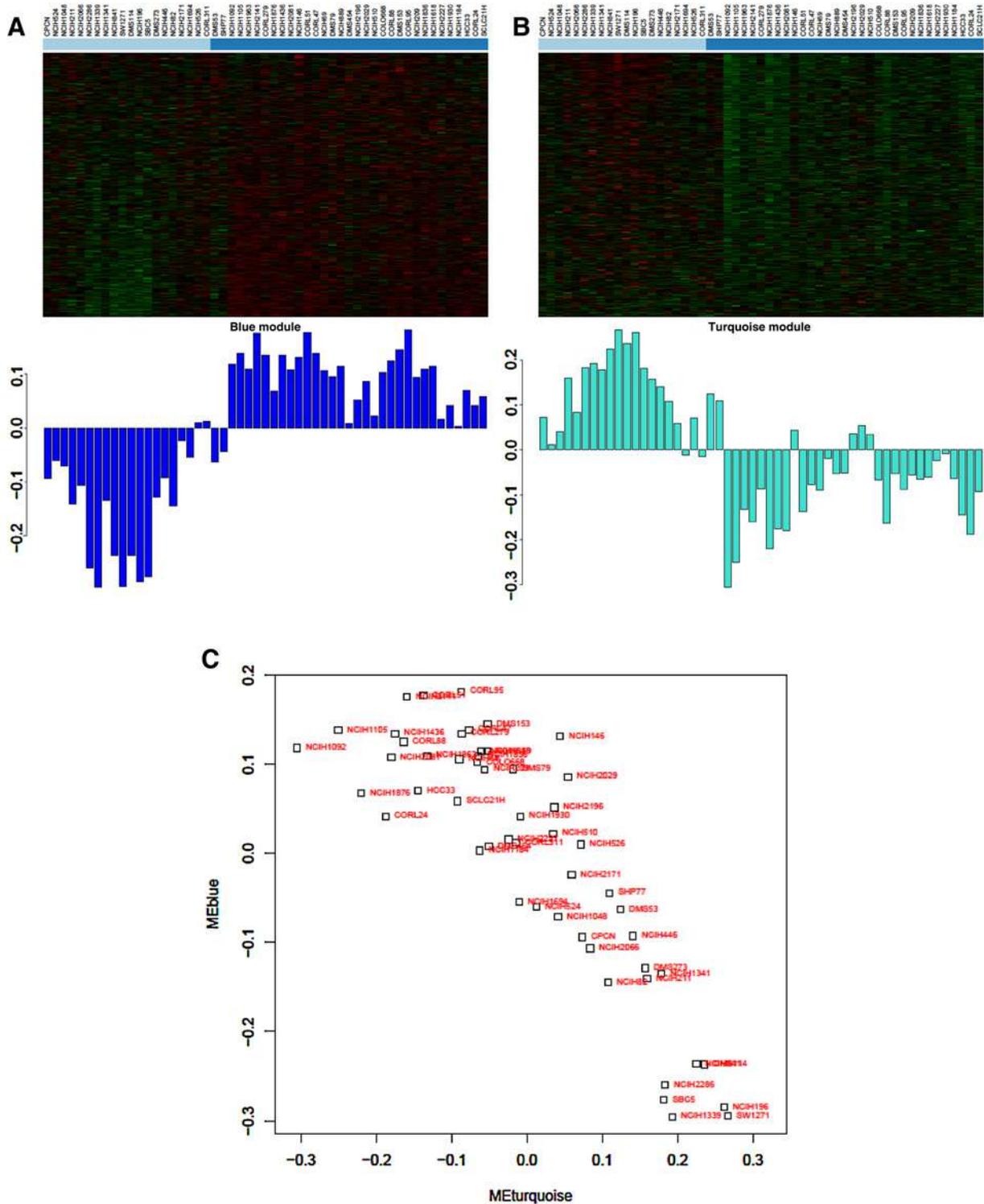
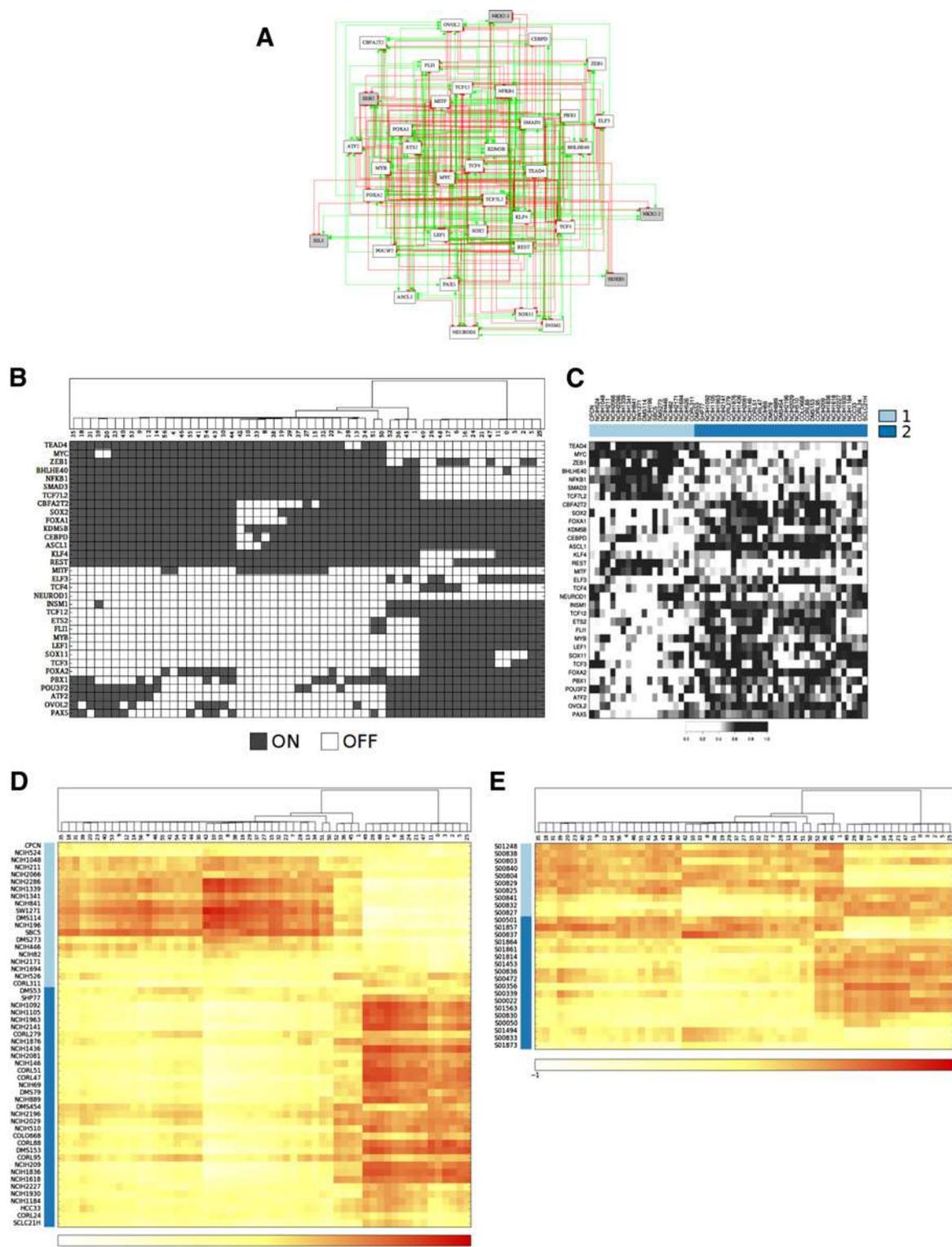
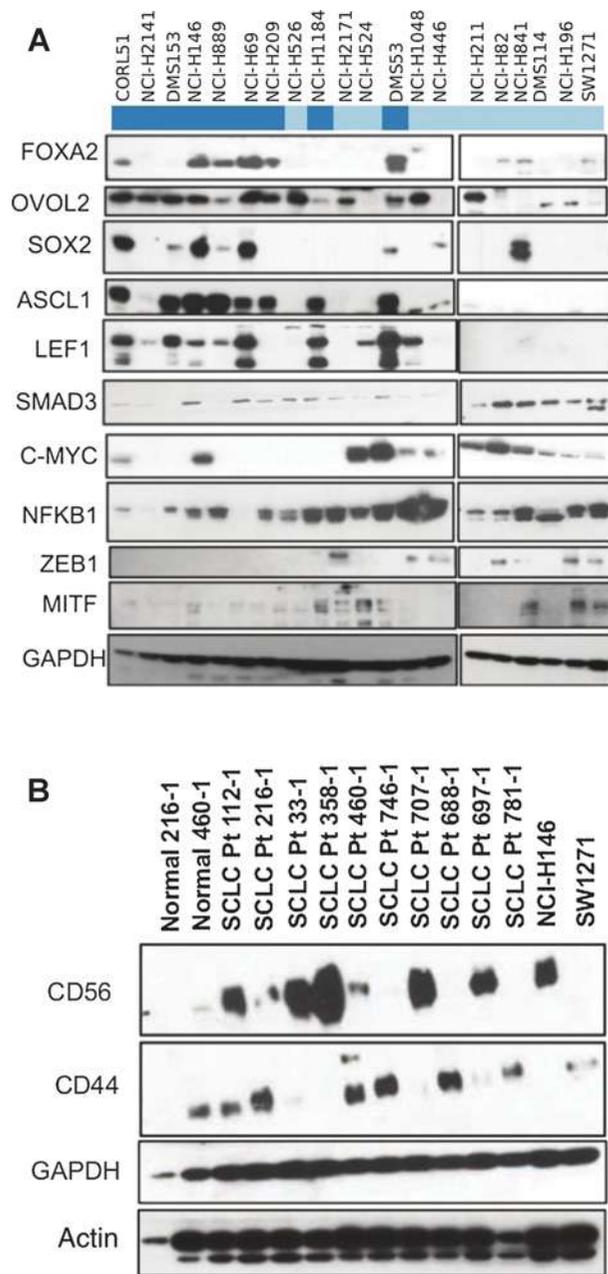


Figure 2. Two anticorrelated gene coexpression networks distinguish phenotypic clusters. **A** and **B**, Top, heatmap view of the Blue (**A**) and Turquoise (**B**) module genes (rows) across 53 SCLC cell lines (columns). Cell lines are ordered and marked as in Fig. 1C consensus clustering. Bottom, the gene expression profile for each cell line is summarized by the eigengene. The Blue and Turquoise modules had significantly different eigengene expression between the two consensus clusters. **C**, The Blue and Turquoise module eigengenes plotted for each cell line reveals anticorrelated expression of Blue vs. Turquoise module (Pearson correlation: -0.86 , $P: 1.6e-16$).

**Figure 3.**

TF network predicts NE and ML attractors. **A**, The 33 TF SCLC regulatory network. Green edges indicate activation, while red indicate inhibition. **B**, Boolean simulations identified 57 stable attractors (columns). TF expression patterns for all attractors is displayed in rows: shaded cells represent TFs that are ON; white represents OFF. **C**, TF expression in the CCLE dataset reveals similar expression patterns to *in silico* attractors. **D**, Correlation score between attractors (columns) and cell lines (rows). Positive correlation, red; no correlation, yellow; negative correlation, white. Attractors 49–25 show high correlation with the NE cell lines, while attractors 35–34 show high correlation with the ML cell lines. Nevertheless, several cell lines are uncorrelated with NE or ML attractors, possibly revealing a hybrid phenotype. **E**, Pearson correlation scores were computed as in **D** between the 57 attractors (columns) and 28 SCLC patient tumor specimens from the CLCGP (rows).

Udyavar et al.

**Figure 4.**

Experimental validation of TF network states in human SCLC. **A**, Five NE TFs (FOXA2, OVOL2, SOX2, ASCL1, and LEF1) and 5 ML TFs (SMAD3, MYC, NFKB1, ZEB1, MITF) validated by Western blot. The NE TFs show higher expression in the NE cell lines (dark blue, as in consensus clustering) than in the ML cell lines (light blue) and vice versa. Nevertheless, several cell lines showed similar levels of expression of both NE and ML TFs, including NCI-H146, NCI-H209, NCI-H1184, DMS53, NCI-H1048, and NCI-H446. Many of these cell lines, such as DMS53, also showed poor correlation with any attractor in Fig. 3 and Supplementary Fig. S8, and were near the center in Fig. 2C, suggesting a non-NE and non-ML phenotype. These cell lines are tentatively denoted as "hybrid" phenotypes. **B**, CD56 (a NE marker) and CD44 (a ML marker) were probed by Western blots in 10 SCLC patient samples. Most patients show expression of only one or the other; however, Pt112-1, Pt216-1, and Pt460-1 show double positive expression of both markers, suggesting the hybrid phenotype may be important in patients as well.

These double-positive cell lines may be either composed of mixed populations of NE and ML cells, or hybrid single cells simultaneously coexpressing features of both phenotypes. We investigated this using single-cell flow cytometry with well-established NE and mesenchymal differentiation biomarkers (Fig. 5 legend). To aid visualization, we defined overall NE and ML scores as the unweighted sum of NE and ML biomarkers, respectively.

In biaxial plots of these scores, NE cell lines consist primarily of NE⁺ML⁻ single cells, while ML cell line single cells were NE⁻ML⁺ (Fig. 5A and C). However, in several cell lines, single cells simultaneously expressed similar levels of both NE and ML biomarkers (Fig. 5B). These results confirm the existence of a novel SCLC NE⁺ML⁺ hybrid phenotype comprised of both NE and ML characteristics.

Other reports of hybrid phenotypes in cancer have associated the hybrid cells with a more stem-like phenotype (44, 45). We measured single-cell expression of CD133, a cancer stem-cell marker, and found no significant difference between the NE, ML, and hybrid cell lines (Supplementary Fig. S11).

Modulation of SCLC phenotypes with chemotherapy or epigenetic drugs

Next, SCLC cell lines were treated with etoposide and cisplatin, and with epigenetic modulators valproic acid (HDAC inhibitor) and 5-azacytidine (DNA methylation inhibitor). Flow cytometry measurement of NE and ML markers (see Fig. 5 legend) was used to characterize how treatment shifted the phenotypic identity of SCLC cells. Biaxial plots of NE vs. ML scores showed phenotypic shifts at the single-cell level for all perturbations (Fig. 6), converging toward hybrid populations. The hybrid phenotype was thus an end state for SCLC cells subjected to stress.

Discussion

In this work, we identified signatures of phenotypic heterogeneity in SCLC, and a set of TFs regulating expression of these genes. Through discrete Boolean model simulations, we showed that a master SCLC TF network naturally settles into states that were identified as NE or ML. This suggests that the NE and ML cell identities can naturally emerge from regulatory dynamics of TF networks, rather than being driven by genomic mutations. Most significantly, by distilling the NE and ML phenotypes to their essential TF drivers, the model also exposed some cell lines that had been improperly classified as NE or ML through transcriptome-based clustering. *In vitro* validation confirmed existence of SCLC cell lines and patient samples that simultaneously express both NE and ML markers and TFs. We verified this hybrid phenotype at the single-cell level in cell lines and showed that cell lines transition toward this hybrid state for survival upon drug treatment.

As resolving cancer heterogeneity can have profound impact on patient care and outcomes (46), a deeper understanding of SCLC heterogeneity, i.e., NE, ML, hybrid and beyond, may translate to benefits in the clinic. Historically, attempts to separate SCLC into subtypes in the clinic were abandoned due to poor reproducibility among pathologists and unclear clinical relevance (47). However, our findings indicate that SCLC heterogeneity is dynamic, because a core TF network specifies both the NE and ML phenotypes, which may confound static associations. The hybrid phenotype may be an additional

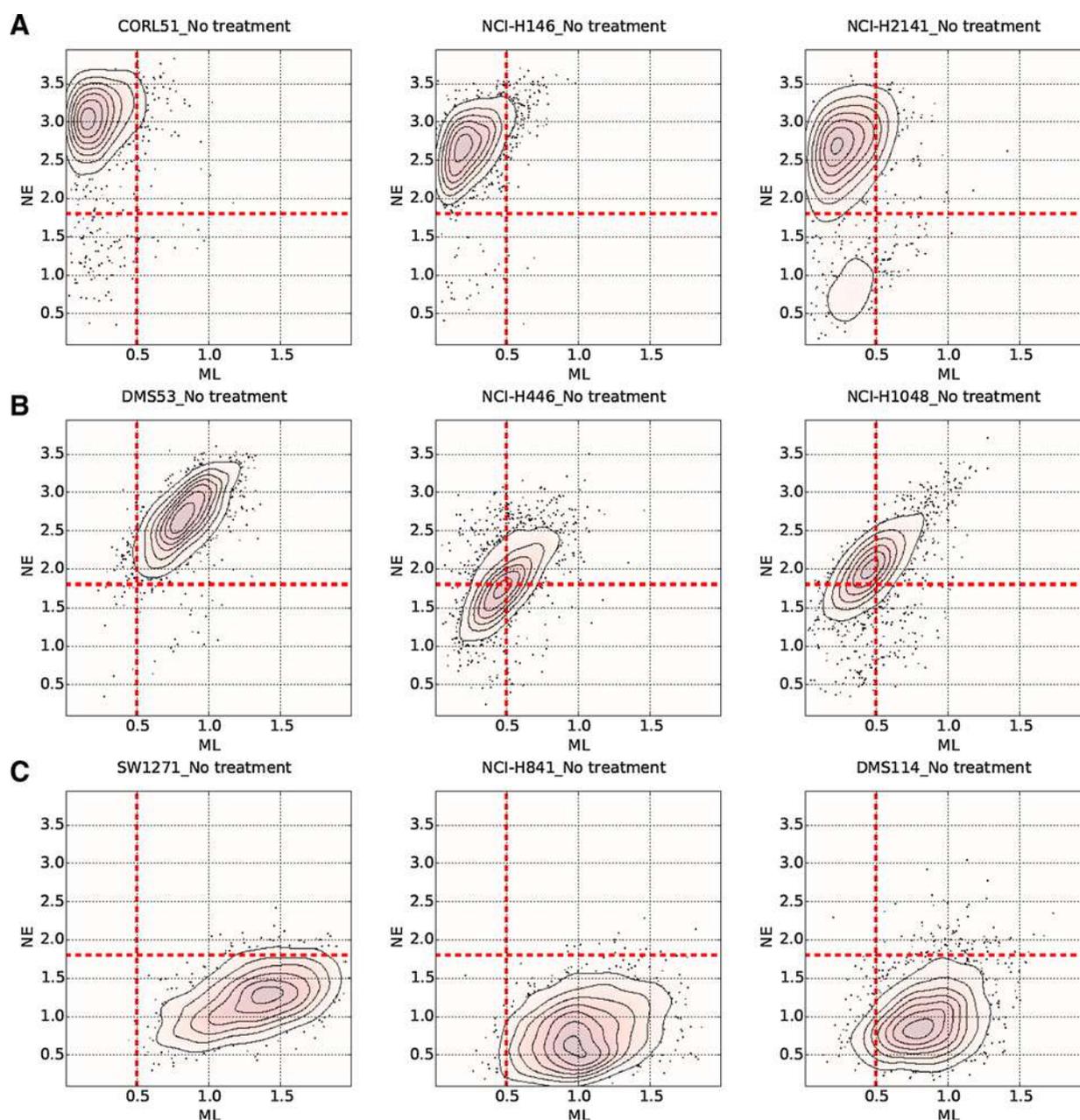


Figure 5.

Single-cell level expression of phenotypic biomarkers in SCLC cell lines reveals hybrid cells. Flow cytometry was performed using four NE surface markers (CD56, CD24, CADM1, and ALCAM) and two ML surface markers (EPHA2 and CD151). Axes represent unweighted sums of normalized NE and ML fluorescence. **A**, CORL51, NCI-H146, and NCI-H2141 were classified as NE by consensus clustering, and single cells show NE⁺ML⁻ phenotype. **B**, DMS53, NCI-H446, and NCI-H1048 showed comparable levels of expression of NE and ML TFs (Fig. 4), Blue- and Turquoise-module eigengenes (Fig. 2), and poor correlation with either NE or ML attractors (Fig. 3 and Supplementary Fig. S8). At the single-cell level, these cells also show similar levels of expression of NE and ML markers, suggesting these cells are in a hybrid NE⁺ML⁺ or NE⁻ML⁻ phenotype, not well described by a NE vs. ML dichotomy. The hybrid state is characterized by cells along the diagonal of these plots. **C**, SW1271, NCI-H841, and DMS114 were consistently classified as ML, and single cells exhibit NE⁻ML⁺ phenotype.

confounder, exacerbated by the tendency of both NE and ML cells to move toward the hybrid phenotype under treatment. We argue that SCLC heterogeneity is best interpreted as states within a phenotypic landscape, and understanding phenotypic mobility within this landscape could provide ties to clinical relevance.

While SCLC has a high mutation rate, no correlations between mutations and distinct subtypes have been observed (9). In this work, we show a non-mutational mechanism for regulation of distinct SCLC cell identities, as attractors of a TF regulatory network. This framework of equating cell types to attractors was first advanced by Waddington in his eponymous epigenetic

Udyavar et al.

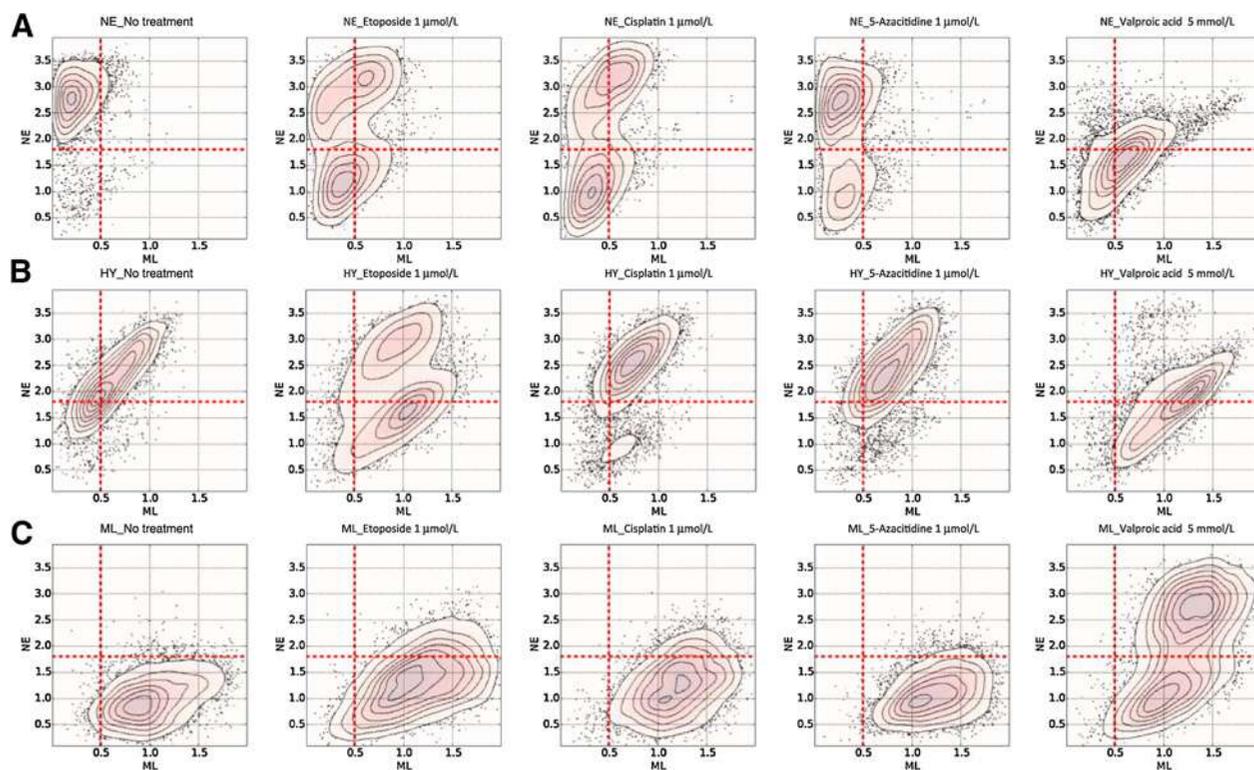


Figure 6.

Phenotypic modulation by treatment pushes cells toward hybrid state. NE, ML, and hybrid (HY) cell lines (from Fig. 5) are combined to show overall phenotypic responses. NE and ML axes are the same as defined in Fig. 5. Each column represents a unique treatment condition: Untreated (column 1); 1 $\mu\text{mol/L}$ etoposide (column 2); 1 $\mu\text{mol/L}$ cisplatin (column 3); 1 $\mu\text{mol/L}$ azacitidine (column 4); 5 mmol/L valproic acid (column 5). Cells were treated for 3 days. **A**, The NE cell lines show a shift toward the NE+ML+ or NE-ML- quadrants. **B**, The hybrid cell lines show no significant shift. **C**, The ML cell lines show a shift toward the NE+ML+ quadrant.

landscape (12), and has because been expounded upon by many mathematical biologists (13). It has further been argued that malignant cancer states and cancer heterogeneity are best described by the concept of attractors (14–16). Significantly, this framework has been shown to provide possible therapeutic strategies by identifying targets *in silico*, which most significantly perturb the attractor states (17). We do not exclude the possibility that mutations can drive SCLC heterogeneous phenotypes, but our results indicate that epigenetic causes should also be considered.

The Boolean modeling approach used here is an established way of obtaining a coarse-grained picture of large network behavior, and is well suited to identifying the stable states of TF networks (18, 19). Our approach was novel in constructing the TF network in a blind fashion from SCLC gene expression datasets, TF databases and literature, and was validated by its ability to reproduce correlations with known phenotypes. Significantly, our model revealed the existence of a previously unrecognized hybrid phenotype, by showing that some cells do not correlate to either the NE or ML attractors. However, we did not identify attractors correlated with the hybrid phenotype. Hybrid EMT phenotypes have been previously reported in NSCLC (48) and lung adenocarcinoma (49), and other groups have recently reported computational modeling of hybrid EMT phenotypes by driving EMT networks with external stimuli (44, 50). Additionally, the Boolean modeling approach cannot capture intermediate levels of expression, and therefore attractors corresponding to the hybrid state

may not be identifiable using this method. Continuum modeling approaches may be needed to better understand the hybrid state. Identification of gene coexpression modules enriched in the hybrid phenotype may also reveal additional relevant TFs. Considering all of these possibilities, current ongoing work in our laboratory is directed at identifying hybrid attractors.

Our study establishes the hybrid phenotype as a refuge for drug-treated SCLC cell lines. We anticipate that this phenotype may play a significant role in the evolution of SCLC tumors under treatment, and possibly in relapse. SCLC TF networks may serve as a guide for interventions aimed at preventing phenotypic transitions into resistant states. While we focused on SCLC here, such transcriptional regulation may play a similar role in maintaining non-mutational heterogeneity in other cancer types, and our approach should be generally useful in uncovering underlying mechanisms.

Disclosure of Potential Conflicts of Interest

A. Califano is cofounder and chief scientific advisor at DarwinHealth, Inc. and has ownership interest (including patents) in the same. J.M. Irish is cofounder and board member of Cytobank, Inc. and received research support from Incyte Corp. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: A.R. Udyavar, D.J. Wooten, L. Estrada, P.P. Massion, V. Quaranta

Development of methodology: A.R. Udyavar, D.J. Wooten, P.P. Massion

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A.R. Udyavar, M. Hoeksema
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.R. Udyavar, D.J. Wooten, M. Bansal, S. Schnell, J.M. Irish, P.P. Massion
Writing, review, and/or revision of the manuscript: A.R. Udyavar, D.J. Wooten, A. Califano, S. Schnell, J.M. Irish, P.P. Massion, V. Quaranta
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A.R. Udyavar, P.P. Massion
Study supervision: A.R. Udyavar, A. Califano, P.P. Massion, V. Quaranta

Acknowledgments

We acknowledge the Vanderbilt Flow Cytometry shared resource lab for the help and assistance in training for the use of flow cytometry instruments. We thank Dr. Charles Rudin (Memorial Sloan Kettering Center, New York, NY) and Dr. Julien Sage (Stanford University, Stanford, CA) for providing us with patient-derived xenograft tumors.

Grant Support

This work was supported by NIH/NCIU54 CA113007-09, Integrative Cancer Biology Program (ICBP, V. Quaranta), Grant 1101CX000242 from the Department of Veterans Affairs, CA90949 from the NCI SPORE program (P.P. Massion), NIH/NCIR00 CA143231 (J.M. Irish), NIHT32 CA009592 (J. Chen), National Center for Advancing Translational Sciences U11-TR000445-06 G. Bernard, and the Vanderbilt-Ingram Cancer Center (P30 CA68485 to J. Pietenpol).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 14, 2016; revised October 7, 2016; accepted November 7, 2016; published OnlineFirst December 8, 2016.

References

- American Cancer Society. Cancer facts and figures 2016. Atlanta, GA: American Cancer Society; 2016.
- Rudin CM, Ismaila N, Hann CL, Malhotra N, Movsas B, Norris K, et al. Treatment of Small-Cell Lung Cancer: American Society of Clinical Oncology Endorsement of the American College of Chest Physicians Guideline. *J Clin Oncol Off J Am Soc Clin Oncol* 2015;33:4106–11.
- Fischer B, Marinov M, Arcaro A. Targeting receptor tyrosine kinase signaling in small cell lung cancer (SCLC): what have we learned so far? *Cancer Treat Rev* 2007;33:391–406.
- Hann CL, Rudin CM. Fast, hungry and unstable: finding the Achilles' heel of small-cell lung cancer. *Trends Mol Med* 2007;13:150–7.
- Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* 2012;12:323–34.
- Carney DN, Gazdar AF, Bepler G, Guccion JG, Marangos PJ, Moody TW, et al. Establishment and identification of small cell lung cancer cell lines having classic and variant features. *Cancer Res* 1985;45:2913–23.
- Sutherland KD, Proost N, Brouns I, Adriaensen D, Song J-Y, Berns A. Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. *Cancer Cell* 2011;19:754–64.
- Calbo J, van Montfort E, Proost N, van Drunen E, Beverloo HB, Meuwissen R, et al. A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. *Cancer Cell* 2011;19:244–56.
- George J, Lim JS, Jang SJ, Cun Y, Ozretić L, Kong G, et al. Comprehensive genomic profiles of small cell lung cancer. *Nature* 2015;524:47–53.
- Almendro V, Cheng Y-K, Randles A, Itzkovitz S, Marusyk A, Ametller E, et al. Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. *Cell Rep* 2014;6:514–27.
- Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature* 2013;501:328–37.
- Waddington CH. The strategy of the genes. London: Allen & Unwin; 1957.
- Huang S. The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology? *BioEssays News Rev Mol Cell Dev Biol* 2012;34:149–57.
- Kauffman S. Differentiation of malignant to benign cells. *J Theor Biol* 1971;31:429–51.
- Huang S, Ernberg I, Kauffman S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin Cell Dev Biol* 2009;20:869–76.
- Li S, Zhu X, Liu B, Wang G, Ao P. Endogenous molecular network reveals two mechanisms of heterogeneity within gastric cancer. *Oncotarget* 2015;6:13607–27.
- Szedlak A, Paternostro G, Piermarocchi C. Control of asymmetric Hopfield networks and application to cancer attractors. *PLoS One* 2014;9:e105842.
- Albert I, Thakar J, Li S, Zhang R, Albert R. Boolean network simulations for life scientists. *Source Code Biol Med* 2008;3:16.
- Wynn ML, Consul N, Merajver SD, Schnell S. Logic-based models in systems biology: a predictive and parameter-free network analysis method. *Integr Biol Quant Biosci Nano Macro* 2012;4:1323–37.
- Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;483:603–7.
- Gautier L, Cope L, Bolstad BM, Irizarry RA. *affy*—analysis of Affymetrix GeneChip data at the probe level. *Bioinforma Oxf Engl* 2004;20:307–15.
- Miller JA, Cai C, Langfelder P, Geschwind DH, Kurian SM, Salomon DR, et al. Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics* 2011;12:322.
- Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinforma Oxf Engl* 2010;26:1572–3.
- Clinical Lung Cancer Genome Project (CLCGP), Network Genomic Medicine (NGM). A genomics-based classification of human lung tumors. *Sci Transl Med* 2013;5:209ra153.
- GEO Accession viewer [Internet]. [cited 2016 May 17]. Available from: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6044>
- Udyavar AR, Hoeksema MD, Clark JE, Zou Y, Tang Z, Li Z, et al. Co-expression network analysis identifies Spleen Tyrosine Kinase (SYK) as a candidate oncogenic driver in a subset of small-cell lung cancer. *BMC Syst Biol* 2013;7Suppl 5:S1.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
- Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLOS One* 2010;5:e13984.
- Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 2006;7 Suppl 1:S7.
- Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, et al. The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 2010;463:318–25.
- Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinforma Oxf Engl* 2010;26:2438–44.
- ENCODE Project Consortium and others. The ENCODE (ENCyclopedia Of DNA Elements) project. *Science* 2004;306:636–40.
- Matys V, Fricke E, Gösling E, Haubrock M, Hehl R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 2003;31:374–8.
- Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2013;D142-7. doi: 10.1093/nar/gkt997.
- Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013;14:128.
- Jourquin J, Duncan D, Shi Z, Zhang B. GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC Genomics* 2012;13Suppl 8:S20.

Udyavar et al.

37. Willadsen K, Wiles J. Robustness and state-space structure of Boolean gene regulatory models. *J Theor Biol* 2007;249:749–65.
38. Derrida B, Pomeau Y. Random networks of automata: a simple annealed approximation. *EPL Europhys Lett* 1986;1:45.
39. Irish JM, Doxie DB. High-dimensional single-cell cancer biology. *Curr Top Microbiol Immunol* 2014;377:1–21.
40. Kotecha N, Krutzik PO, Irish JM. Web-based analysis and publication of flow cytometry experiments. *Curr Protoc Cytom Editor Board J Paul Robinson Manag Ed Al* 2010;Chapter 10:Unit10.17.
41. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
42. Downen JM, Fan ZP, Hnisz D, Ren C, Abraham BJ, Zhang LN, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 2014;159:374–87.
43. Gilbert SF. *Developmental biology*. Sunderland, MA: Sinauer; 2013.
44. Jolly MK, Boareto M, Huang B, Jia D, Lu M, Ben-Jacob E, et al. Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Front Oncol* 2015;5:155.
45. Xue J, Zhu Y, Sun Z, Ji R, Zhang X, Xu W, et al. Tumorigenic hybrids between mesenchymal stem cells and gastric cancer cells enhanced cancer proliferation, migration and stemness. *BMC Cancer* 2015;15:793.
46. Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer* 2013;108:479–85.
47. Travis WD. Update on small cell carcinoma and its differentiation from squamous cell carcinoma and other non-small cell carcinomas. *Mod Pathol* 2012;25:S18–30.
48. Lecharpentier A, Vielh P, Perez-Moreno P, Planchard D, Soria JC, Farace F. Detection of circulating tumour cells with a hybrid (epithelial/mesenchymal) phenotype in patients with metastatic non-small cell lung cancer. *Br J Cancer* 2011;105:1338–41.
49. Schliekelman MJ, Taguchi A, Zhu J, Dai X, Rodriguez J, Celiktas M, et al. Molecular portraits of epithelial, mesenchymal, and hybrid states in lung adenocarcinoma and their relevance to survival. *Cancer Res* 2015;75:1789–800.
50. Steinway SN, Zaňudo JGT, Michel PJ, Feith DJ, Loughran TP, Albert R. Combinatorial interventions inhibit TGF β -driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *Npj Syst Biol Appl* 2015;1:15014.

Correction: Novel Hybrid Phenotype Revealed in Small Cell Lung Cancer by a Transcription Factor Network Model That Can Explain Tumor Heterogeneity

Akshata R. Udyavar, David J. Wooten, Megan Hoeksema, Mukesh Bansal, Andrea Califano, Lourdes Estrada, Santiago Schnell, Jonathan M. Irish, Pierre P. Massion, and Vito Quaranta



In the original version of this article (1), Fig. 1 was incorrectly replaced with Fig. 2. The error has been corrected in the latest online HTML and PDF versions of the article. The publisher regrets this error.

Reference

1. Udyavar AR, Wooten DJ, Hoeksema M, Bansal M, Califano A, Estrada L, et al. Novel hybrid phenotype revealed in small cell lung cancer by a transcription factor network model that can explain tumor heterogeneity. *Cancer Res* 2017;77:1063-74.

Published online March 1, 2019.

doi: 10.1158/0008-5472.CAN-19-0069

©2019 American Association for Cancer Research.

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Novel Hybrid Phenotype Revealed in Small Cell Lung Cancer by a Transcription Factor Network Model That Can Explain Tumor Heterogeneity

Akshata R. Udyavar, David J. Wooten, Megan Hoeksema, et al.

Cancer Res 2017;77:1063-1074. Published OnlineFirst February 28, 2017.

Updated version Access the most recent version of this article at:
doi:[10.1158/0008-5472.CAN-16-1467](https://doi.org/10.1158/0008-5472.CAN-16-1467)

Supplementary Material Access the most recent supplemental material at:
<http://cancerres.aacrjournals.org/content/suppl/2016/12/03/0008-5472.CAN-16-1467.DC1>

Cited articles This article cites 44 articles, 5 of which you can access for free at:
<http://cancerres.aacrjournals.org/content/77/5/1063.full#ref-list-1>

Citing articles This article has been cited by 6 HighWire-hosted articles. Access the articles at:
<http://cancerres.aacrjournals.org/content/77/5/1063.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://cancerres.aacrjournals.org/content/77/5/1063>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.