QUANTIFYING PHENOTYPIC HETEROGENEITY IN SMALL-CELL LUNG CANCER: IMPLICATIONS FOR SUBTYPE CLASSIFICATION AND TREATMENT

By

Akshata Ramrao Udyavar

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical and Physical Biology

May 2015

Nashville, Tennessee

Approved by:

Dr. Vito Quaranta

Dr. Pierre Massion

Dr. Alissa Weaver

Dr. Jonathan Irish

Dr. David Cortez

Dr. Bing Zhang

ORIGINAL PUBLICATIONS

1. Ocak S, Yamashita H, **Udyavar AR**, Miller AN, Gonzalez AL, Zou Y, Jiang A, Yi Y, Shyr Y, Estrada L, Quaranta V, Massion PP. DNA copy number aberrations in small-cell lung cancer reveal activation of the focal adhesion pathway. Oncogene. 2010 Dec 2; 29(48):6331-42.
2. **Udyavar AR**, Hoeksema MD, Clark JE, Zou Y, Tang Z, Li Z, Li M, Chen H, Statnikov A, Shyr Y, Liebler DC, Field J, Eisenberg R, Estrada L, Massion PP, Quaranta V. 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; 7 Suppl 5:S1.
3. Franco OE, Tyson DR, **Udyavar AR**, Konvinse K, Estrada L, Hayward SW, Quaranta V. Cooperative interactions between heterogeneous tumor cells promotes cancer progression (*under review*).
4. **Udyavar AR**, Wooten DJ, Hoeksema M, Bansal M, Califano A, Estrada L, Irish J, Massion PP, Quaranta V. Distinct transcriptional programs drive phenotypic heterogeneity in Small-Cell Lung Cancer (*manuscript in preparation*).

PATENTS

1. **Udyavar A**, Estrada L, Quaranta V, Massion PP. SYK as a target in Small-cell lung cancer (SCLC). U.S. Patent 61/829,789, filed May 31, 2013.

To my beloved family and the most supportive husband one could ever ask for, Aniket Borwankar:

Thank so much for being my pillars of support throughout this journey.

ACKNOWLEDGEMENTS

There are many people who have made significant contributions to getting me where I am today. First and foremost, I would like to sincerely thank Dr. Vito Quaranta for being an amazing mentor, and so patient and kind with me. He has been an amazing role model and father figure to me when I joined Vanderbilt in 2008 as a research assistant in his lab, in only my second year in the United States. He took a chance with hiring me then and in 2010, he supported my transition from a RA to a graduate student in his lab.

I would like to thank Dr. Pierre Massion, who has been a wonderful PhD co-mentor and a clinical mentor while pursuing the diploma in the Certificate Program in Molecular Medicine (CPMM). It is his undying passion to diagnose and treat lung disease and cancer patients, that has truly made me appreciate and respect the work I do even more. Shadowing him in the clinic was an eye-opening experience for me and made me realize the gaps that exist in bench science and translational medicine. Working with his lab has motivated me to make my research more translationally relevant. Thank you for incorporating me in your lab and making invaluable patient samples available to my research.

My deepest gratitude to Dr. Lourdes Estrada who was the first person I met at Vanderbilt and has been an amazing role model, mentor and a friend to me. Thanks so much for always having your office door open for me, to listen to my problems, calming me down and providing solutions. I have always appreciated you being my go-to person and sounding board in the lab.

It has been an amazing journey for me in both the Quaranta and Massion laboratories. The senior people in the lab including Dr. Darren Tyson, Shawn Garbett, Dr. Katherine Jameson, Dr. Mohamed Hassanein, Dr. Jun Qian and Dr. Jonathan Lehman have always been there for advice, support and prompt help whenever I needed it, be it learning a programming language or a new experimental protocol, or just bouncing off crazy ideas. I have learnt so much from each and everyone of you. I have made some life-long friends and had fun interacting with each and everyone in both labs. I would especially like to thank Megan Hoeksema, Bradford Harris, Keisha Hardeman and David Wooten for making significant contributions to this thesis. David Wooten and Dr. Leonard Harris have been awesome resources for learning concepts about attractor landscapes and network modeling. Jing Hao and Yong Zou have been great pleasures to work with and been very helpful throughout this journey.

I have also learnt a great deal of statistics and R programming from the Biostaticians in the group - Dr. Ming Li, Pencheng Lu, Dr. Heidi Chen and Dr. Yu Shyr. Dr. Josh Fessell has been a wonderful resource of exciting ideas and help with the metabolism aspects of the project. I would like to thank Dr. Hal Moses and Dr. Yu Shyr for making time for me and being such great resources for career advice.

I would like to also thank the awe-inspiring and talented faculty members that are part of my thesis committee. Dr. Bing Zhang and Dr. Alissa Weaver have been my first sources of inspiration to start learning bioinformatics and studying network biology in 2010 as I watched them conceptualize the Program Project Grant (PPG) on systems biology of lung cancer together with our lab members. Dr. Jonathan Irish and his laboratory members (Deon Doxie, Nalin Leelathian, Hannah Polinsky, Kirsten Diggins,

Cara Woogsland) have just been an absolute pleasure to work with and introduced me to the idea of high-throughput quantitate flow cytometry. Dr. Dave Cortez has inspired me to think about the big picture ideas of my project and helped in bouncing off ideas.

This work would not have come this far if it had not been for a great group of collaborators established during this journey. Dr. Steve Horvath at UCLA, the developer of co-expression network analysis package (WGCNA), has been instrumental in me learning R programming and application of co-expression network analysis in my thesis. Dr. Julien Sage (Stanford University), Dr. Kwok-Kin Wong (Dana Farber Cancer Institute) and Dr. Charles Rudin (MSKCC) have been impressive resources for patient and mouse tumor samples. I would like to thank Dr. Kimberly Stegmaier and her lab members for their advice and sharing the SYK constructs and inhibitors with me. I would also like to thank Dr. Dana Pe'er and lab members (Columbia University) for advice and help on flow cytometry data analysis tools. I would sincerely like to thank Dr. Mukesh Bansal and Dr. Andrea Califano (Columbia University) for the assistance and advice on using the network analysis tools developed by their lab.

My journey wouldn't have been as fun-filled and exciting if it hadn't been for my life outside of the lab. I have been extremely lucky to gain a family of friends during my time in Nashville. These folks have been the coolest friends and always been there for me in my happy and tough times. Finally, I would like to thank my mother Sangeeta and father Ramrao who have been my role models in life, and have taught me to be perseverant, persistent and a perfectionist. Their pain-staking efforts and undying support has helped me come to the United States and be where I am today. My brothers Amit and Ameya, brother-in-law Kedar and sisters-in-laws Usha, Akshata and

Sital, have always been there for me and encouraged me throughout my life. I would also like to thank my mother-in-law Sulochana and father-in-law Dilip for supporting my career goals throughout the past 8 years, and for being here every year to help me in whatever way they could. Last but the most important, my first friend in the United States and soul mate, Aniket Borwankar, has been my rock ever since we got married. I have been very fortunate to have him around me, always giving his 100% to our relationship and the well-being of our family. The sacrifices he has made in this journey together have been simply astounding. He has taught me the importance of relationships and unconditional love, and that it is OK to take breaks from work once in a while and enjoy the simple joys of life. I could not have done any of this without his presence in my life. And of course, I could not have made it through long nights of work or writing without the unconditional love and joyful company I received every day by my beloved furry babies Buddy and Casper.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AC	Atypical carcinoids
ADC	Adenocarcinoma
ADS	Adenosquamous
ALK	Anaplastic lymphoma receptor tyrosine kinase
ASCL1	Achaete-scute complex homolog 1
AT2	Alveolar type II epithelial cell
BAC	Bronchioalveolar carcinoma
BASCs	Bronchioalveolar stem cells
BC	Basaloid carcinoma
BCR	B-cell antigen receptor
BRAF	v-raf murine sarcoma viral oncogene homolog B
CA	Carcinoids
CCCP	Carbonyl cyanide m-chlorophenylhydrazone
CCLE	Cancer cell line encyclopedia
CCND1	Cyclin D1
CD	Cluster of differentiation
CDF	Cumulative distribution function
CGA	Chromogranin A
CGRP	Calcitonin Gene-related peptides
CI	Confidence interval
CT	Computed tomography
CTCs	Circulating tumor cells
CYB5A	Cytochrome B5
DDC	Dopa decarboxylase
DDR2	Discoidin domain receptor tyrosine kinase 2
DIPNECH	Diffuse idiopathic pulmonary neuroendocrine cell hyperplasia
DPI	Data processing inequality
EGFR	Epidermal growth factor receptor
EML4	Echinoderm microtubule associated protein like 4

EMT	Epithelial-to-mesenchymal transition
ERBB2	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2
ES	Embryonic stem
ES	Extensive stage
FDR	False discovery rate
FFPE	Formalin-fixed paraffin embedded
FGFR1	Fibroblast growth factor receptor 1
FPKM	Fragments per kilobase of exon per million
GEMM	Genetically engineered mouse models
GO	Gene Ontology
GRP	Gastrin releasing peptide
GS	Gene significance
GSEA	Gene-set enrichment analysis
HDAC	Histone deacetylase
H&E	Hematoxylin-eosin
HGNT	High-grade neuroendocrine tumors
ITAM	immunoreceptor tyrosine based activation motif
IVL	Involvulin
KEAP1	kelch-like ECH-associated protein 1
KEGG	Kyoto Encyclopedia of Genes and Genomes
kME	Intramodular connectivity
KRAS	Kirsten rat sarcoma viral oncogene homolog
KRT14	Cytokeratin 14
KRT7	Keratin 7
LC	Large-Cell Carcinoma
LCNEC	Large-Cell Neuroendocrine Carcinoma
LGALS7	Galectin-7
LS	Limited stage
MEs	Module eigengenes
MET	Mesenchymal-to-epithelial transition
c-MET	Hepatocyte Growth Factor Receptor, MET photo-oncogene

ML	Mesenchymal-like
MRI	Magnetic resonance imaging
MYCL	v-myc avian myelocytomatosis viral oncogene lung carcinoma derived homolog
NCAM1	Neural cell adhesion molecule 1, also known as CD56
NE	Neuroendocrine
NEBs	Neuroepithelial bodies
NET	Neuroendocrine tumors
NF1	neurofibromin 1
NSCLC	Non-small-cell lung cancer
NSE/ENO2	Neuron-specific enolase
PTEN	Phosphatase and tensin homolog
PDX	Patient derived xenograft
PET	Positive emission tomography
PI3K	Phosphatidylinositol 3-kinase
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha
PNECs	Pulmonary neuroendocrine cells
RET	Ret proto-oncogene
ROS1	c-ros oncogene 1 , receptor tyrosine kinase
RTK	Receptor tyrosine kinase
SCC/SQ	Squamous cell carcinoma
SCLC	Small-cell lung cancer
SFTPB	Surfactant protein B
SFTPC	Surfactant protein C
SFTPD	Surfactant protein D
SHH	Sonic hedgehog
SOX2	SRY (sex determining region Y)-box 2
SRS	Somatostatin receptor scintigraphy
SSHN	SCLC specific hub network
STK11	Serine threonine kinase 11
SYK	Spleen tyrosine kinase

TC	Typical carcinoids
TMAs	Tissue microarrays
TNM	Tumor-Node-Metastasis
TT	Targeted therapies
TTF-1	Thyroid transcription factor 1
VPA	Valproic acid
WGCNA	Weighted gene co-expression network analysis
WHO	World Health Organization

CHAPTER I

INTRODUCTION

Lung cancer: An ensemble of multiple phenotypes

Lung cancer is the second most common cancer type (14% of all cancers), and is the leading cause of cancer-related deaths (26-29%) in both men and women each year across the United States (Siegel et al., 2012). Cigarette smoking remains the leading risk factor for this disease, in addition to asbestos exposure. Lung cancer is diagnosed usually via physical examination, a chest X-ray or computed tomography (CT) scan. The Tumor-Node-Metastasis (TNM) staging is used for lung cancer prognosis, where higher stage presents worse prognosis. Early diagnosis is key in lung cancer as early stage (IA) patient survival rates are close to 70% (Wardwell et al., 2005; Hassanein et al., 2012). Recent lung cancer screening trials suggest that early diagnosis of high-risk patients result in a significant decrease in mortality rates (Humphrey et al., 2013). Once a lung nodule is detected, a tissue biopsy is used for pathology-based assessment of the cancer type and stage. Although 96% of the diagnosed lung nodules are generally benign, additional noninvasive methods such as diagnostic and prognostic biomarkers in high-risk individuals are warranted for detection of malignant lung tumors (Hassanein et al., 2012; Massion et al., 2014).

Histology-based classification of lung cancer

Based on histology, lung cancer is broadly classified as non-small-cell lung cancer (NSCLC), small-cell lung cancer (SCLC) or Carcinoids (CA). NSCLC have worse survival rates than carcinoids but better than neuroendocrine tumors (NET) (Figure 1.1B, details in next section).

Non-small cell lung cancer (NSCLC)

NSCLC, 80-85% of all lung cancers, is further classified as Adenocarcinoma (ADC), Squamous Cell Carcinoma (SCC/SQ) and Large-Cell carcinoma (LC) (including Large-Cell Neuroendocrine Carcinoma LCNEC) (Figure 1.1A). 5-year survival rate of NSCLC is less than 15%, primarily attributed to late diagnosis in advanced stages (Clinical Lung Cancer Genome Project (CLCGP)Network Genomic Medicine NGM, 2013). Surgical resection via lobectomy and adjuvant platinum-based chemotherapy is the standard of care in early-stage (stages I-II) NSCLC patients. Advanced stage NSCLC (III-IV) get treated with combination of chemotherapy and radiation (Zarogoulidis et al., 2013).

ADC, the most common lung cancer (40%), present as peripherally located lesions with ductal/glandular morphology and prominent nucleoli (Figure 1.1C). It is strongly associated with smoking, although a significant number of non-smokers also develop ADC. These tumors show strong expression of thyroid transcription factor 1 (TTF-1), surfactant proteins, napsin A and keratin 7 (KRT7) (Table 1.1).

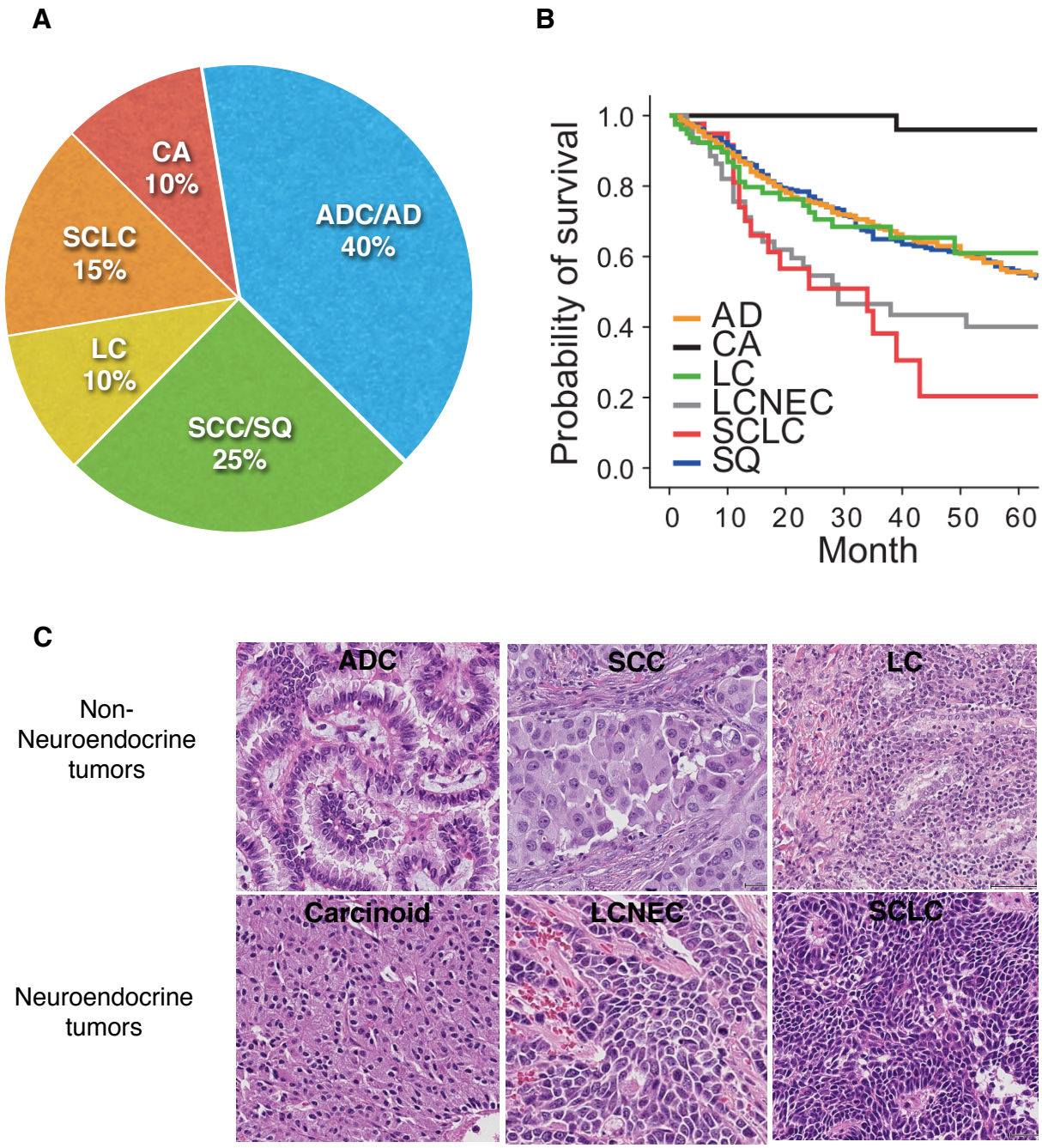


Figure 1.1: Histology based classification of lung cancer.

(A) Prevalence of the different subtypes of lung cancer (percentages obtained from cancer.gov).

(B) 5-year survival statistics of lung cancer subtypes (Clinical Lung Cancer Genome Project (CLCGP)Network Genomic Medicine NGM, 2013).

(C) Hematoxylin-eosin (H&E) sections of various lung cancer subtypes. Top panel shows non-neuroendocrine and bottom panel comprises of neuroendocrine lung tumors.

Tumour type	Negative	Positive
SCLC	p63	NCAM, CK7, TTF1
LCNEC	p63	NCAM, CK7, TTF1±
Adeno	CK20*, NEmarkers**	CK7*, SurfApoA/B, NapsinA, TTF1
SCC	TTF1, NEmarkers±	CK5/6, p63
LC	p63±, CK5/6	TTF1±, CK7, CK14±
*CK20 can be positive and CK7 negative in centrally located and mucinous AC. **Neuroendocrine markers: most useful NCAM (CD56), less ChromograninA, Synaptophysin.		

Table 1.1: Histology-based classification of lung cancer.

Markers used for identification of various histological subtypes of lung cancer (Popper et al., 2011).

ADC is a highly heterogenous disease, simply by histology itself, ranging from well differentiated, non-invasive, non-mucinous bronchioalveolar carcinoma (BAC) with lepidic pattern and ~93% 5-year survival, to poorly differentiated micropapillary and mucinous ADC with poor 5-year survival (38-39%) (Brambilla et al., 2001; Chen et al., 2014, Russell et al., 2011).

SCC, the second most common lung cancer (25%), is more strongly associated with smoking than ADC. These tumors exhibit a squamous differentiation pattern ('flat scale-like') with keratinization, express cytokeratin 5 and 6, p63 and SOX2 and are TTF-1 negative (Figure 1.1C, Table 1.1). SCC are subclassified as (1) basaloid (BC), (2) papillary, (3) clear cell and (4) small-cell variants (Perez-Moreno et al., 2012; Chen et al., 2014).

LC, 10% of all lung cancers, are composed of large-cells and diagnosed as having low expression of ADC, SCC or SCLC markers (Figure 1.1C, Table 1.1). It is very difficult to diagnose LC as there are no specific biomarkers for this disease. Futhermore,ADC/SCC protein biomarkers (Table 1.1) tend to reassign LC (~60%) into ADC or SCC subtypes. Also, depending on the number of sections investigated, higher is the probability of ADC/SCC differentiation patterns (Barbareschi et al., 2011; Popper, 2011).

Adenosquamous (ADS) tumors, a rare lung cancer (0.4-4%), are a mixture of ADC and SCC in a minimum of 10% of the tumor field, classified as a separate entity by 2004 World Health Organization (WHO) criteria. These tumors express both above mentioned ADC and SCC markers (Table 1.1). Stage I disease has worse survival than either ADC and SCC (Filosso et al., 2011).

Due to high level of variability in marker expression and histological features, there is a dire need for better ways to classify lung cancer subtypes.

Neuroendocrine tumors (NET)

Neuroendocrine tumors (NET) are a distinct histological subgroup of lung cancer comprising of low-grade typical carcinoids (TC), intermediate-grade atypical carcinoids (AC), high-grade NET - large-cell neuroendocrine cancer (LCNEC), and small-cell lung cancer (SCLC) (Figure 1.1C). These comprise of 2%, 0.2%, 3% and 15-20% of all lung cancers respectively (Table 1.2). These tumors present with common features such as (1) neuroendocrine morphology defined by organoid nesting, palisades, rosettes, or a trabecular pattern of cellular organization, (2) faint cytoplasm and nucleoli, 'salt-and-pepper' pattern of chromatin, and (3) expression of neuropeptides and neuronal markers such as chromogranin A, synaptophysin, neuron-specific enolase (NSE/ENO2), gastrin releasing peptide (GRP) and NCAM1/CD56 (Gustafsson et al., 2008; Rekhtman et al., 2010; Travis et al., 2010). Most NET also express TTF-1, an ADC-specific marker (Sturm et al., 2002) (Table 1.1). In general, ADC and SCC are considered non-neuroendocrine tumors due to the lack of expression of neuroendocrine markers. 95-100% of SCLC and LCNEC patients have a smoking history, while 50% carcinoid patients are smokers. Stage I-II TC and AC have a 5-year survival of 75-98% and 75-100% respectively, while early stage SCLC and LCNEC still have a poor prognosis (Table 1.2) (Asamura et al.,2006).

Features	Typical carcinoid TC	Atypical carcinoid AC	Large-cell neuroendocrine carcinoma LCNEC	Small-cell lung cancer SCLC
Incidence *	2%	0.2%	3%	15-20%
5-year survival rate**,#	92-100%	61-88%	15-57%	5%
Percent smokers #	33%	64%	98%	97%
Average age at diagnosis #	40-50	50-60	68	50-70
Mitosis**	<2 mitosis per 2 mm^2	2-10 mitosis per 2 mm^2	≥ 11 mitosis (High rate of mitosis)	60-80 mitosis (High rate of mitosis)
Proliferation rate**	≤ 5%	5-20%	50-100%	80-100%
Necrosis ++,**	No necrosis	Punctate necrosis	High levels of necrosis	High levels of necrosis
Tumor grade ++	Low	Intermediate	High	High
Morphology ++	Well differentiated	Well differentiated	Poorly differentiated	Poorly differentiated
Lymph node metastasis #	4-14%	35-64%	40%	90%
Distant metastasis#	1.5% Liver, bone	10%	65%	60-70%
3p locus alterations ++	< 25%	> 50%	> 50%	> 75%
p53 alteration++	4%	29%	80%	75%
Telomerase activity ++	< 10%	-	~ 90%	~ 90%
Symptoms ++,**	Shortness of breath, hemoptysis, cough, pneumonia	Shortness of breath, hemoptysis, cough, pneumonia	Chest pain, Shortness of breath, hemoptysis, cough, pneumonia, weight loss	Fatigue, cough, pain, Shortness of breath, hemoptysis, weight loss
Paraneoplastic syndrome ++,**	Cushing's syndrome; Carcinoid syndrome	Cushing's syndrome; Carcinoid syndrome	-	Cushing's syndrome; autoimmune encephalomyelitis

Table 1.2: Histology-based classification of neuroendocrine tumors of the lung (NET).

[+] - Carcinoid morphology :
Neuroendocrine morphology :
* - from (Rekhtman et al., 2010)
** - from (Travis et al., 2010)
[++] - from (Gustafsson et al., 2008)
[#] - from (Swarts et al., 2012)

Table 1.2: Histology-based classification of neuroendocrine tumors of the lung (NET).

Table and facts adapted from (Gustafsson et al., 2008; Travis et al., 2010; Rekhtman et al., 2010; Swarts et al., 2012)

Pulmonary neuroendocrine cells (PNECs), distributed as single cells throughout the lung or form clusters (termed neuroepithelial bodies, NEBs), are postulated to be the putative cell of origin of NET. Several conditions such as fibrosis, interstitial lung disease, and obliterative broncheolitis, induce hyper-proliferation of these PNECs leading to a condition termed Diffuse idiopathic pulmonary neuroendocrine cell hyperplasia (DIPNECH), a rare precancerous state. These generate lesions < 5mm in size are referred to as 'carcinoid tumorlets'. Carcinoids are typical stage I tumors, well differentiated with a low mitotic index and lesion size >5mm, exhibiting benign behavior and have the best survival amongst all lung cancers (Figure 1.1B, Table 1.2). TC and AC differ in presence of necrosis, mitotic index and tumor metastasis, whereby TC, being a localized mass, are mostly treated via surgical resection (lobectomy) (Table 1.2). Being resistant to chemotherapy and radiation, AC with lymph node metastasis are difficult to treat and primarily managed with surgery. Unfortunately, no treatment regimen other than surgery exists for these patients (Gustafsson et al., 2008; Travis et al., 2010; Rekhtman et al., 2010; Gallego et al., 2012).

SCLC and large-cell neuroendocrine carcinoma (LCNEC), a variant of LC, are considered less differentiated, high-grade neuroendocrine tumors of the lung (HGNT) due to their aggressive nature and have the worst prognosis among all lung cancers (Figure 1.1B and C, Table 1.2) (Gustafsson et al., 2008). These tumors present with high mitotic index and loss of p53 and Rb, major nodal and distant metastatic spread, substantial necrosis. Unlike carcinoids, surgical approaches are not an option for patients with these these tumors. LCNEC exhibits a similar neuroendocrine phenotype as SCLC, but can be distinguished by its NSCLC features such as visible nucleoli,

larger size and cytoplasm, and clumpy chromatin (Rekhtman et al., 2010). High-grade NET are treated with chemotherapy and thoracic radiation (in early stage disease), although treatment regimens for LCNEC are not well studied (Travis et al., 2010). SCLC patients present with brain metastasis, hence prophylactic cranial irradiation is recommended for delaying recurrence (Rekhtman et al., 2010).

Once a pulmonary nodule is identified, computed tomography (CT) of the chest are useful to diagnose NET. Analog-based imaging such as somatostatin receptor scintigraphy (SRS) and 68Ga-DOTA-TOC positive emission tomography (PET) are useful to distinguish NET from other lung cancers since majority of NET express somatostatin receptors. For high-grade NETs, bone and brain magnetic resonance imaging (MRI) are performed for detection of distant metastasis (Gustafsson et al., 2008).

Based on histology of fine-needle biopsies or surgical resections, it is difficult to distinguish SCLC from carcinoids, and SCLC from LCNEC, especially due to crush artifact of SCLC and difficulty in assessment of mitotic index/necrosis in small sections of the tumor. Furthermore, 10-20% NSCLC tumors also exhibit expression of neuroendocrine markers, which complicates the distinction of NET from NSCLC (Gustafsson et al., 2008; Travis et al., 2010). The presented evidence further highlights the need for continuing efforts towards identification of specific biomarkers and signatures to guide disease classification and management.

Genomics-based classification of lung cancer: advent of personalized medicine

For most of the lung cancer patients, both NSCLC and SCLC, the standard of care is either surgery for early-stage disease and chemotherapy coupled with thoracic radiation in late-stage disease. The response rates for chemotherapy (cisplatin-etoposide) in NSCLC patients are less than 30% leading to a median 5-year survival of less than 10% (Demedts et al., 2009).

The concept of 'oncogene addiction' was coined a decade ago by Bernard Weinstein (Weinstein, 2002), where tumor maintenance is dependent on the mutated or overexpressed form of the oncogene such that targeted inhibition of this 'addiction' resulted in diminished proliferation and tumor growth. Several examples of oncogene addiction were reported in genetically engineered mouse models of leukemia, lymphoma, and melanoma (Felsher et al., 1999; Tran et al., 2008; Chin et al., 1999). This led to a sudden increase in studies in lung cancer for the identification of targetable oncogenes.

With the advent of novel genetic screening strategies including mutation screens, exome and RNA sequencing, significant strides have been made in genomics-based classification of lung cancer. Global and subtype-specific alterations can be found in patients (Figure 1.2A), validating the observed histological differential characteristics (Figure 1.1C). All lung cancers carried chromosomal gains in 5p, and p53 was the most commonly mutated gene (53%) (Clinical Lung Cancer Genome Project (CLCGP)Network Genomic Medicine NGM, 2013).

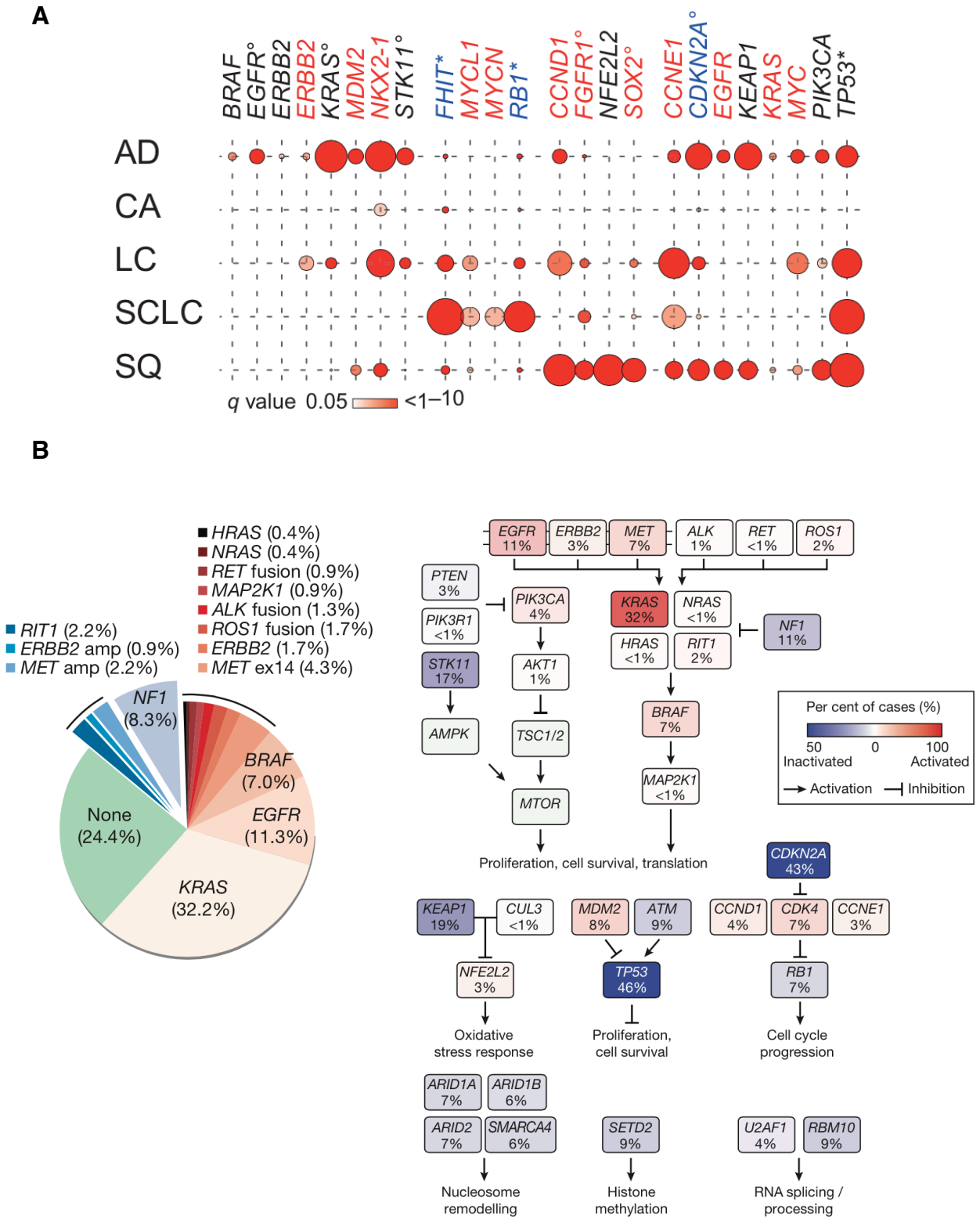


Figure 1.2: Genomics based classification of lung cancer.

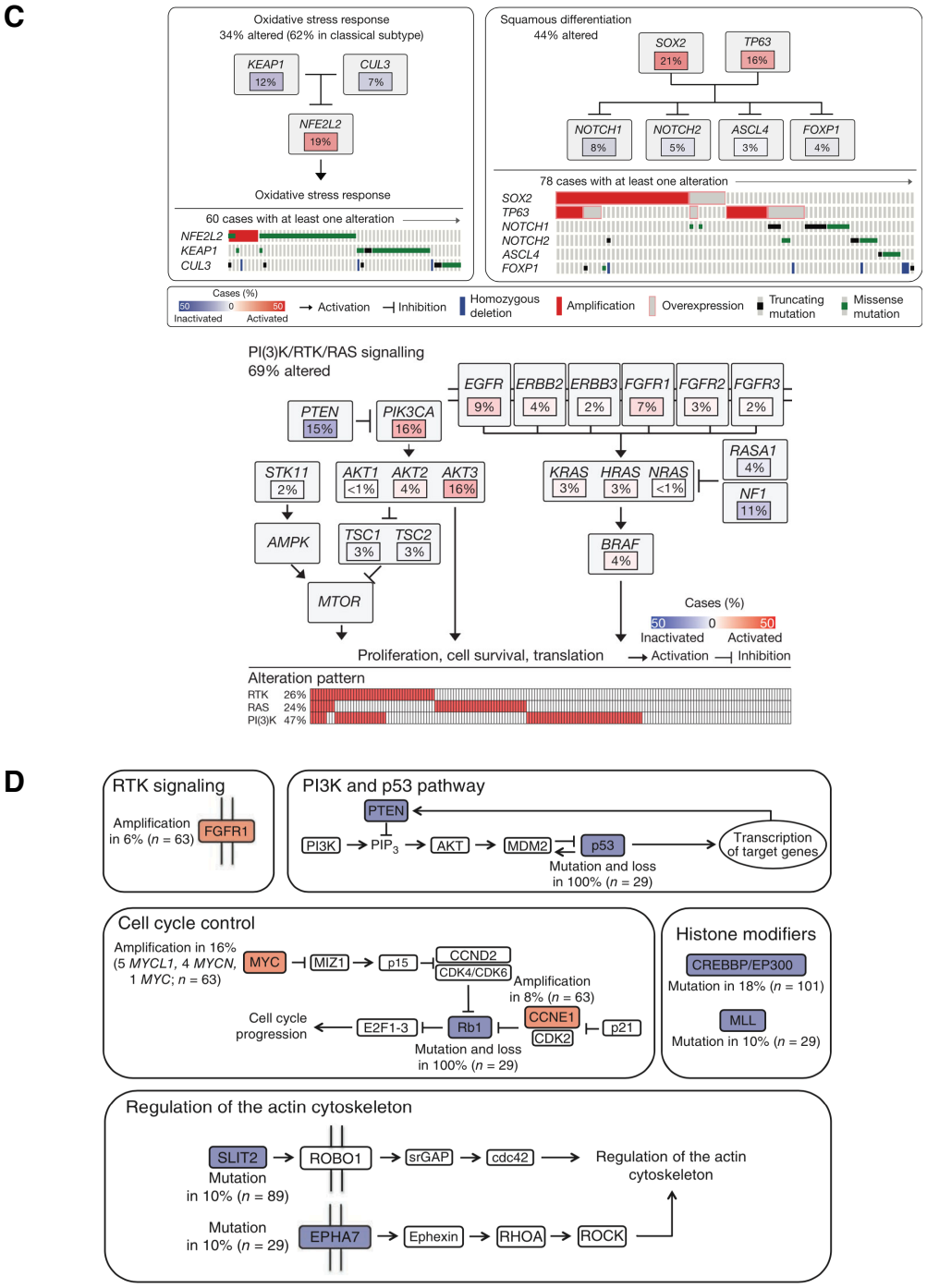


Figure 1.2: Genomics based classification of lung cancer.

(A) Frequency of chromosomal amplifications (red), losses (blue) and mutations (black) in tumor suppressors and addicting oncogenes in histological subsets of lung cancer (Clinical Lung Cancer Genome Project (CLCGP)Network Genomic Medicine NGM, 2013). Significantly altered targetable pathways in lung (B) ADC (Network, 2014), (C) SCC (Hammerman et al., 2012) and (D) SCLC (Peifer et al., 2012).

On the other hand, specific alterations in kinases such as EGFR, KRAS, ERBB2, ALK, STK11 were observed only in ADC, but not in other lung cancers (Figure 1.2A) (Clinical Lung Cancer Genome Project (CLCGP)Network Genomic Medicine NGM, 2013). Interestingly, ADC patients exhibited non-overlapping alterations (activating mutations, amplifications) in KRAS (32.2 %), EGFR (11.3%), BRAF (7%), ROS1/ALK/ RET (4-5%), c-MET (2.2%), RIT1 (2.2%) and ERBB2 (0.9%) (Figure 1.2B, left panel). Newly discovered alterations included ERBB2 and loss-of-function mutations in KEAP1 and NF1 in previously oncogene-negative patients. Together with exome sequencing and reverse-phase protein array (RPPA), 76%, 25%, 63%, 64%, 22% and 49% exhibited RTK/RAS/RAF, PI3K-mTOR, p53, cell cycle regulation, oxidative stress, and chromatin-RNA splicing factors pathway alterations respectively have been discovered (Figure 1.2B, right panel) (Network, 2014).

Although majority of late stage ADC are still treated with chemotherapy as standard-of-care, patients with EGFR and EML4-ALK fusions have been successfully treated with EGFR (gefitinib, erlotinib)(Pao et al., 2004) and EML4-ALK (crizotinib) (Gaughan and Costa, 2011; Pao et al., 2010) specific inhibitors as first-line treatment, resulting in significant improvements in survival. However, currently, 24.4% of ADC still lack a targetable oncogenic driver (Network, 2014); KRAS subset still remains non-targetable (Riely et al., 2009), and most other newly discovered mutations such as BRAF, ERBB2, and ROS1/RET fusions, are still in the experimental stage (Chen et al., 2014; Pao et al., 2010). Furthermore, the response to targeted therapies is short-lived and several resistance mechanisms to EGFR TKI (Chong and Jänne, 2013; Engelman

et al., 2007; Guix et al., 2008) and EML-ALK TKI (Lovly and Pao, 2012) have been discovered.

SCC has lacked any targetable oncogenic drivers, until recently when specific alterations have been discovered including (1) amplifications in several chromosomal regions including SOX2, MYCL1, EGFR, FGFR1 and CCND1, (2) mutations in NFE2L2, DDR2 and FGFR3, and (3) deletions in PTEN, FOXP1, and NF1 (Clinical Lung Cancer Genome Project (CLCGP)Network Genomic Medicine NGM, 2013) (Figure 1.2A). Others have also reported 3q amplification in SCC, which contains SOX2, a lineage-specific oncogene for SCC, P63, a well-accepted marker for SCC (Table 1.1), and PIK3CA, a key component of the PI3kinase pathway (Bass et al., 2009; Yamamoto et al., 2008). Serine-threonine kinase STK11/LKB1 inactivation has also been found in 19% of SCC patients, and when combined with PTEN deletion gives rise to murine SCC lung tumors (Ji et al., 2007; Xu et al., 2014). In addition, oxidative stress response (34%), squamous differentiation (44%) and PI3K-AKT (47%, regulated by LKB1, PTEN) pathways were altered in a significant subset of SCC patients, suggesting the potential importance of targeting these pathways (Figure 1.2C)(Hammerman et al., 2012).

However, these findings have not yet translated into clinical trials catered to SCC patients. FGFR1 inhibitors (PD173074) have proven to be efficacious in NSCLC cell lines with FGFR1 amplifications (Weiss et al., 2010). TKIs developed for ADC have been futile in this disease since it lacks ADC-specific alterations (Figure 1.2A). Mutation rate in SCC is much higher than that of ADC (Scagliotti et al., 2013). EGFR mutations distinct from those of ADC are seen in SCC patients, that are potentially targetable (Hammerman et al., 2012).

SCLC lacks all of these above ADC- and SCC-specific alterations but harbors specific deletions of p53, FHIT and Rb1 (Figure 1.2A). Oncogenes such as MYCL1, MYCN, Cyclin E1 (CCNE1) and FGFR1 can also be amplified. Separately, Peifer et.al described 100% of SCLC patients exhibited mutation and loss of both tumor suppressors p53 and Rb, underscoring the relevance of genetically engineered mouse models (GEMM) of SCLC with Cre-mediated knockout of these 2 tumor suppressors that also exhibit amplifications in MYCL1 and MYCN (Park et al., 2011a; Peifer et al., 2012; Sutherland et al., 2011). In addition, loss-of-function mutations in PTEN, CREBBP/EP300, MLL, FHIT EPHA7 and SLIT2 have been identified (Figure 1.2D) (Peifer et al., 2012). Interestingly, p53-Rb-PTEN Cre-knockout GEMM suggests PTEN loss to be a late event in tumor evolution, thereby promoting progression (McFadden et al., 2014). Twenty-six percent (26%) of SCLC patients display SOX2 amplification (Rudin et al., 2012).

In terms of targetable RTK altered pathways, FGFR1 amplification was the only significantly altered receptor tyrosine kinase (RTK) in ~6% SCLC patients (Peifer et al., 2012; Schultheis et al., 2013). Unfortunately, the majority of SCLC patients remain unsuitable for currently available targeted therapies due to lack of single targetable oncogenic drivers.

LCNEC shares several genomic features with SCLC such as mutations in p53, Rb1 and EP300, including the worst survival amongst all lung cancers (Clinical Lung Cancer Genome Project (CLCGP)Network Genomic Medicine NGM, 2013). Interestingly, carcinoids carried no significant oncogenic alterations, while LC shared common alterations with both NSCLC (ADC, SCC) and SCLC, exhibiting high level of

diversity (Figure 1.2A) (Clinical Lung Cancer Genome Project (CLCGP)Network Genomic Medicine NGM, 2013).

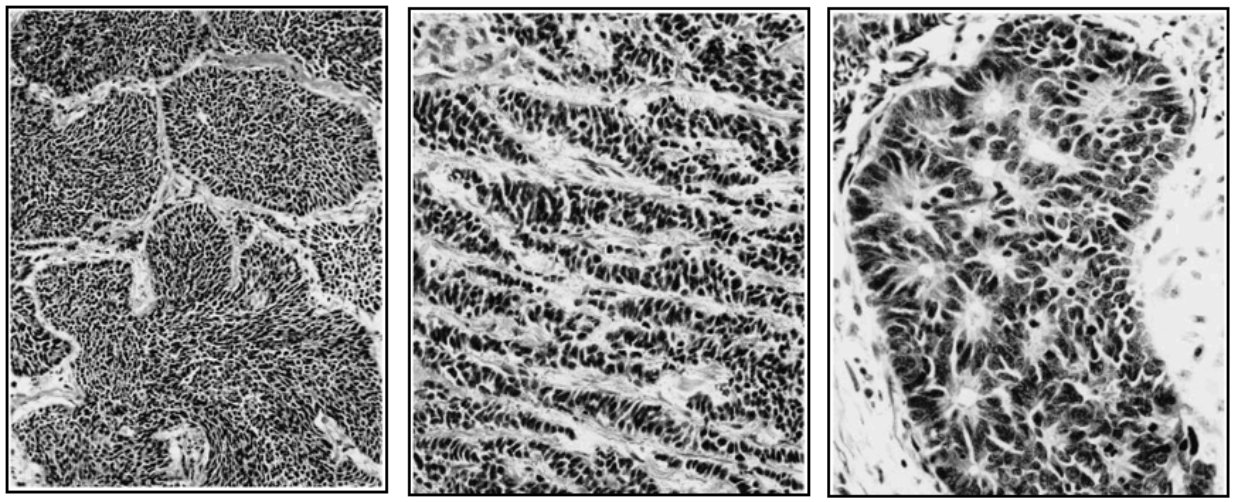
In summary, lung cancer subtypes (especially well-differentiated early stage tumors) can be distinguished via histology, biomarkers via immunohistochemistry and genetic alterations. Some tumor types have significantly benefited from targeted therapies. However, most lung tumors present as late-stage, invasive and poorly differentiated disease, and mixed phenotypes (SCLC-NSCLC or ADC-SCC) are difficult to distinguish from one another. Thus, better biomarkers and targeted therapies are urgently needed in effective personalized medicine in lung cancer.

Untamed biology of Small-cell lung cancer (SCLC)

SCLC comprise 15-20% of all lung cancers and its is found almost exclusively in heavy smokers. Patients with SCLC face an unfortunate prognosis of less that 5% 5-year survival rate (Figure 1.1B, Table 1.1). SCLC is classified by histology from NSCLC by its small nuclei, crush artifact and palisade-fusiform cellular patterns, and neuroendocrine marker expression (Figure 1.3). It is staged as limited stage (LS) and extensive stage (ES) disease. LS-SCLC (Stage I-III of TNM, 30% of SCLC) is confined to one-side of the lung (i.e. hemithorax), lacks malignant pleural effusion and can be covered in one radiation therapy port. ES-SCLC (Stage IV of TNM, 70% of SCLC) does not satisfy these specifications and exhibits highly metastatic behavior (Rosti et al., 2006).

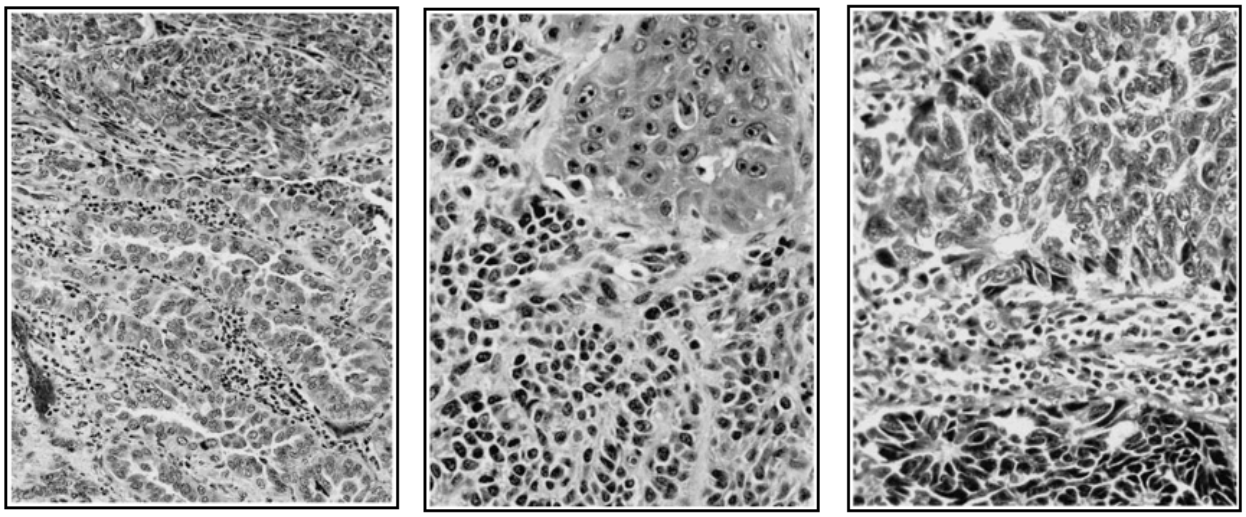
Prior to 1999 World Health Organization (WHO) guidelines, SCLC was classified as pure, mixed (with LC or LCNEC) and combined (with ADC or SCC). Currently, it is

Pure SCLC



Palisade pattern | Trabecular pattern | Rosette pattern

Combined SCLC



SCLC + ADC | SCLC + SCC | SCLC + LC

Figure 1.3: Histology based diagnosis of SCLC.

Modified from (Nicholson et al., 2002)**.**

(A) Pure SCLC - patterns observed in tumor sections.
(B) Combined SCLC - SCLC mixed with NSCLC phenotypes.

classified as either pure SCLC and combined SCLC containing atleast 10% NSCLC cells that includes LC, ADC and SCC (Figure 1.3) (Hirsch et al., 1988; Nicholson et al., 2002). Other than histopathology, to date no genetic biomarkers specific to SCLC or molecular subtypes have been identified in SCLC as it is the case for NSCLC (Figure 1.2B). In addition, other than tumor stage, no other histopathological features are predictive of prognosis. Since most SCLC are usually diagnosed based on a small biopsy or fine-needle biopsy, it may be difficult to accurately quantify amount and extent of mitosis and necrosis, the key features that distinguish between the various NETs. Furthermore, due to the mixed histologies of SCLC, one of the major issues in the clinic today is distinction of SCLC from both NSCLC and low-grade NET (Travis, 2012).

Gene-expression profiling based approaches have led to successful subtype predictions in leukemia (Yeoh et al., 2002), breast (Lehmann et al., 2011; Sørlie et al., 2001), prostate (Lapointe et al., 2004) and colorectal (Sadanandam et al., 2013) cancers that have given rise to subtype-specific therapeutic strategies. These unsupervised analyses have been also recently applied to identify molecular subtypes in NSCLC such as ADC and SCC that match the mutational subtypes (Hammerman et al., 2012; Network, 2014; Wilkerson et al., 2010; 2012). Gene expression profiling studies for SCLC, based on differential expression, have yielded highly independent signatures with minimal overlap of 5-10 genes. Hence, no reproducible signature exists for SCLC in the clinic.

Cell of origin of lung neuroendocrine tumors: PNECs

Neuroendocrine tumors of the lung are thought to arise from pulmonary neuroendocrine cells (PNECs) that are sparsely distributed throughout the lung and also form small clusters termed neuroepithelial bodies (NEBs). PNECs (and NEBs) are specialized epithelial cells that are part of the parasympathetic nervous system, which sense oxygen via NADPH oxidase (NOX) receptors, secrete neuropeptides, namely, bombesin, serotonin (5-hydroxytryptophan, 5-HT), Calcitonin Gene-related peptides (CGRP), Dopa decarboxylase (DDC), and express neuronal markers such as Chromogranin A (CGA), Neuron specific enolase (NSE/ENO2), Neural cell adhesion molecule (NCAM1/CD56). These neuropeptides are secreted in response to hypoxia or oxidative stress, and act in an autocrine and paracrine fashion to regulate breathing reflexes (Buttigieg et al., 2012; Gustafsson et al., 2008).

Two hallmark tumor suppressors p53 and Rb1 are mutated and lost in 75-90% and close to 100% SCLC patients respectively. Other genetic alterations that have been described in several reports are amplifications in the Myc family of oncogenes (c-Myc, N-Myc, L-Myc) and loss of PTEN (Byers and Rudin, 2014; Peifer et al., 2012; Rudin et al., 2012). Mouse models designed by Cre-mediated knockout of p53 and Rb, generate tumors (in about 273 days) that resemble SCLC in terms of histology, neuroendocrine marker expression, L-Myc and N-MYC amplification and propensity for liver metastasis (Sutherland et al., 2011). In addition, p53/Rb/p130 and p53/Rb/PTEN triple knockout mouse models have yielded SCLC tumors that have reduced latency (McFadden et al., 2014; Park et al., 2011a).

Current treatment options for SCLC

Standard of care treatment

Effective translational discoveries in SCLC have been slow and disappointing partly explained by: (1) our lack of understanding of early events due to unavailability of early-stage disease tissue; (2) drug toxicity-related issues; (3) paucity of tissue for genomic studies since these patients are diagnosed by fine-needle biopsies and rarely surgically resected; and (4) no re-biopsy protocols in place for insight into the rapid recurrence developed with standard-of-care treatment. Chemotherapy remains the standard of care for most, if not all, SCLC patients.

Strikingly, median survival of untreated disease is 11-14 weeks (LS-SCLC) and 5-7 weeks (ES-SCLC) respectively (WM et al., 2006). Standard of care first-line treatment for LS-SCLC consists of combination chemotherapy (cisplatin-etoposide) and thoracic/cranial irradiation for prevention of local and distant metastases, while ES-SCLC are treated with combination chemotherapy alone due to widespread systemic disease. Despite initial overall response rates are >80% in LS-SCLC and >60% in ES-SCLC, the median survival of these patients is 14-20 and 8-13 months respectively (Figure 1.4A) (Demedts et al., 2009; Rosti et al., 2006).

The initial response rates are dampened by early recurrence and widespread metastasis leading to an overall survival rate of 5-10% and 2-5% in LS-SCLC and ES-SCLC, respectively (Rosti et al., 2006). Second-line treatment with temozolomide (TMZ) has recently shown 38% complete/partial responses in SCLC patients with brain metastatic lesions (Pietanza et al., 2012).

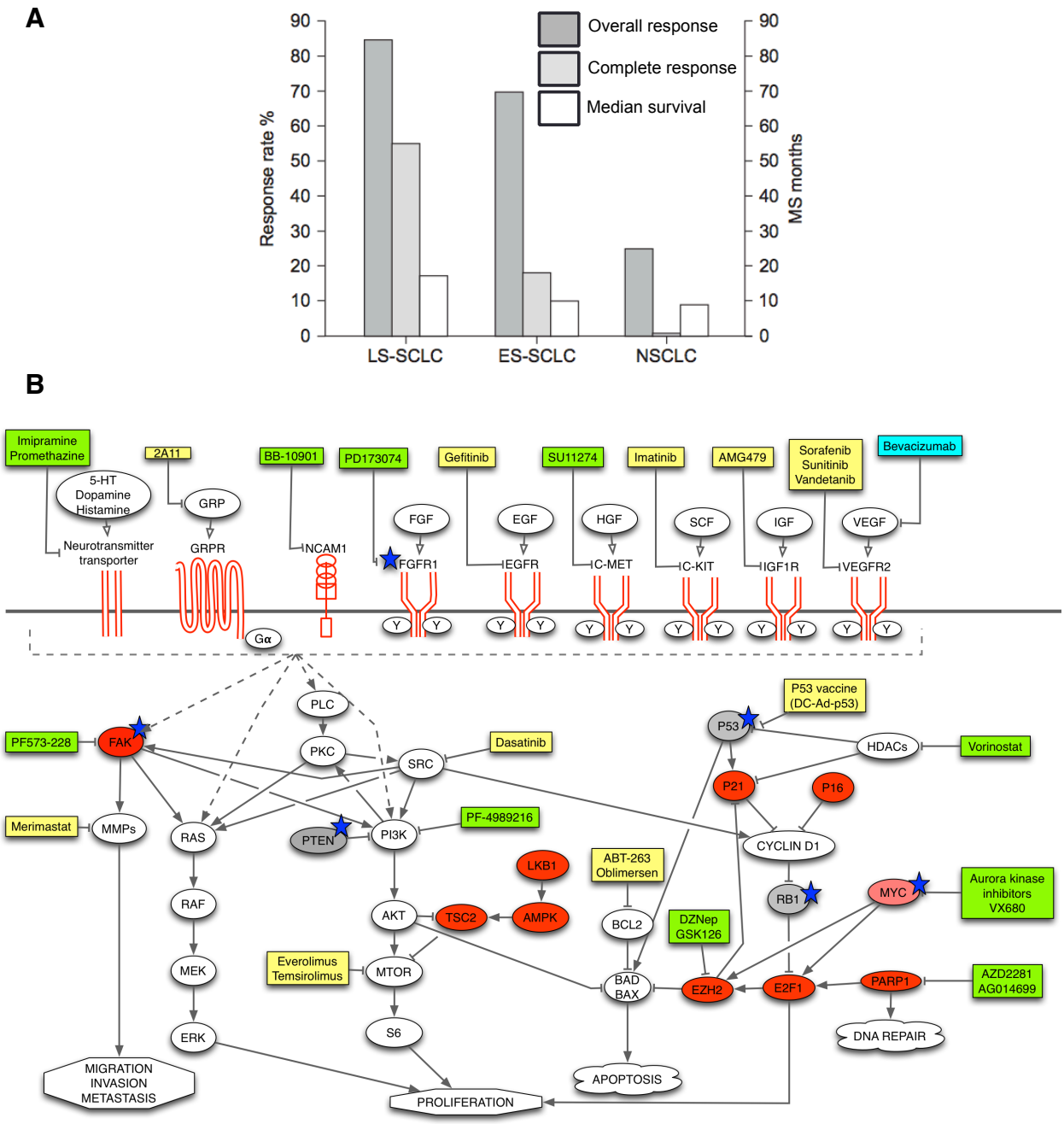


Figure 1.4: Current status of therapeutic strategies for SCLC patients.

Figure 1.4: Current status of therapeutic strategies for SCLC patients.

(A) Response rates and median survival of SCLC patients to standard of care combination chemotherapy (cisplatin-etoposide) (Demedts et al., 2009).

(B) Overview of signaling pathway alterations studied in SCLC. Overexpressed targets (including receptors) and tumor suppressors (loss in SCLC) are indicated in red and grey respectively. Amplified/ mutated targets are indicated by the blue star. Targeted therapies failed in clinical trials are shown in yellow, while new therapies entering clinical trials are indicated in green. Promising targets in Phase II trials are indicated in turquoise (Byers et al., 2012; Jahchan et al., 2013; Ocak et al., 2010; Sato et al., 2013; Sos et al., 2012; Zhang and He, 2013).

Targeted therapies in SCLC : failures and hope for the future

Several attempts have been made to target cell signaling and angiogenesis pathways in SCLC, but these have failed in clinical trials as summarized in Figure 1.4B. Inhibitors of invasion/metastasis and angiogenesis (MMPs, VEGFR) therapeutics have failed in SCLC due to high toxic profiles and no added survival advantage. However, Bevacuzimab, an anti-VEGF antibody, shows some promise in phase II clinical trials. Despite overexpression of c-Kit in in 28-73% SCLC, clinical trials conducted to date with Imatinib, a c-Kit/BCR-Abl/PDGFR inhibitor, have not shown therapeutic benefits. EGFR mutations, found in NSCLC, are not found in SCLC. Dasatinib, a Src-family kinase inhibitor and mTOR inhibitors have also failed to show any improvements in survival. Gastrin-releasing peptide (GRP), a neuropeptide and its receptor GRPR are overexpressed in a large percentage SCLC. However, targeting GRP via 2A11 only showed response in 1 out of 13 patients (Hann and Rudin, 2007; Hohla and Schally, 2010; Zhang and He, 2013).

Being associated with heavy smoking, SCLC has an extremely high mutation rate and genomic instability. Thus alterations are found in many passenger oncogenes, but very few are targetable driver oncogenes. Thus far, FGFR1 amplifications have been found in 6% SCLC, and some FGFR1-amplified SCLC cell lines are responsive to FGFR1-targeted inhibition (Figure 1.4B). Amplification of MYC-family transcription factors, seen in 20% SCLC patients, sensitizes SCLC cell lines to Aurora kinase inhibitors (Sos et al., 2012). 8% and 16.8% of SCLC cell lines and patients, respectively, show mutations in PIK3CA, which predicts response to PIK3CA inhibitor PF-4989216, and induces apoptosis *in vitro* and *in vivo* (Walls et al., 2014).

c-MET is expressed in a subset of SCLC cell lines triggering ongoing clinical trails with its inhibitors (Hann and Rudin, 2007; Wang et al., 2011b; Zhang and He, 2013). Proteomic profiling using reverse-phase protein arrays (RPPA) demonstrated that DNA-damage response, LKB1-AMPK, and cell-cycle pathways are elevated in SCLC identifying novel therapeutic targets such as PARP1, Chk1 and EZH2 (Figure 1.4B) (Byers et al., 2012). SCLC cell lines expressing PARP1 and low PI3K pathway activity were shown sensitive to drug/siRNA-mediated inhibition *in vitro* and *in vivo* (Cardnell et al., 2013). EZH2 also seems to be promising target in SCLC (Byers et al., 2012; Hubaux et al., 2013). Drug repositioning approaches using SCLC gene expression data have led to identification of pathway-based class of drugs such as tricyclic antidepressants (imipramine) that were validated in cell line and patient-derived xenografts (PDX) models (Figure 1.4B)(Jahchan et al., 2013).

'Transcriptional addiction' targeting is slowly entering SCLC field with the discovery of Sonic hedgehog (Shh), a developmental pathway important in lung development and injury response (Velcheti and Govindan, 2007). Shh was found to be constitutively active in SCLC tumor initiation and pathway inhibition prevented tumor initiation and delayed resistance to cisplatin-etoposide (Park et al., 2011b). ASCL1, a lineage survival TF important for PNEC development and differentiation, was found to be overexpressed in ~72% of SCLC, 13-85% in other NET and a subset of NSCLC with neuroendocrine features (NSCLC-NE, 10%) (Jiang et al., 2003). ASCL1 was co-expressed with stem-cell markers CD133 and ALDH1A1 in subpopulations in SCLC PDX, and its knockdown in SCLC cell lines inhibited colony formation in soft agar (Jiang et al., 2009). Also, inhibition of BCL2, a downstream target of ASCL1, was effective in

NSCLC cell lines with NET features (Augustyn et al., 2014). SOX2 amplification was identified in a subset of SCLC patients (27%) that might be potentially sensitive to SOX2 inhibition (Rudin et al., 2012). Recently, CDK7 inhibitor THZ1, identified through a drug screen in mSCLC cell lines, suppressed growth of SCLC cell lines *in vitro* and *in vivo*, via downregulation of key neuroendocrine transcription factors like SOX2, SOX4, INSM1, NeuroD1 and ASCL1 (Christensen et al., 2014).

These findings highlight the role of transcriptional networks in SCLC, and suggest that the TF itself or inhibition of druggable targets downstream of the TF could offer viable options for treatment of SCLC.

Inter-tumor heterogeneity

Inter-tumor heterogeneity is defined by patient-to-patient variability driven by different intrinsic factors such as genetic and racial predispositions, sex and extrinsic factors like environment and exposure to carcinogens (asbestos, radiation, UV, smoking). This type of heterogeneity is typically described as genetic alteration- or gene expression-based classification as observed in breast cancer (Her2, luminal ER/PR+, triple negative/basal-like), leukemia and NSCLC (Burrell et al., 2013; Network, 2014; Sørlie et al., 2001; Yeoh et al., 2002). These subtypes are the basis to identify subsets of patients that would be more likely to respond to a particular targeted therapy (Pao et al., 2010), an approach that has made significant impact on patient outcomes. Genetic heterogeneity is only recently coming into light with SCLC, as subsets of patients with distinct genomic alterations (FGFR1, MYC, PTEN, SOX2) are being identified (Pietanza and Ladanyi, 2012).

Intra-tumor heterogeneity

Intra-tumor heterogeneity can arise from both genetic and non-genetic mechanisms that ultimately constitute the cellular makeup or "phenotype", a.k.a the hallmarks of cancer (Almendro et al., 2013).

Genetic heterogeneity

Genetic variability arises from high genomic instability in cancer leading to clonal diversity explained by a "linear evolution" where genetic events occur in a single dominant clone, or "branched evolution" where several clones arise and survive simultaneously (Swanton, 2012). Clonal evolution is spatially (due to differences in angiogenesis, hypoxia, microenvironment in primary and metastatic niches) and temporally regulated, evident by multiple sampling of the tumor at various locations within the primary and metastatic sites (Almendro et al., 2013; Marusyk et al., 2012), recently described in NSCLC patients (de Bruin et al., 2014; Zhang et al., 2014b) and in mouse models of SCLC (McFadden et al., 2014). Even in the same location, neighboring cells exhibit genetic alterations of distinct receptors (EGFR, PDGFRa) in glioma (Szerlip et al., 2012).

Non-genetic heterogeneity : the theory of attractor landscapes

Non-genetic heterogeneity can arise from dynamics in the gene regulatory networks (GRN) even within genetically similar cancer cells. Genetically distinct cell types such as KRas mutant or EGFR mutant generate the same ADC phenotype,

suggesting that they might posses the same underlying GRN. A GRN enforces (or at least can enforce) the existence of multiple distinct states which are stable over time. In dynamical systems theoretical framework, these states can be thought of as existing at the bottom of some "basin", such that all cells within the basin have a natural tendency to fall toward the bottom, stable state. This stable GRN state is referred to as an 'attractor' a.k.a phenotype of a cell that is at the bottom of the basin, which is heritable over multiple cell generations/divisions (Figure 1.5A). For a 2-node (A and B genes) GRN, 4 possible network states can be achieved, the 2 stable attractors can be found where A>>B or B<<A (Figure 1.5A). Thus, for an n-node GRN, there are 2^n possible states. However, in many cases, most of these states will be transient - only visited as the system moves toward a small subset of states which are stable. In the specific example shown in Figure 1.5A, which has 2 nodes, and therefore 2^n = 2^2 = 4 possible states, only 2 of which are stable.

Non-genetic heterogeneity can be deterministic or stochastic depending on the 'noise' in the system. In normal cells which are genetically identical and the microenvironment is relatively homogeneous, the gene regulatory networks (GRN) driving a certain phenotype are stable due to low 'noise' (Figure 1.5B). In cancer, 'noise' is driven by genetic clonal variation and disorganized microenvironmental conditions (hypoxia, aberrant angiogenesis, stroma and tumor-infiltrating immune cells), which promotes a noisy GRN leading to a more plastic phenotype where state transitions are more easily achieved (Figure 1.5B)(Marusyk et al., 2012).

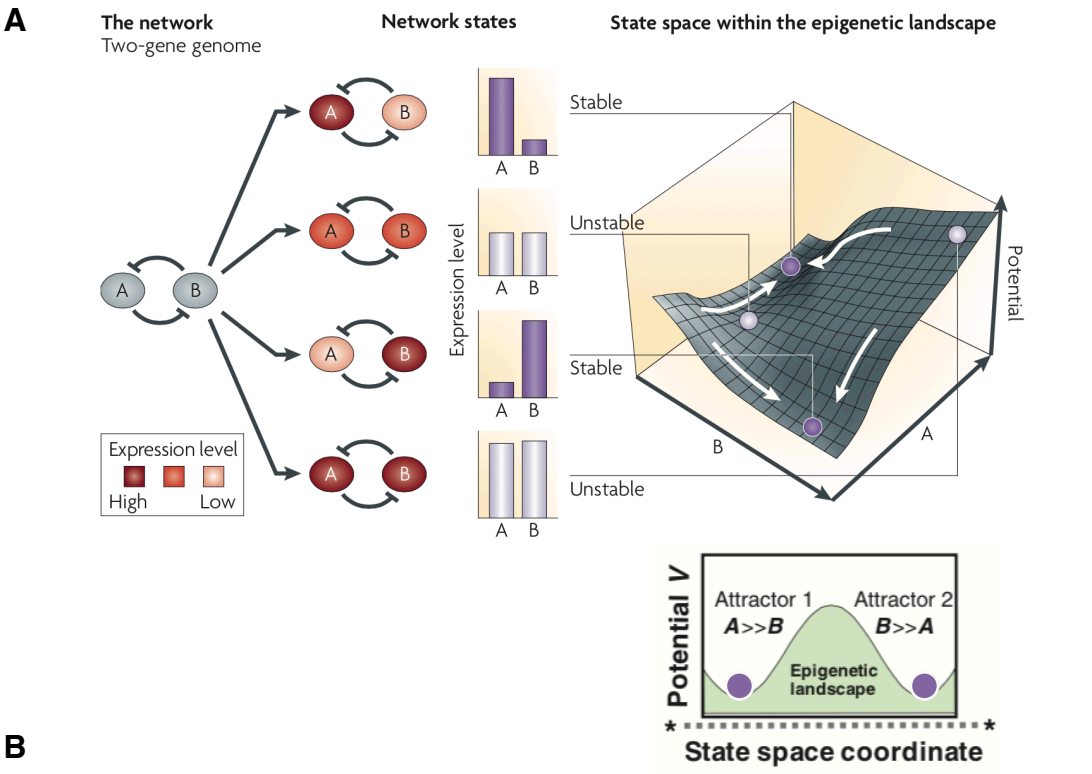


Figure 1.5: Attractor landscape view of non-genetic heterogeneity in normal tissue and cancer.

2-node gene regulatory network giving rise to stable attractor states 1 and 2. Modified from (Brock et al., 2009; Huang, 2009).

Noise in GRN and state transitions in normal and tumor. From (Marusyk et al., 2012).

Normal development of the human body starts from 1 single cell (1 attractor state) and differentiates into several distinct cell types over time simply by fluctuations in the network without acquisition of any mutations, and may involve chromatin-mediated (epigenetic) genome-level transitions. This phenomena of normal cellular differentiation is explained by Waddington's epigenetic landscape as a hierarchical process where an embryonic stem (ES) cell rolls down a mountain full of valleys signifying different degrees of differentiated cell phenotypes (Brock et al., 2009). However, these stable differentiated states, which are endpoints in this hierarchical process, a.k.a basins (e.g. fibroblasts) can be kicked out of their stable state or 'basin' into a dedifferentiated induced pluripotent stem cell-like state (iPS) (Wernig et al., 2007) or transdifferentiated state (e.g. neurons) (Chanda et al., 2014) by manipulation of the GRN via expression (or 'activation') of a cocktail of transcription factors, underlining the plasticity of these states given a strong push. These changes in stable phenotypes such as dedifferentiation or transdifferentiation are referred to as state transitions in an attractor landscape.

Phenotypic state transitions in lung injury and cancer

Normal lung is extremely heterogenous with several distinct differentiated cell types which upon injury can enter cell cycle and be reprogrammed (Kotton and Morrisey, 2014) (Figure 1.6A). Upon lung injury, normal PNECs, cell of origin of murine SCLC, can transdifferentiate and repopulate the other cell types such as Clara and ciliated epithelia (Figure 1.6A). Interestingly PNEC hyperplasia is also induced during lung development and via smoking (Lommel, 2001; Song et al., 2012). Similarly,

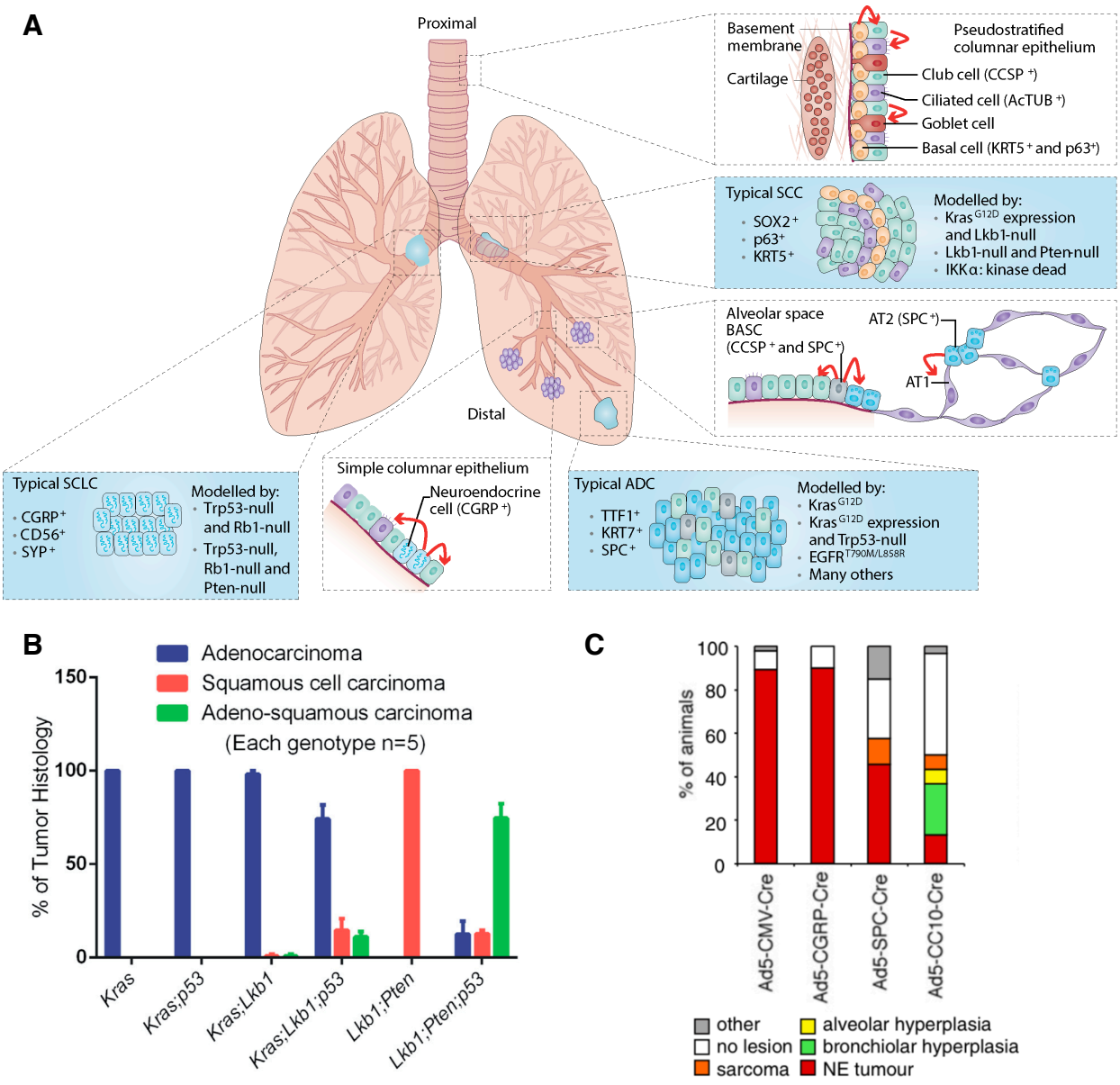


Figure 1.6: Heterogeneity of phenotypic states in normal and neoplastic lung.

(A) Normal lung consists of various cell types in different regions of the lung. Types of lung cancer are indicated along with the associated genetic murine models. Phenotypic state transitions (transdifferentiation) in normal lung or during lung injury are indicated by the red arrows. Modified from (Chen et al., 2014) and (Kotton and Morrisey, 2014). (B) shows distinct tumor histologies arising from pairing various combinations of KRas mutation, and loss of LKB1, p53 and PTEN. Each genetic lesion generated distinct types of NSCLC tumors. Adapted from (Xu et al., 2014). (C) shows percentage of mice that developed different types of tumors, specifically NE tumors resembling SCLC by Cre-mediated knockout of p53 and Rb in different cell types. CGRP, SPC and CC10 promoters target PNECs, AT2 and Clara cells respectively. Adapted from (Sutherland et al., 2011).

mature, functional alveolar type 2 (AT2) cells (cell of origin of lung ADC), in the lung maintain a 'bifunctional stem-cell' state (both differentiated and regenerative function) whereby they transdifferentiate into alveolar type 1 (AT1) over time, especially after AT1 injury (Desai et al., 2014) (Figure 1.6A). In addition, $p63^{+}KRT5^{+}$ basal cells can differentiate into ciliated, AT1 and AT2 cells upon injury (Kotton and Morrisey, 2014; Rock et al., 2011). Similar pathways and mechanisms participate during lung injury and potentially in cancer development (Chen et al., 2014), which might explain the heterogeneous phenotypic states.

Oncogenic $KRas^{G12D}$ induction combined with targeted deletion of either LKB1 (mutated in SCC patients), PTEN or TGFBR2 specifically in basal cells generated tumors exhibiting both SCC and ADC tumors, while LKB1-PTEN combined knockout yielded SCC (Figure 1.6B) (Chen et al., 2014; Malkoski et al., 2013; Xu et al., 2014). In addition, a unique subpopulation of stem-like alveolar epithelial cells (similar to AT2), upon $KRas^{G12D}$ induction, could generate ADC, SCC or sarcomatoid carcinoma, depending on the cross-talk between SOX2 and TGFbeta pathway components (Ischenko et al., 2014). Also, SOX2, a lineage-specific oncogene for SCC, overexpression in ADC cell lines promotes squamous-like differentiation including expression of SCC-specific markers such as P63 and KRT6 (Bass et al., 2009). Cre-mediated knockout of p53 and Rb in Clara cells (CC10-Cre), AT2 cells (SPC-Cre) and PNEC (CGRP-Cre), each yielded NET tumors resembling SCLC with ~20% , 50% and 90% incidence rate. This strongly suggests that each cell type is capable of producing SCLC, although PNECs were the most efficient (Figure 1.6C) (Sutherland et al., 2011). Interestingly, SCLC mouse models generated by Cre-mediated knockout of p53 and Rb

in PNECs, exhibit phenotypic heterogeneity where each mSCLC tumor contains neuroendocrine (NE) and non-NE states, which co-operatively, not singly, promote liver metastasis. Mutant KRas induction converts NE to non-NE differentiated states (Calbo et al., 2011). However, KRas mutation is very rare in human SCLC, so if these non-NE cells are in fact ADC or a novel state in human SCLC remains to be seen. Combined SCLC with NSCLC (30% of SCLC tumors) and NSCLC with NE features (10% of NSCLC), are seen commonly in the clinic (Nicholson et al., 2002; Travis, 2010), indicating that the above described phenomena might be taking place in human tumors leading to increased intra-tumor heterogeneity.

This phenotypic heterogeneity and plasticity is evident in other cancers as well. In breast cancer, luminal, basal and stem-like states (sorted via flow cytometry based on surface markers EpCAM, CD24 and CD44) can stochastically transition between states to re-equilibrate to the original heterogeneous population over time (Gupta et al., 2011). Glioblastoma cells can be reprogrammed into cancer stem cell-like phenotype can be achieved via induction of a set of 4 TFs (Suvà et al., 2014).

Tools to study tumor heterogeneity : a systems-level approach

The extent of intra-tumor/phenotypic heterogeneity is typically measured by genetic alterations, gene expression, surface markers, proliferation, signaling and transcription factors. Gene expression, RNAseq whole genome/exome sequencing and proteomics have been used to delineate inter-tumor heterogeneity and identify classifications in various cancers including lung cancer. With evolving technologies, there is an increasing need for data analysis pipelines to decipher relationships between

gene expression and genetic alterations. Consensus clustering and unsupervised co-expression network based methods are a few examples (Horváth et al., 2006; Shi et al., 2010; Wilkerson et al., 2010; Zhang et al., 2014a; Zhu et al., 2013). A network-based understanding of the impact of genetic alterations on gene-regulatory networks will further our understanding for identifying network-level targets (Creixell et al., 2012). One such example is sensitization of triple-negative breast cancer cell lines to DNA damaging agents by dynamic inhibition of the EGFR network (Lee et al., 2012).

Only recently, techniques such as RNA-seq are now being applied at a single-cell level to study intra-tumor heterogeneity in normal differentiation hierarchies (Desai et al., 2014; Treutlein et al., 2014) and cancer evolution (Patel et al., 2014). Quantitative imaging of genomic aberrations using immunoFISH coupled with surface markers have been used to elucidate genomic and phenotypic diversity in breast cancer patients at a single-cell level (Almendro et al., 2014b). High-throughput fluorescent flow cytometry, and more recently mass-cytometry/CyTOF have been applied for extensive phenotypic characterization and study of signaling network-level differences in cancer subpopulations to identify novel therapeutic strategies (Bodenmiller et al., 2012; Irish et al., 2010). Imaging-based techniques also help study drug sensitivities in distinct subpopulations.

These above methods provide a snapshot of the tumor heterogeneity at a given point in time. As we now know that tumor evolution and drug response is a highly dynamic process, quantitative techniques are warranted to follow the tumor cells over time and study time-dependent drug responses. Time-lapse quantitative imaging techniques such as live-cell reporters of signaling, transcription factor activity and cell

cycle, and analysis tools are currently being developed (Albeck et al., 2013; Loo et al., 2007; Miwa et al., 2014; Singh et al., 2010; Spencer et al., 2009; Tyson et al., 2012).

Impact of tumor heterogeneity on therapeutic response and resistance

Treatment decisions are generally based on the initial histology- and genetics-based assessment of a small biopsy, which might overlook some of the spatially separated genetic and non-genetic alterations at the primary and metastatic sites. Drug treatment significantly influences intra-tumor heterogeneity by shifting the balance between sensitive and resistant genetic clones (Figure 1.7A). Recurrent tumors exhibit some similar but also distinct genetic alterations (Johnson et al., 2014). EGFR T790M mutant or c-MET amplified clones are present in the initial tumor at low levels, which expand and lead to acquired resistance to EGFR TKIs (Bean et al., 2007; Engelman et al., 2007; Inukai et al., 2006).

Non-genetic mechanisms can also modify drug response dynamics to generate drug resistant and phenotypically distinct subpopulation, i.e. an alternate attractor state (Figure 1.7A and 1.8). The goal of treatment regimens should be to identify drug-sensitive attractor states which would induce growth-arrest and/or apoptosis, and use drugs to 'push' cells towards these alternative states (Huang, 2013) (Figure 1.8).

EGFR-TKI drug response and resistance could be dynamically altered in a reversible manner by modifying chromatin states (HDAC inhibitors) or via IGF-1R inhibition (Sharma et al., 2010). One patient with surgically resected ADC recurred with SCLC and Rb mutation suggesting transdifferentiation of ADC promoted by Rb loss (Peifer et al., 2012). In addition, 14% of NSCLC patients acquired EGFR TKI resistance

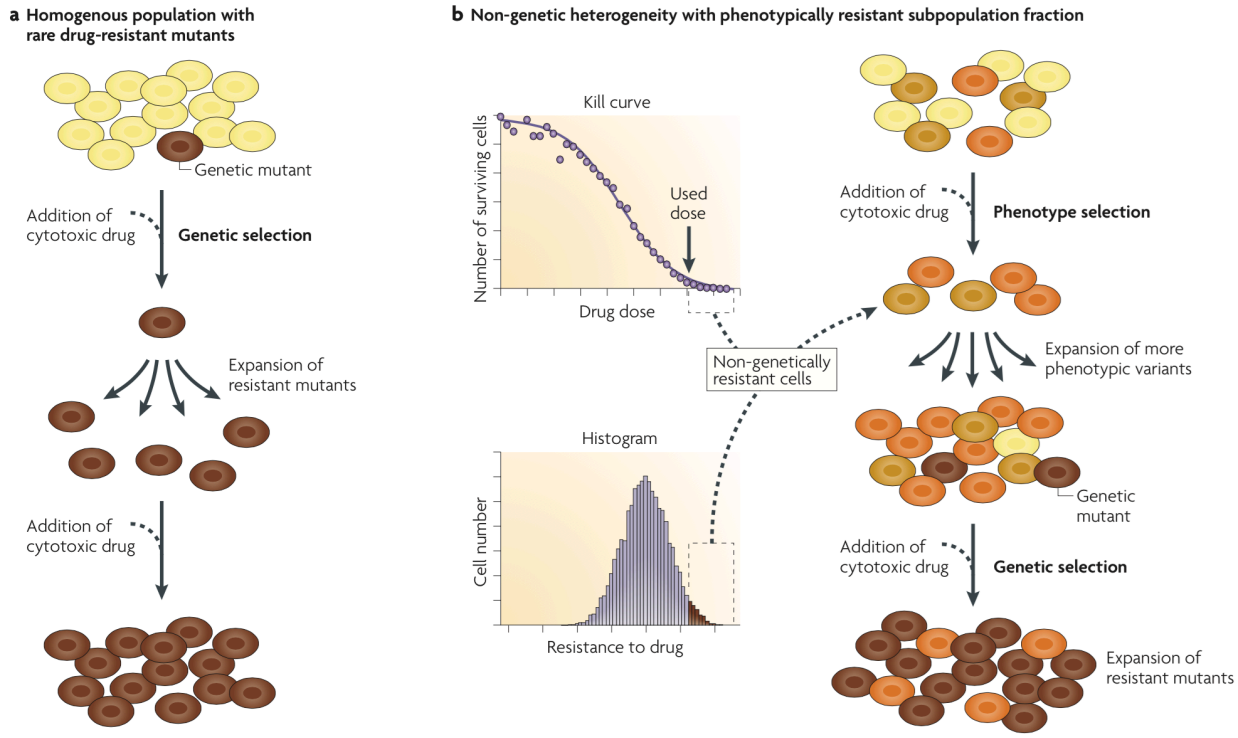


Figure 1.7: Impact of genetic and non-genetic heterogeneity on therapeutic resistance in cancer .
From (Brock et al., 2009).

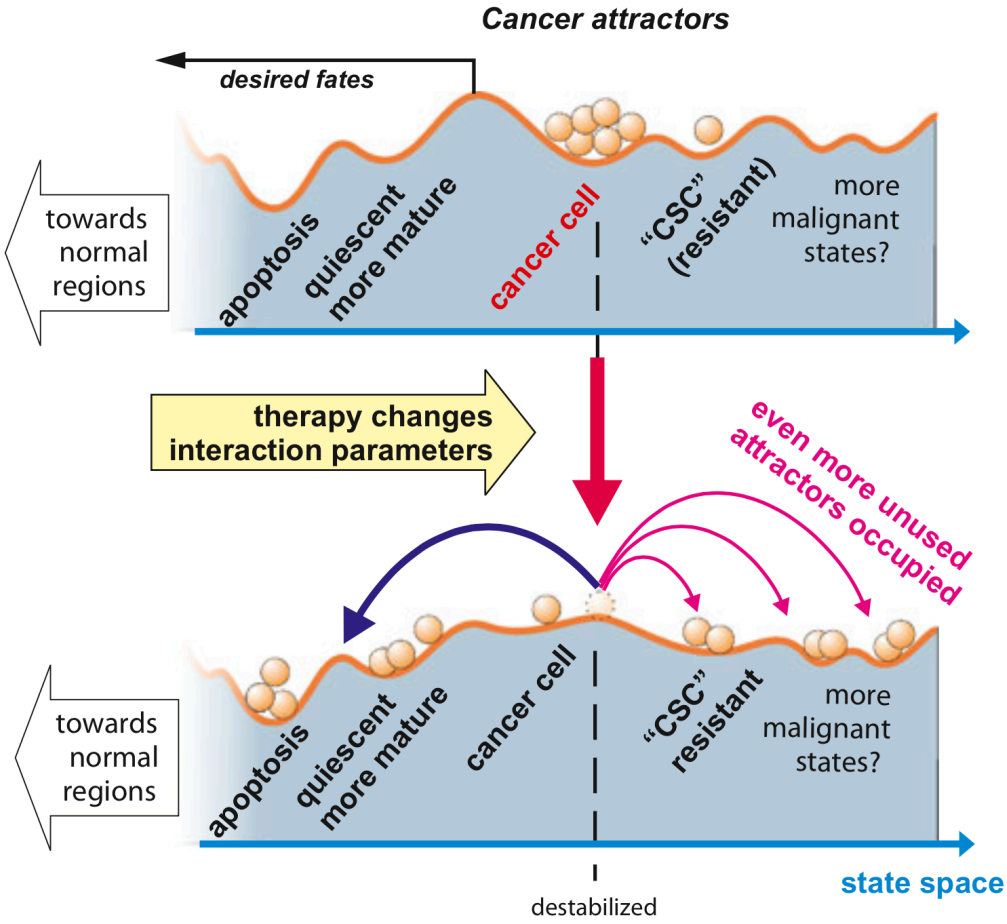


Figure 1.8: Theoretical model depicting impact of drug treatment on attractor states in cancer. From (Huang, 2013).

by switching to SCLC phenotype while maintaining the TKI-sensitizing EGFR mutation (Sequist et al., 2011). Epithelial-to-mesenchymal transition (EMT) and mesenchymal-to-epithelial transition (MET) are other examples of non-genetic heterogeneity that regulate invasion/metastasis and drug resistance, mediated by network-level changes in gene expression (Polyak and Weinberg, 2009; Tam and Weinberg, 2013). A drug-resistant mesenchymal phenotype was discovered in both ADC and SCC exhibiting phenotypic plasticity (Basu et al., 2010; Byers et al., 2013). Subtype-specific state-transitions were also induced chemotherapy treated breast cancer patients (Almendro et al., 2014a).

Summary and Dissertation Overview

NSCLC is a heterogeneous disease by histology and molecular features. Genomics-based studies has yielded a new level of classification of NSCLC with the identification of several overexpressed or mutated addicting oncogenes that can be successfully targeted in “molecular subtyped” NSCLC patients. On the other hand, SCLC suffers from several pitfalls in diagnosis and treatment with the mixed histologies with NSCLC and low-grade NET. There is no reproducible signature for SCLC diagnosis in the clinic. With increasing availability of high throughput cancer genomic datasets, several key issues are progressively becoming evident: 1) differential gene expression analysis yields a list of genes pertinent to a phenotype or outcome, however, the function of these genes is difficult to interpret since many might not be functionally related; 2) the gene expression signatures based on differential expression analysis, as a whole, are difficult to reproduce across datasets; 3) the number of potential

therapeutic targets (ranked according to differential expression scores) is ever growing, making target prioritization challenging due to the lack of functional insight.

Moreover, SCLC is still considered a monolithic disease with respect to its phenotype and treatment, where every patient is treated with a one-size-fits-all combination chemotherapy regimen. These patients die within a year after treatment due to widespread fatal metastasis. A better understanding of SCLC biology is needed on a global scale to better treat and manage this disease. Unlike NSCLC, single RTK targeted therapies (TT) haven't been clinically successful in SCLC patients, and it doesn't seem to be not driven by genetically altered oncogenic drivers, except for a couple such as FGFR1, MYC, which are currently being tested in clinical trials.

There is also a lack of understanding of global network-level changes that lead to rapid recurrence and disease progression with chemotherapy in SCLC. Also, one gene target approach might not be sufficient to attain sustained outcomes to a single RTK TT and resistance is inevitable as seen in NSCLC. It is important to realize that cellular signaling occurs in the form of networks, not a single gene or pathway. GRN involving transcription factors and their dynamics, regulate non-genetic heterogeneity. Thus, multiple genetic and non-genetic mechanisms allow cancer cells to evade treatment such as mutations in parallel signaling pathways, branched evolution and non-genetic phenotypic state transitions. Hence, combination therapies that study the response dynamics of cancer cells need to be the future of cancer treatment. A similar integrative systems biology approach is warranted in the study of SCLC etiology and discovery of novel targeted therapeutics.

I hypothesize that network-level changes in gene and protein expression might be playing a role in SCLC biology, which need to be studied in more detail. The primary goal of my thesis is to obtain a global understanding of SCLC biology using quantitative systems-biology approaches in order to identify and prioritize deregulated hubs in SCLC network that can serve as biomarkers and potentially novel therapeutic targets. Within this goal, I propose to identify network-based signatures for SCLC that would potentially delineate heterogeneity in SCLC using a combination of high-throughput genomic and proteomic datasets and identify potential subsets that would be sensitive to novel network-level targeted inhibition. Furthermore, I aim to understand the dynamics of the gene regulatory networks that drive SCLC biology in the context of stable attractor states using a experimental-bioinformatic-computational modeling framework. The ultimate goal is to understand the role of transcriptional networks in regulating SCLC phenotype, and use network-based biomarkers for quantitative monitoring of treatment response and resistance.

CHAPTER II

MATERIALS AND METHODS

Microarray data normalization

Public datasets on the Affymetrix platform GSE6044 (Rohrbeck et al., 2008), GSE4824 (Lockwood et al., 2008) were downloaded from GEO (Barrett et al., 2010) as CEL files. Cancer cell line encylopoedia (CCLE) dataset (Barretina et al., 2013) was downloaded from Broad Institute (http://www.broadinstitute.org/software/cprg/?q=node/11). Data were normalized and median centered using quantile RMA normalization using Affy Bioconductor package (Gautier et al., 2004) in R (Team, 2012). Agilent datasets, GSE11969 (Takeuchi et al., 2006) and our own Agilent dataset, were Lowess-normalized and median centered using GeneSpring (Chu et al., 2001). Probe-level data for all the datasets was converted to gene-level data by probe merging using the collapseRows function (Miller et al., 2011). Probes with no known gene symbols were removed from further analyses to reduce the dimensionality of the dataset.

WGCNA and Network analysis

The co-expression network analysis was performed in R using the WGCNA package as previously described and summarized in Figure 2.1 (Horváth et al., 2006; Winden et al., 2009). Briefly, all genes in the training dataset (GSE6044) or the Cancer cell line encyclopedia dataset (CCLE) were used to build an unsupervised co-expression based similarity matrix via Pearson's correlation coefficient.

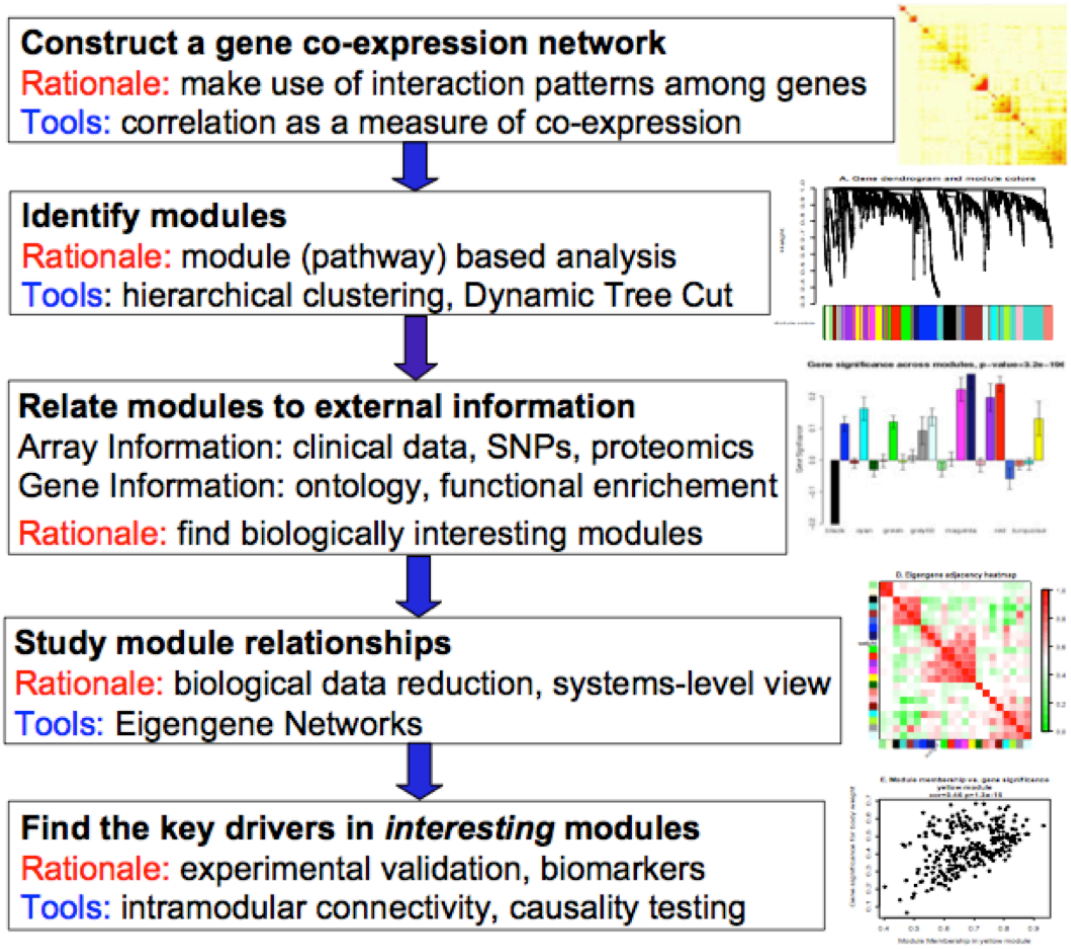


Figure 2.1: Weighted Gene-Coexpression Network Analysis (WGCNA) overview.

Adapted from (Langfelder and Horvath, 2008).

The similarity matrix was converted to a weighted adjacency matrix by raising it to a power β (β = 6) to amplify the strong connections and penalize the weaker connections (Langfelder and Horvath, 2008). Modules were generated using unsupervised average-linked hierarchical clustering with a cut-off of 0.9. This cut-off was chosen to minimize a large number of modules with very few genes, that is, less than 20 modules containing at least 100 genes. Each module is a hierarchical gene network.

Gene significance (GS): defined as $GSi = |cor(x_i, T)|$, indicates correlation of a x_i node expression profile to a phenotypic trait T, a binary trait variable across m samples (Langfelder and Horvath, 2008). In this case, phenotypic trait is lung tissue type - ADC, SCC, SCLC, and NL. Network hubs are defined as highly connected genes within a network, having high intramodular connectivity.

Intramodular connectivity is a measure of module eigengene-based connectivity (kME) (or module membership), defined as $K_{cor,i}^{(q)} = cor(x_i, E^{(q)})$, where $E^{(q)}$ is the module eigengene or 1st principal component of module q . Module hubs that have high GS are hubs that are significantly correlated to a phenotypic trait (Langfelder and Horvath, 2008), in our case, SCLC phenotype.

To filter hubs significantly correlated to SCLC phenotype and identify a SCLC specific hub network (SSHN), we used high values of GS, kME and differential expression (SCLC vs normal lung NL). To classify SCLC from other lung cancer types, unsupervised clustering of the SSHN genes was performed by bootstrapping analysis using pvclust package (Suzuki and Shimodaira, 2006). Bootstrapping analysis provides confidence values for the stability of each cluster derived by hierarchical clustering, via

resampling of the data. Heatmaps were generated using the gplots package (Ben Bolker et al., 2012).

For classification performance estimation, we used nested repeated 5-fold cross-validation procedure (Statnikov et al., 2005). The inner loop of cross-validation was used to determine the best parameters of the classifier (i.e., values of parameters yielding the best classification performance for the validation dataset). The outer loop of cross-validation was used for estimating the classification performance of the model that was built using the previously found best parameters by testing with an independent set of samples. To account for variance in performance estimation, we repeated this entire process (nested 5-fold cross-validation) for 10 different splits of the data into 5 cross-validation testing sets and averaged the results. Linear support vector machine is used as the classifier in our analysis, and the error penalty parameter was selected based on the nested cross-validation procedure.

Pathway analysis

Functional enrichment analysis of the SCLC hub network (SSHN) was performed using Webgestalt (Zhang et al., 2005) in Chapter III. This tool statistically compares the enrichment of SSHN genes with pathways contained in various databases such as Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG). Functional category enrichment in Webgestalt was tested by the hypergeometric test and multiple comparison corrections were made using Benjamini & Hochberg method (Benjamini and Y, 1995; Zhang et al., 2005).

Comparative Pathway enrichment analysis of the Blue and Turquoise modules was performed using BINGO and Enrichment map (Merico et al., 2010) and network visualization was conducted using Cytoscape (www.cytoscape.org) (Chapter IV). To summarize the expression data at the pathway level, we first transformed the data using a Z-score by subtracting the mean and dividing by the standard deviation of each gene across all samples. Next, in each sample we took the average expression of all genes that are simultaneously represented in a given pathway and a module (blue or turquoise) as the represented score for that module pathway combination.

RNAseq data generation and analysis

Tissue samples (20 samples: 10 with SCLC, 5 with SCC, and 5 normal bronchial brushings) were collected from the Vanderbilt University Medical Center and the University of Liverpool Hospital. Research protocols were approved by both institutions' Institutional Review Board. Total RNA was extracted from fresh frozen tumors and bronchial brushings by the RNeasy Kit (Qiagen, CA USA) according to the manufacturer's protocol. Whole transcriptome analysis (RNA-seq) was carried out by next-generation sequencing using Illumina platform in the lab of Vanderbilt Genome Sciences Resource. Next-generation sequencing methodology has been applied to sequence RNA from 20 tissue samples. Due to staged sequencing of samples, two technologies have been utilized: Illumina GAIIX and Illumina Hi-Seq. Sequencing runs from Illumina GAIIX (for 11 samples) were produced with 43bp reads and data was preprocessed using CASAVA 1.7 software. Sequencing runs from Illumina Hi-Seq (for 9 samples) were produced with 51bp reads and data was preprocessed with CASAVA 1.8

software. To make data from two platforms comparable, we have trimmed the last 8bp on each Illumina Hi-Seq read. Using 20 FASTQ data files (after Illumina Pass Filtering) with 43bp reads, we performed alignment using TopHat (v1.4.1), Bowtie (v0.12.7.0), and Samtools (v0.1.18) software. We experimented with two alignment approaches: with two seeds of 21bp and with one seed of 25bp. Since both alignment approaches led to very similar results (in terms of number and percentage of pass filter aligned reads and gene correlations with phenotypes in Fragments Per Kilobase of transcript per Million (FPKM) mapped reads data), we decided to use alignment with one seed of 25bp. Given aligned data, we computed gene expression FPKM (fragments per kilobase of exon per million fragments mapped) values using Cufflinks (v1.3.0) software and performed additional upper quintile normalization of Cufflinks. Using the resulting normalized gene expression dataset, we have assessed whether 287 SSHN genes are associated with SCLC vs. normal OR SCLC vs. SCC brushings by a two-sample t-test at 5% alpha level adjusted for multiple comparisons using the method (Byers et al., 2012).

Shotgun Proteomics

Shotgun proteomic analysis was performed from archival formalin fixed paraffin embedded tissues for pools of 5 ADC, 5 SCC, 5 SCLC as well as 5 non-cancerous alveolar lung and 5 bronchial epithelium tissue using our previously published methods (Sprung et al., 2009). Briefly, following deparaffinization with Sub-X, rehydration with ethanol-water, and protein solubilization in ammonium bicarbonate and trifluoroethanol, proteins were reduced, alkylated and digested overnight with trypsin. Tryptic peptides

were separated by isoelectric focusing using ZOOM IPGRunner IEF strips (Invitrogen) with an immobilized pH gradient of 3.5-4.7 (Slebos et al., 2008). LC-MS/MS analyses were performed on an LTQ-XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with an Eksigent nanoLC 1D plus pump and Eksigent autosampler (Dublin, CA) as described previously (Sprung et al., 2009). MS/MS spectra were processed for protein identifications using a data analysis pipeline described previously (Ma et al., 2009; Tabb et al., 2007; 2011). False positive peptide-spectrum matches were estimated by reversed database search (Peifer et al., 2012) and held at 5%. Further filtering to require at least one identified spectrum per sample across all analyses maintained a protein false discovery rate (FDR) (Benjamini and Y, 1995) below 5%. To compare protein expression differences between different histology groups (for example, SCLC vs. Normal), we applied our quasi-likelihood model and analysis software QuasiTel to analyze spectral count data (Li et al., 2010). The quasi-likelihood model, with no restriction on the distribution assumptions, is appropriate for modeling count data with overdispersion and/or underdispersion issue that is frequently observed in spectral count data. Multiple comparison adjusted p values (quasi-FDR) were calculated by incorporating the FDR method described previously (Benjamini and Y, 1995).

Transcriptional regulatory network construction

ARACNe is an information theoretic algorithm for the inference of gene regulatory networks using a large compendium of gene expression profiles. It identifies statistically significant gene-gene co-regulation by mutual information and then it

eliminates any indirect regulation where two genes are co-regulated through one or more intermediate genes using a well-established data transmission theory called data processing inequality (DPI) (Margolin et al., 2006). To generate a SCLC-specific transcriptional network, we used gene expression profiles from 53 SCLC cell lines from the CCLE resulting in 27224 interactions among 8706 nodes comprising TFs and non-TF genes. This SCLC-specific network was generated with the bootstrap version of ARACNe (1) algorithm, which helps reduce false negative connections in the network using the following parameters: pvalue = 10^{-7}, dpi = 0 and 100 bootstraps. To evaluate if genes in Blue and Turquoise module are enriched for targets of any specific TF, we checked for independent overlap between TF targets from the ARACNe inferred SCLC gene regulatory network and the 1179 and 3471 genes in Blue and Turquoise module, respectively, using Fisher's Exact Test (FET) (Carro et al., 2010). We selected all TFs to be candidate master regulators if the FET p-value was ≤ 0.05, leaving 96 and 207 TFs for Blue and Turquoise module, respectively. Of these, 23 TFs were common to both modules (Figure 3.5).

Furthermore, we independently validated the ARACNe predicted regulatory TFs with CHIP-Seq and TF-target binding site prediction databases such as CHEA (Lachmann et al., 2010), ENCODE (Landt et al., 2012), TRANSFAC (Matys et al., 2003), JASPAR (Mathelier et al., 2014) using the EnrichR tool (http://amp.pharm.mssm.edu/Enrichr/) and literature databases such as Pubmed (http://www.ncbi.nlm.nih.gov/pubmed/) and Glad4U (http://bioinfo.vanderbilt.edu/glad4u/). With these filtration steps, we identified a list of 76 likely TF regulators of NE and/or ML differentiation. Finally, to identify TFs which play an active role in maintaining the

phenotypic differences between the SCLC cell lines, we considered only heterogeneously expressed TFs (median absolute deviation above the 50th percentile) yielding a final list of 38 TFs that we used to build a boolean network model for SCLC. We then extracted interactions between these TFs from manual curation of literature, ChIP-X databases (ChEA (Lachmann et al., 2010), ENCODE (Landt et al., 2012)) and TF-binding motif predictions (TRANSFAC (Matys et al., 2003), JASPAR (Mathelier et al., 2014)).

Boolean network model simulation and analysis

To infer a directed network of interactions, transcriptional regulation was determined using Enrichr (Chen et al., 2013) which queries TRANSFAC (Matys et al., 2003), JASPAR (Mathelier et al., 2014), ChEA (Lachmann et al., 2010), and ENCODE (Maher, 2012) to produce a directed interaction network. Estimates of regulatory edge weights were inferred from the literature where possible; otherwise they were assigned as +/-1 if the nodes were positively or negatively correlated, respectively. The TFs that do not have incoming edges (5) were eliminated.

The TF network was simulated as a Boolean network where each node was either ON (active) or OFF (inactive). Nodes were updated with a random order asynchronous scheme, such that the list of TFs was randomly shuffled and each node updated exactly once, in order, before the whole list was shuffled again. If the total weight of active regulators (where activators have positive weight and repressors have negative weight) was positive, the node was switched on; if the total weight was negative, the node was switched off; if the total weight was zero, the node did not

change. A state transition network was constructed where each of the 2^33 possible states was updated 30 times to sample the possible asynchronous update orders, and an edge was drawn from the initial state to the updated state. The attracting components algorithm from NetworkX (http://networkx.github.io/documentation/latest/reference/index.html) was used to identify attractors from the directed state transition graph. Once the 33 core transcription factors were simulated, the 5 leaf TFs (those without dynamic feedback loops) were then assigned based on the current network state.

A correlation score was calculated for each attractor against each cell line. Using the CCLE expression data for a given cell line, expression of the transcription factors were linearly scaled to be between 0 and 1, and Pearson's r was calculated between the attractor vector and the cell line vector.

Consensus clustering analysis

Unsupervised consensus clustering analysis was performed using ConsensusClusterPlus package in R (Wilkerson and Hayes, 2010) on either 53 SCLC cell line dataset (Barretina et al., 2013) or 28 SCLC patients (Clinical Lung Cancer Genome Project (CLCGP)Network Genomic Medicine NGM, 2013), with 80% sub-sampling of both genes and samples with a N=1000. The distance metric used was 1 - Spearman correlation with k-means distance function. Genes used were 1179 and 3471 of the Blue and Turquoise modules respectively identified in 53 SCLC cell lines from the CCLE dataset. This method is typically used for unsupervised class/subset discovery in a large group of patients. It uses the cumulative distribution function (CDF) plots to

denote the confidence of class prediction where the best-fit line is defined as one that is flat and linear between consensus score 0 and 1. The consensus matrix heatmaps are plotted as samples versus themselves (Wilkerson and Hayes, 2010).

Antibodies and Reagents

Antibodies used for western blotting include GAD1/2, EphA2, PDGFR (Cellsignal®), E-cadherin (BD Biosciences), Vimentin (SantaCruz), OVOL2, CBFA2T2, SOX2, ASCL1, POU5F1, GAPDH, SOX2, FYN, beta-actin (Sigma®). Fluorophore-conjugated primary antibodies were used for flow cytometry – CD56 BV605, CD151 PE, CD24 BUV395 (BD Biosciences), CD44 Pacific blue, E-cadherin Pacific blue (Cellsignal®), CADM1 A647 (MBL), EPHA2 A488 (R&D Systems), CD133 PE-Cy7 (Biolegend), Vimentin (Santacruz).

Drugs used for viability and flow cytometry studies were procured from Sigma (Cisplatin, Etoposide) and Selleck-Chem (Valproic acid, Trichostatin A, 5-azacitidine) and all high concentration stocks were made in Dimethyl sulfoxide (DMSO - Sigma®) and stored at -80.

Cell culture

All SCLC cell lines were purchased from ATCC (www.atcc.org) and grown in company recommended media conditions. NCI-H146, NCI-H209, NCI-H69, NCI-H524, NCI-H526, NCI-H211, NCI-H82, NCI-H196, NCI-H446 were grown in RPMI 1640 (GIBCO®) and 10% FBS. NCI-H2141, NCI-H1184, NCI-H2171, NCI-H1048, NCI-H841, SW1271 were grown in ATCC recommended HITES media (DMEM-F12 supplemented

with insulin, transferrin, selenium, 5% FBS, 2.5 mM Glutamine, beta-estradiol, hydrocortisone) while DMS153, DMS53, DMS114 were grown in Waymouth media (GIBCO®) supplemented with 10% FBS. HBECKT was grown in ATCC recommended Keratinocyte serum free media.

Xenograft mouse studies

Patient derived xenograft (PDX) are human tumors that never touch plastic and are only propagated in athymic nude-Foxn1$^{nu/nu}$ mice (Jackson laboratories). The PDXs LX-22 and NJ-H29 were obtained from Charles Rudin laboratory (Memorial Sloan Kettering Center) and Julien Sage lab (Stanford University) respectively. These were obtained as frozen vials of cells that were thawed at 37 degrees and injected at 1:1 ratio in nude mice (Jackson labs) with matrigel (BD). For subsequent analysis of tissue, the tumor was dissected for (1) immunohistochemistry by formalin fixation, (2) protein by snap freezing in liquid nitrogen. For propagation and flow cytometry analysis, tumor tissue was minced with sterile razor blade, filtered through 70um mesh filter, washed several times in serum-free RPMI media and spun down at 1000rpm for 5 minutes at room temperature. Red blood cells were lysed using ACK lysis buffer (GIBCO). Cells were either frozen in RPMI media with 10% FBS and 5%DMSO (Sigma) or used for flow cytometry analysis immediately.

Western blotting

All cell lines were plated for 2 days in complete medium to achieve equilibrium in signaling states. For siRNA experiments, 400,000 cells were transfected using

Dharmafect 4 transfection reagent and siRNA (Dharmacon®) in 6-well plates. Cells were incubated for either 3 or 7 days followed by lysate preparation and western blotting process as detailed below. Lysates were prepared by spinning cells down 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 30secs and centrifuged at 15000 rpm for 15mins (at 4˚C). The protein concentration was quantified using BCA assay (Pierce®). Lysates were boiled for 10minutes at 100 degrees with 1X NuPage sample buffer (Molecular Probes®) and run on 8% or 4-12% Tris-glycine gels (Molecular Probes®). Semi-dry transfer was performed followed by blocking with 1X Casein-TBS. Blots were imaged using chemiluminescence or Odyssey. The band intensities were quantified using ImageJ and plotted in R (www.r-project.org).

Tissue microarray immunostaining and analysis

Two TMAs of SCLC specimens were prepared from formalin-fixed paraffin-embedded (FFPE) tissue blocks following previously reported methods (Renshaw et al., 2005). Pathology blocks were retrieved from the archives of the Department of Pathology at Vanderbilt University Medical Center, Nashville VA Medical Center and St-Thomas Hospital in Nashville, Tennessee. They were obtained between 1996 and 2008 from 85 patients who had surgery or bronchoscopy prior to medical treatment. SCLC diagnosis was confirmed on hematoxylin and eosin-stained sections by an experienced lung cancer pathologist (RE). The study was approved by Institutional Review Boards at each medical center. The Syk/Fyn/BRCA1 IHC was examined in two to five spots for

each TMA. The intensity of staining was scored as 0-no staining, 1-weak, 2-moderate, and 3-strong and the percentage of area stained was also measured. The IHC score was determined by multiplying intensity score to the percentage area stained. The highest score among the spots was used for the unsupervised clustering analysis of protein expression. Tumor images were captured by brighfield microscopy using the Leica SCN400 system (Leica Biosystems®) at 20X magnification. Kaplan-Maier survival analysis was performed using survival package (Therneau and Grambsch, 2000) in R.

Bright field imaging

Cells were plated overnight in 10cm dishes and imaged the next day. High resolution brightfield images of SCLC cell lines were captured using Leica microscope using 20X magnification lens.

Cellavista viability assay

10000 cells were plated in 100ul of complete medium (RPMI 1640 containing 10% Fetal bovine serum) in each well of a 96-well plate with Dharmafect 4 and siRNA mixture (Thermo-Scientific). The reagent dilutions and transfection procedures were performed as per the manufacturer's protocol. Cells were incubated at 37˚C until each timepoint. At each timepoint, cells were transferred to a BD Falcon 96-well black clear bottom imaging plate and live-dead viability dyes (calcein – live cells; ethidium homodimer – dead cells) and hoescht 33342 for total nuclei (Invitrogen®) were added in complete medium. The cells were incubated with the dyes for 15mins at 37˚C followed by imaging using the Cellavista high-throughput imaging microscope (SynenTec,

Elmshorn, Germany). The Roche cell viability protocol was used to image and quantify the cells in 3 colors as per manufacturer's instructions. The output generated from this algorithm included total cell number, viable cell count, percent live/dead cells, etc. The data plotting and statistics were done using R (Team, 2012). The viability growth curves statistics were generated using a linear regression growth model (Kutner, 2005). Multiple comparison of treatments were derived using ANOVA and Tukey's method (Kuehl, 2000; Kutner, 2005). The p-values for percent dead at day 5 were generated using a paired t-test, pairing across, N=4, experimental replicates. IC50 values were calculated by fitting nonlinear regression models using drc package in R.

Oxygen consumption assay

1-2 million cells were plated in T75 flasks and allowed to grow exponentially in complete media for 2 days. On day 2, the cells were collected in 15ml conical tubes, washed with serum-free glucose-free RPMI media 1-2 times and spun down for 5 minutes at 1000rpm. Cells (approximately 1 million) were then added into each of the two chambers of the O2k Oxygraph (O2k Oxygraph, Oroboros Instruments, Innsbruck, Dr. Josh Fessel's lab). The oxygraph uses a closed system polarographic oxygen electrode to measure real-time changes in oxygen concentrations and oxygen consumption rates with extremely high sensitivity. Oxygen flux rates are calculated in real time as the negative derivative with respect to time of the oxygen concentration curve, so that decreases in oxygen concentration are expressed as positive flux rates. Additionally, fluxes are normalized to cell number so that the readout is a specific flux of oxygen expressed as pmol O2 consumed/sec/million cells. To probe specific aspects of

mitochondrial function, specific combinations of substrates and inhibitors were added sequentially, and oxygen fluxes as well as changes in flux rates in response to these stimuli were quantified. Substrates and inhibitors added were as described below:

Digitonin (typical working concentration 4μM) – Selectively permeabilizes the plasma membrane to allow for efficient diffusion of substrates into and out of the cells, while leaving the mitochondrial membrane intact. This removes the effects of transporters, diffusion barriers from plasma membrane, etc. It also allows endogenous substrates to leak out of the cell, so that we can see the true effect of addition of the substrates and inhibitors of interest.

Glucose (5.5mM) and glutamine (Gln, 2mM) - Cells were in glucose-free media, so first glucose was added as a substrate, followed by glutamine.

Glutamate (G, 10mM) and malate (M, 0.5mM) – Fuels for Complex I in the mitochondria.

ADP (2.5mM)- Induces state 3 respiration, maximizing oxygen consumption linked to ATP synthesis.

Succinate (Suc, 10mM)- Fuel for succinate dehydrogenase, which is Complex II in the electron transport system and also part of the TCA cycle. In the presence of glucose, glutamine, glutamate, malate, ADP, and succinate, the electron transport system is essentially fully engaged, and oxygen consumption linked to ATP production (OXPHOS-linked O2 flux) is maximal.

Carbonyl cyanide m-chlorophenylhydrazone (CCCP, 0.5-1μM) – A protonophore that destroys the proton gradient across the inner mitochondrial membrane. This uncouples

ATP synthesis from oxygen consumption and effectively pushes the oxygen flux rate to the maximum possible for the cell.

Flow cytometry data generation and analysis

For steady-state expression of surface markers, 1-2million 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 at 37 degrees for the indicated time points and then collected for flow experiments.

Cells were dissociated using TryplE (GIBCO®) for 10-15 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 (Table 2.1) 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 cytometer instrument at the Vanderbilt Flow cytometry shared resource core. First fluorescent channels were compensated using anti-mouse IgK beads (BD Biosciences) that were tagged with fluorescent antibodies. This compensation was applied to all the samples to eliminate spectral overlap of fluorescent channels.

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 (Figure 2.2A). This was followed by gating for Alexa700 negative viable cells, where A700 positive

Marker	Fluorophore	Type	Catalog
CD24	BUV395	Neuroendocrine	BD biosciences
CD56	BV605	Neuroendocrine	BD biosciences
CADM1	A647	Neuroendocrine	MBL
CD151	PE	Mesenchymal	BD biosciences
CD44	Pacific Blue	Mesenchymal	Cellsignal
EPHA2	A488	Mesenchymal	R&D Biosystems
CD133	PE-Cy7	Stem-cell	Biolegend
Viability	A700	Viability	Molecular Probes

Table 2.1: Surface marker panel for quantifying phenotypic heterogeneity in SCLC cell lines and PDX models.

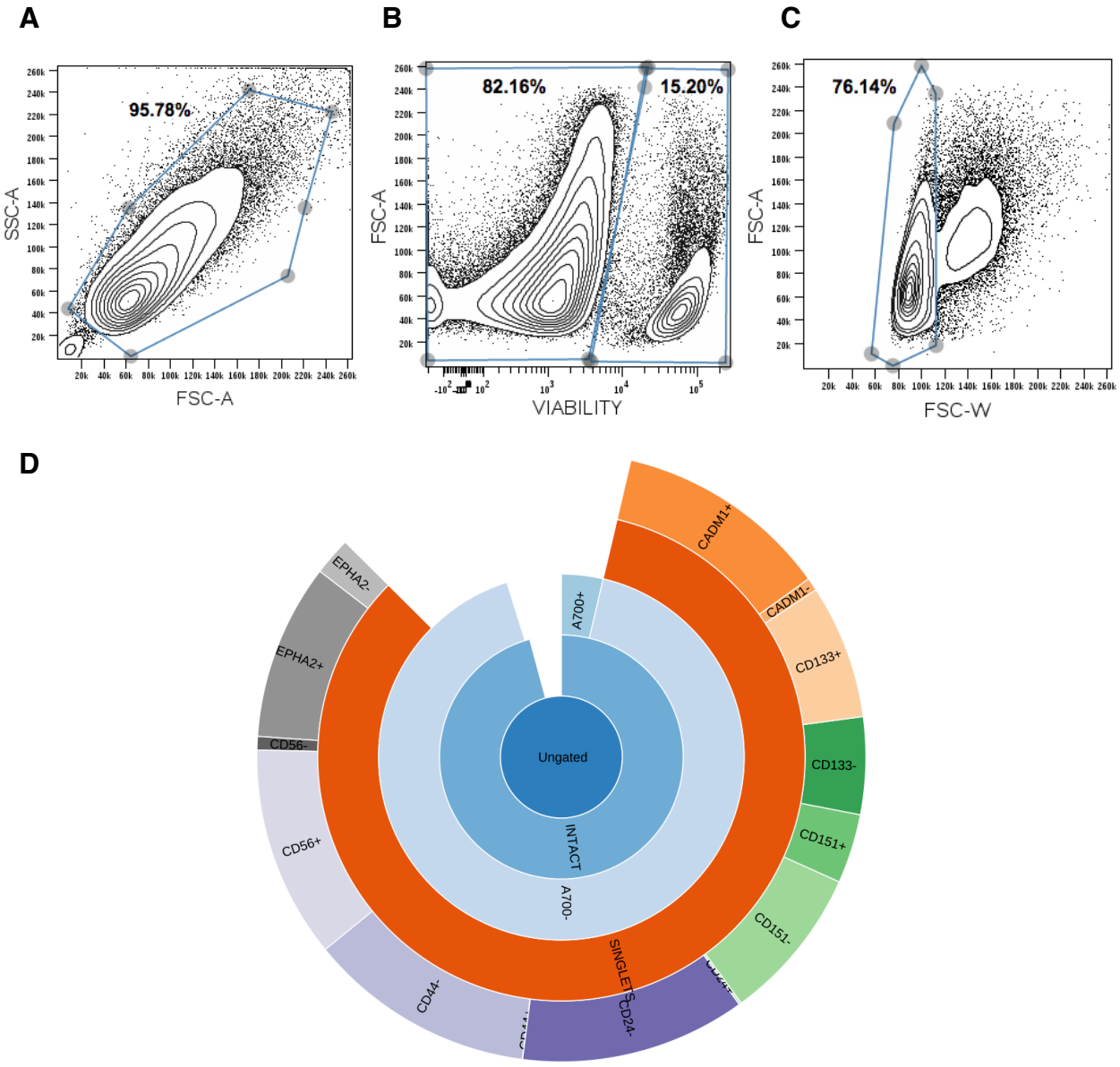


Figure 2.2: Gating scheme for viable singlet cells in flow cytometry experiments.

Analysis generated in Cytobank (https://irishlab.cytobank.org). (A) Gating of intact cells based on FSC-A and SCC-A. (B) Biaxial plot of A700 (viability) on X-axis and FSC-A on Y-axis where low A700 staining (left gate) denotes viable cells (A700-) and A700 positive staining (right gate) indicates dead/dying cells. (C) shows gating of singlet cells based on FSC-A and FSC-width plot where singlets and doublets (cells stuck together/dividing cells) can be detected via their separate density distributions. (D) shows a population sunburst plot where all subsequent gates of fluorescent markers are created on intact-viable singlet cell populations.

populations are dead/dying cells (Figure 2.2B). Finally the A700 negative viable cells are further gated to include only singlet populations (Figure 2.2C). At least 20000-30000 viable singlet cells were collected per sample. All subsequent gates on fluorescent markers are made on viable singlet cells depicted in the sunburst plot (Figure 2.2D).

The raw cytometer intensity readouts for fluorescent channels were first converted to log scale by using the asinh() function in Cytobank® (https://irishlab.cytobank.org/cytobank) or R (www.rproject.org) by dividing by a co-factor of 150. Data analysis (gating and plots) and visualization were conducted in Cytobank and cyt tool in Matlab (http://www.c2b2.columbia.edu/danapeerlab/html/cyt.html). Box-and-whisker plots and its statistics (Kruskal-Wallis test) were conducted using WGCNA package in R (Horváth et al., 2006).

NbClust package was used for estimation of the optimal number of clusters based on 30 indices that provide confidence in the cluster number estimation to be used for k-means clustering analysis in R (Charrad et al., 2014). k-means clustering analysis and overlay visualization plots (on CD56 and CD151 in Figure 4.12) were performed using the cyt tool in Matlab. These FCS files that now contained k-means cluster assignment for each cell were used for SPADE analysis.

High-dimensional data reduction techniques using clustering algorithms such as SPADE was conducted in the web interface of Cytobank®. SPADE algorithm uses density-based downsampling the high-dimensional flow cytometry data to include rare events, followed by agglomerative clustering to generate a minimum spanning tree by clustering the data into a large set of nodes (typically 200) (Figure 2.3).

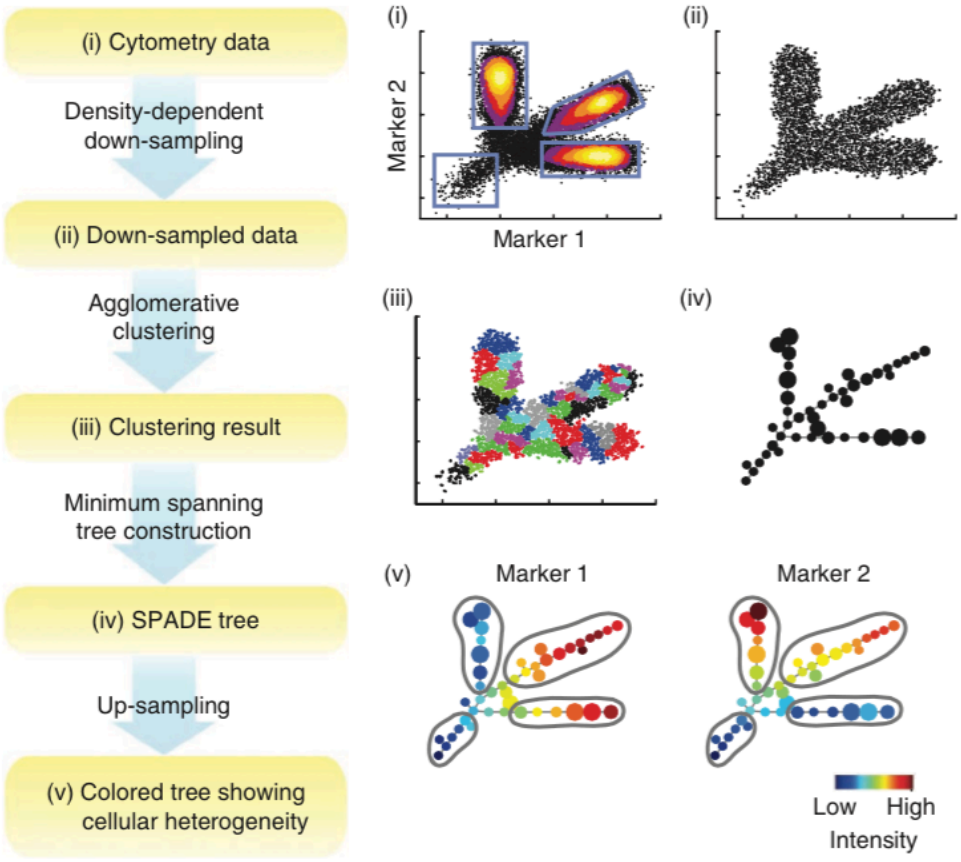


Figure 2.3: SPADE method summary.

Adapted from (Simonds et al., 2011).

We used 6 markers - 3 NE and 3 ML (Table 2.1) to run SPADE with the default setting of k=200 and singlet gated viable cells (10000 randomly sampled events per sample) 22 SCLC cell lines, including 2 PDXs. The same settings were used for drug treatment samples as well. Manual gating or 'bubbles' of NE and ML phenotypic states was performed in Cytobank itself via co-expression of either NE or ML markers respectively within each cell line. The nodes labeled as NE had high co-overexpression of NE markers CD56, CD24 and CADM1 and low expression of ML markers CD44, CD151, and EPHA2, and vice versa for ML nodes. These manually added NE and ML bubbles were further validated by overlay of the k-means predicted NE and ML clusters on the SPADE plots.

Fluorescent barcoding for signaling experiments

Cells were stimulated with 24 distinct stimuli shown in Table 2.2, each for 15minutes, followed by fixation with 2% PFA and permeabilization with 1ml 100% ice-cold methanol as described above. Cells were then washed with 1ml PBS and spun down for 5 minutes at 2000rpm and room temperature. Cells were then resuspended into 190ul PBS and added to each of the barcoding wells containing 10ul total of barcoding mix (Pacific blue PB and Pacific orange PO) dyes (Molecular Probes), thus assigning a specific sample a fluorescent barcode (ratio of PB and PO dyes) (Table 2.2). Cells were incubated with barcoding mix for 30 minutes in the dark, followed by washing with PBS containing 1% Bovine serum albumin (BSA). The cells from all 12 wells were mixed into 1 tube and then washed followed by spinning down at 2000rpm for 5 minutes. So cells from the 2 tubes, each now containing 12 samples, and each

with a specific barcode, were distributed into 3 staining panels shown in Table 2.3. Cells were stained with fluorescent conjugated antibodies for 30 minutes in dark at room temperature. Samples were washed with PBS and run on BD Fortessa cytometer instrument at the Vanderbilt Flow cytometry shared resource core. First fluorescent channels were compensated using anti-mouse IgK beads (BD Biosciences) that were tagged with fluorescent antibodies. This compensation was applied to all the samples to eliminate spectral overlap of fluorescent channels. 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 (Figure 2.4A). This was followed by gating for viable cells to include only singlet populations (Figure 2.4B). At least 500,000 viable singlet cells were collected per sample. All subsequent gates on fluorescent markers are made on viable singlet cells. A barcoded sample containing 12 distinct populations (stimulation conditions) is shown in Figure 2.4C. These data were analyzed in Cytobank (https://irishlab.cytobank.org/cytobank).

Samples	Barcode*	Barcoding panel 1	Barcoding panel 2
1	PO L1 : PB L1	Unstimulated 1	Unstimulated 2
2	PO L1 : PB L2	IL-2 20ng/ml	INFalpha 20ng/ml
3	PO L1 : PB L3	IL-3 20ng/ml	INFgamma 20ng/ml
4	PO L1 : PB L4	IL-4 20ng/ml	EGF 100ng/ml
5	PO L2 : PB L1	IL-6 20ng/ml	HGF 20ng/ml
6	PO L2 : PB L2	IL-7 20ng/ml	IGF 20ng/ml
7	PO L2 : PB L3	IL-8 20ng/ml	NGF 20ng/ml
8	PO L2 : PB L4	IL-9 20ng/ml	SCF 50ng/ml
9	PO L3 : PB L1	IL-10 20ng/ml	G-SCF 20ng/ml
10	PO L3 : PB L2	IL-13 20ng/ml	PDGF-BB 25ng/ml
11	PO L3 : PB L3	IL-15 20ng/ml	NRG-1 20ng/ml
12	PO L3 : PB L4	IL-21 20ng/ml	H2O2 3.3mM

Barcode* : defined by levels of Pacific orange (PO) and Pacific blue (PB) as given below.

Levels	Pacific orange concentration (ug/ml)	Pacific Blue concentration (ug/ml)
L1	0.34859	0.03115
L2	3.48592	0.24924
L3	20.91556	1.49704
L4	-	7.48922

Table 2.2: Signaling stimuli conditions and fluorescent barcoding panel.

Channel	Staining Panel 1	Staining Panel 2	Staining Panel 3
Alexa 488	P-LCK	P-EGFR	P-4EBP1
PE	P-AKT1/2	P-SRC	P-STAT3
Alexa 647	P-SHP2	P-PLCg	P-S6
Per-CP Cy5.5	P-ERK1/2	P-STAT6	P-STAT1
PE-Cy7	P-P38	P-NFKB	P-STAT5

Table 2.3: Signaling antibody marker panels for barcoding experiments.

P- denotes Phosphorylated form of the protein. All antibodies were purchased from BD Biosciences.

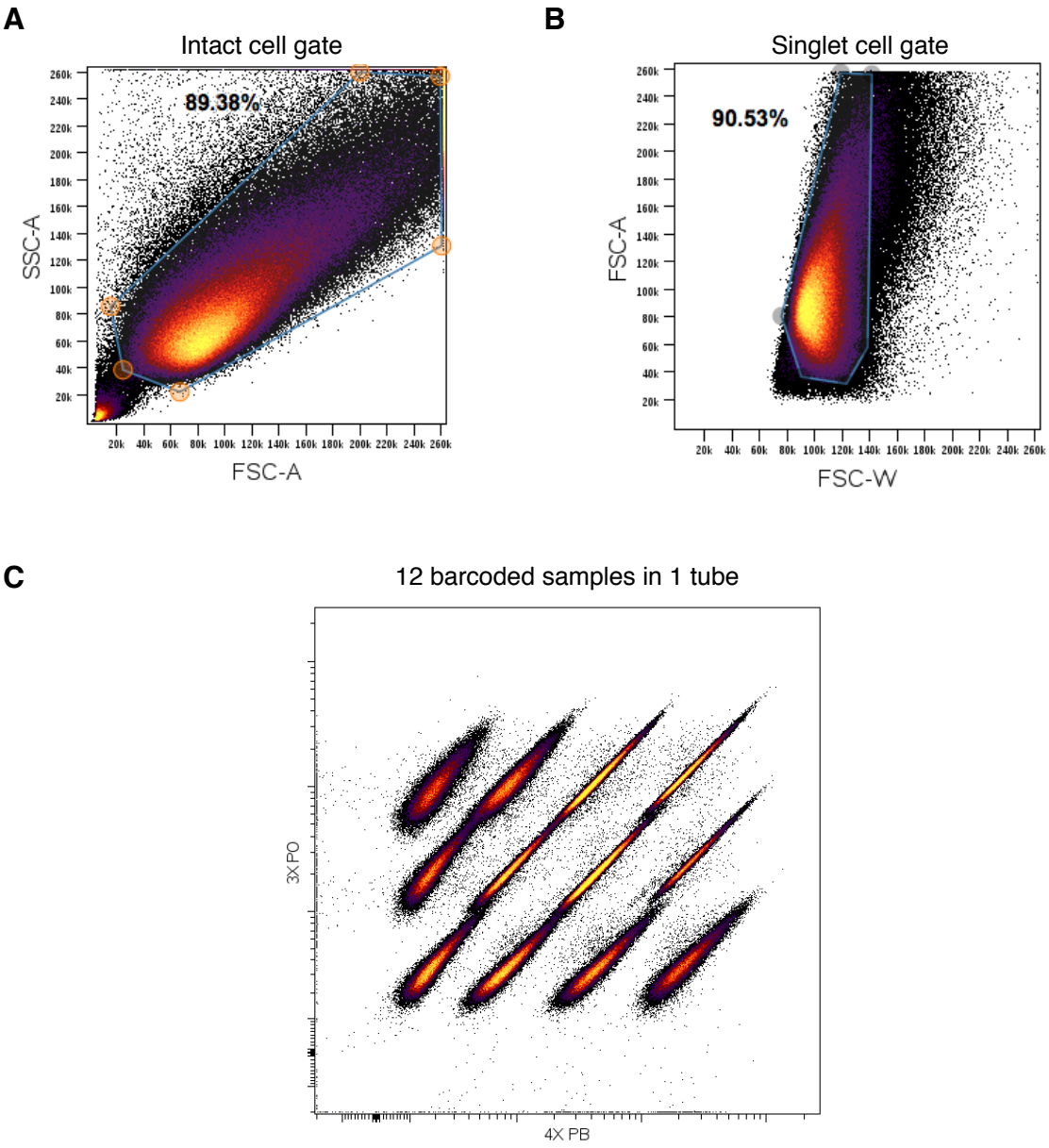


Figure 2.4: Barcoding signaling experiment gating scheme.

CHAPTER III

GENE CO-EXPRESSION NETWORK ANALYSIS IDENTIFIES SPLEEN TYROSINE KINASE (SYK) AS A CANDIDATE ONCOGENIC DRIVER IN A SUBSET OF SMALL-CELL LUNG CANCER[1]

Abstract

Background: Oncogenic mechanisms in small-cell lung cancer (SCLC) remain poorly understood leaving this tumor with the worst prognosis among all lung cancers. Unlike other cancer types, sequencing genomic approaches have been of limited success in small-cell lung cancer, i.e., no mutated oncogenes with potential driver characteristics have emerged, as it is the case for activating mutations of epidermal growth factor receptor in non-small-cell lung cancer. Differential gene expression analysis has also produced SCLC signatures with limited application, since they are generally not robust across datasets. Nonetheless, additional genomic approaches are warranted, due to the increasing availability of suitable small-cell lung cancer datasets. Gene co-expression network approaches are a recent and promising avenue, since they have been successful in identifying gene modules that drive phenotypic traits in several biological systems, including other cancer types.

Results: We derived an SCLC-specific classifier from weighted gene co-expression network analysis (WGCNA) of a lung cancer dataset. The classifier, termed SCLC-

[1] This work has been published under Udyavar AR, Hoeksema MD, Clark JE, Zou Y, Tang Z, Li Z, Li M, Chen H, Statnikov A, Shyr Y, Liebler DC, Field J, Eisenberg R, Estrada L, Massion PP, Quaranta V. 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; 7 Suppl 5:S1.

specific hub network (SSHN), robustly separates SCLC from other lung cancer types across multiple datasets and multiple platforms, including RNA-seq and shotgun proteomics. The classifier was also conserved in SCLC cell lines. SSHN is enriched for co-expressed signaling network hubs strongly associated with the SCLC phenotype. Twenty of these hubs are actionable kinases with oncogenic potential, among which spleen tyrosine kinase (SYK) exhibits one of the highest overall statistical associations to SCLC. In patient tissue microarrays and cell lines, SCLC can be separated into SYK-positive and -negative. SYK siRNA decreases proliferation rate and increases cell death of SYK-positive SCLC cell lines, suggesting a role for SYK as an oncogenic driver in a subset of SCLC.

Conclusions: SCLC treatment has thus far been limited to chemotherapy and radiation. Our WGCNA analysis identifies SYK both as a candidate biomarker to stratify SCLC patients and as a potential therapeutic target. In summary, WGCNA represents an alternative strategy to large scale sequencing for the identification of potential oncogenic drivers, based on a systems view of signaling networks. This strategy is especially useful in cancer types where no actionable mutations have emerged.

Introduction

Small-cell lung cancer (SCLC) represent up to 15 % of lung cancers and pose a major challenge as we are unable to diagnose it early, its most aggressive clinical behavior and the lack of lasting benefit from therapy. Patients presenting with this neuroendocrine tumor of the lung have a dismal 5% 5-year survival rate. Although SCLC is highly sensitive to chemotherapy and radiation, it invariably recurs with fatal

widespread metastasis (Rosti et al., 2006). In contrast to non-small cell lung cancer (NSCLC), to date no specific genetic biomarkers or molecular subtypes have been identified in SCLC (Ettinger, 2006). Gene expression profiling has had limited success in SCLC stratification for the purpose of personalized treatment. Although recent advances in genomic analysis of SCLC have identified potential driver mutations in SCLC (Peifer et al., 2012; Sos et al., 2012; Staaf et al., 2012), there remains an unmet need for approaches that can stratify SCLC patients and/or uncover viable molecular targets in SCLC.

To meet this challenge, we turned to weighted gene co-expression gene network analysis (WGCNA), a recently introduced bioinformatics method that captures complex relationships between genes and phenotypes. The distinct advantage over other methods, such as differential gene expression, is that WGCNA transforms gene expression data into functional modules of co-expressed genes without any prior assumptions about genes/phenotypes, providing insights into signaling networks that may be responsible for phenotypic traits of interest (Horváth et al., 2006; Shi et al., 2010; Winden et al., 2009). In lung cancer, its potential remains unexplored.

Our WGCNA analysis of a public lung tumor dataset (Rohrbeck et al., 2008) revealed a module of co-expressed genes specific to SCLC. After filtering, the SCLC-specific module was reduced to a SCLC-specific hub network (SSHN) signature that classified SCLC from other lung cancer types in several public and in-house tumor datasets (including independent high-throughput screening techniques such as RNAseq and shotgun proteomics), and in lung cancer cell lines. SSHN was enriched for hubs in signaling networks known to be associated with SCLC pathogenesis, including cell

cycle, oxidative stress response and DNA damage response. As a proof of concept, we chose to validate oncogenic kinase hubs (20 kinase genes) within SSHN, as they provide special translational relevance as potential candidates for targeted therapy and also play key roles in various hallmarks of cancer. Among the twenty, spleen tyrosine kinase (SYK), a previously undescribed target in SCLC, exhibited one of the highest overall statistical associations with the SCLC phenotype, based on WGCNA gene significance (GS, see Methods) and overexpression in shotgun proteomics, and was therefore selected for further validation as a target.

SYK has been previously investigated most extensively in the context of lymphocyte development and as a therapeutic target in hematologic malignancies. SYK activation leads to several downstream events that promote cell survival, including activation of phosphatidylinositol 3-kinase (PI3K) and AKT, and the phosphorylation of multiple signaling proteins (Buchner et al., 2009; Hahn et al., 2009; Prinos et al., 2011). In B-cells, it transduces tonic signaling by physical interaction with the immunoreceptor tyrosine-based activation motif (ITAM) of the B-cell antigen receptor (BCR) complex (Woyach et al., 2012), positively regulating survival and proliferation during development and immune response. SYK is also associated with the Fc receptor in B-cells, which instead has opposite effects to the BCR (Chen et al., 2008; Mócsai et al., 2010). The balance of regulation on survival and proliferation downstream of SYK is influenced by redox signaling: NADPH oxidase, in close proximity to BCR, can produce peroxide that inhibits phosphatase action on BCR-activated SYK, reinforcing tonic signaling (Reth, 2002). Another important function of SYK is response to oxidative stress where SYK gets activated and promotes pro-survival pathways (Takano et al.,

2002). B-cells die in response to SYK knock-down and fail to develop in SYK-deficient mice (Mócsai et al., 2010). Together, these observations have formed a rationale for SYK-targeted therapy in hematological malignancies with small molecule kinase inhibitors (Cheng et al., 2011; Friedberg et al., 2010; Hahn et al., 2009). SYK has not been studied in the context of lung neuroendocrine (NE) cells, the SCLC cells of origin, whose oxygen sensing functions, in analogy with BCR, rely on redox signaling (Buttigieg et al., 2012).

To our knowledge, SYK has not been proposed before as an oncogenic driver or candidate target in SCLC. Based on our WGCNA results, we investigated this possibility. We determined that 11 out of 33 SCLCs were SYK-positive by immunostaining in patient tissue microarrays (TMAs). Moreover, SYK knock-down reduced proliferation and survival in SYK-positive SCLC lines. We propose that SYK is an oncogenic driver in SCLC and that SYK expression may be developed as a companion biomarker for SYK targeted therapy.

Results

Identification of a SCLC-specific co-expression module

To identify a hierarchical network view of co-expressed genes across lung cancer subtypes, we applied WGCNA to a public dataset (GEO ID: GSE6044 – 33 untreated patients) comprised of 5 normal, 9 adenocarcinoma (ADC), 9 squamous cell carcinoma (SCC) and 9 SCLC lung cancer tissue specimens (Rohrbeck et al., 2008). An unsupervised correlation similarity matrix was built based on pairwise correlations between genes.

Unsupervised average linkage hierarchical clustering of all genes in this dataset resulted into 13 modules (Figure 3.1A) labeled by color and each comprised of mutually exclusive co-expressed genes. Genes with no distinct module assignment are grouped in a grey module by WGCNA. None of these modules were identified using any pre-assigned phenotype or gene bias. To ensure that modules were not being detected by chance, we simulated a random dataset containing same number of samples and genes as our test dataset. Only two modules were generated from the random dataset, turquoise and grey (with the grey module containing the vast majority of genes), indicating that WGCNA module identification in our test dataset is in fact driven by meaningful gene co-expression patterns (Figure 3.2).

Following the unsupervised module generation, individual gene correlations to a specific phenotype (normal lung, ADC, SCLC, SCC) were quantified by gene significance (GS). The average GS of all genes within each module is summarized in Figure 3.1B. This analysis unveiled positive or negative correlation of certain modules with specific lung cancer subtypes, or normal lung. The brown and purple modules appeared to be ADC specific, and included previously identified ADC markers cytochrome B5 (CYB5A) or surfactant protein B, C and D (SFTPB, SFTPC, SFTPD), respectively(Meyerson et al., 2004). Yellow, pink, orange and light cyan modules were SCC specific and included involvulin (IVL), cytokeratin 14 (KRT14), and galectin-7 (LGALS7) (Dakir et al., 2008; Levitt et al., 1990; Meyerson et al., 2004) (Appendix A - Download file).

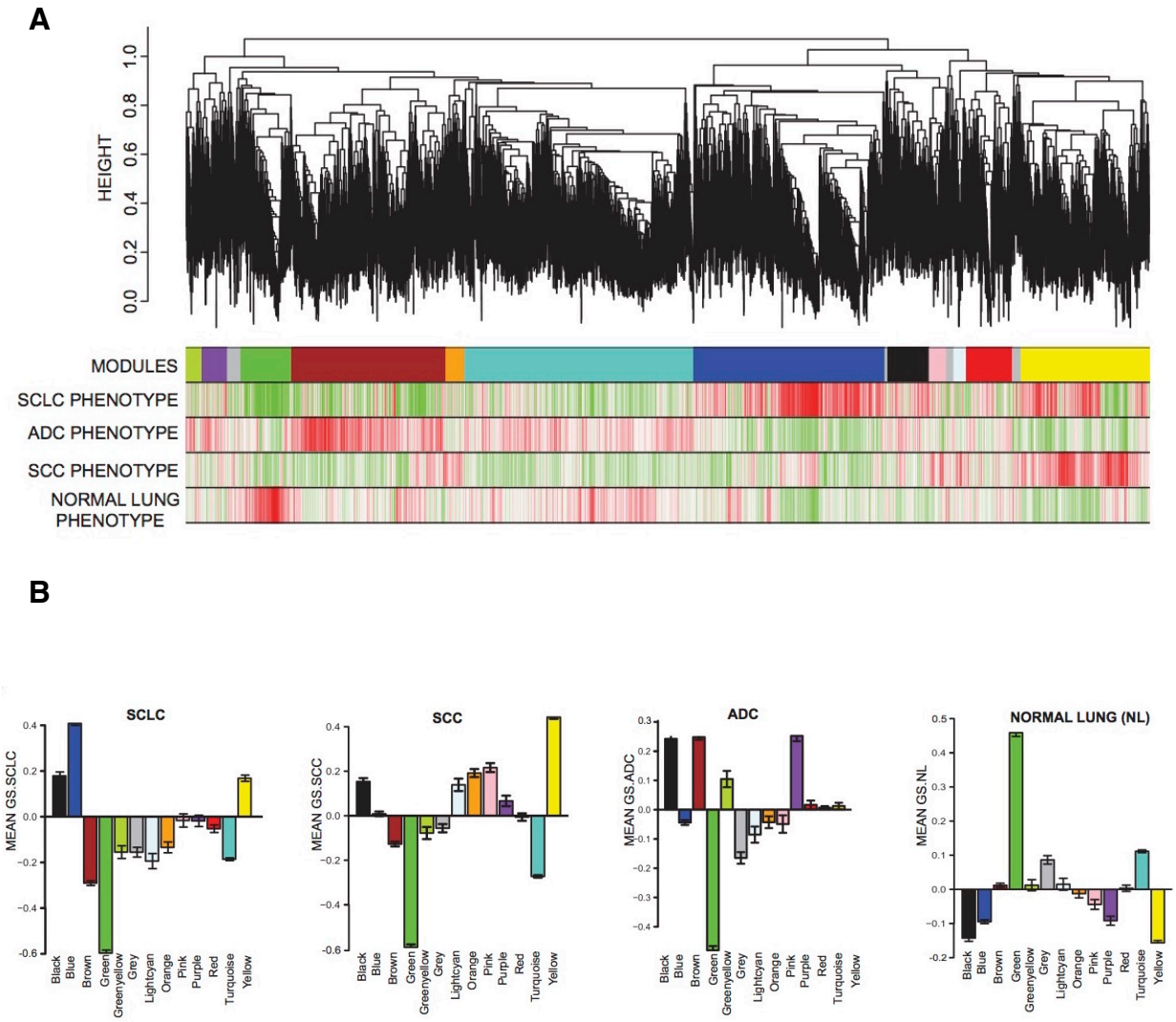
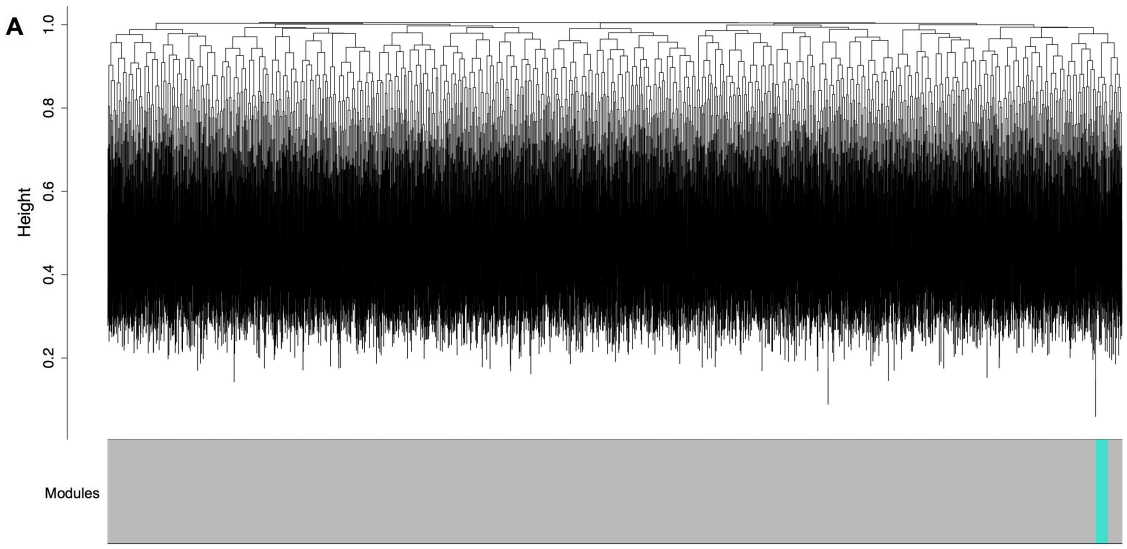


Figure 3.1: Identification of SCLC-specific modules using WGCNA.

(A) In the hierarchical dendrogram, lower branches correspond to higher co-expression (height = Euclidean distance). The 13 identified modules were coded by the colors indicated below the dendrogram. Below, red and green lines indicate positive or negative correlations, respectively, with lung tumor types on the left. (B) Average 'gene significance' (GS) of genes within a specific module summarized in the barplot for each lung tissue type (left to right: SCLC, SCC, ADC, and NL). The blue module is associated solely with SCLC.



B

Number of clusters/modules	Number of datasets
8	1
7	1
6	6
5	24
4	74
3	167
2	277
1	285
0	165

Figure 3.2: Absence of modules/clusters in a control WGCNA analysis of a simulated random dataset.

1000 random datasets were simulated in R to mirror the test dataset GSE6044 (8500 genes, 33 samples) (Rohrbeck et al., 2008), and was subjected to the exact analysis. (A) A representative dendrogram is shown (each line is a gene). Essentially all genes merged into the grey module, which is reserved by WGCNA to genes not assigned to any module. (B) Shows the number of random simulated datasets from the N=1000 that detected a certain number of modules. The overall p-value for this simulation analysis is less than 0.001, which is highly significant, indicating that our 13 modules detected in GSE6044 are meaningful and relevant to the biology of these tumors.

The green module contained genes positively correlated to the normal lung phenotype and negatively correlated with all tumor subtypes (SCLC, ADC, and SCC), making it a "normal lung module" (Appendix A -Download file). The blue module was specific to SCLC (Figure 3.1). Accordingly, it contained genes that have already been associated with SCLC progression such as Achaete-scute complex homolog 1 (ASCL1), Neural cell adhesion molecule 1 (NCAM1/CD56), Thyroid transcription factor-1 (TTF-1) and Insulinoma associated-1 (INSM1) (Arriola et al., 2008; Hiroshima et al., 2006) (Appendix A -Download file).

Identification and validation of a SCLC-specific hub network (SSHN) of co-expressed genes across genomic and proteomic platforms.

To identify and validate a network of co-expressed genes that is specific to SCLC, we focused on the blue module. The SCLC-specific blue module (1696 genes; Figure 3.1) is comprised of co-expressed up-regulated genes across SCLC tumors. Each module is arranged in the form of a hierarchical network (due to hierarchical clustering used to obtain the modules, Figure 3.1A dendrogram). Therefore, each module consists of a few highly connected "hubs" (genes that have high intramodular connectivity kME) as well as many genes with fewer connections. The rationale behind building hub-based networks is to narrow down the list of relevant candidates, based on the assumption that highly connected hubs are more vulnerable targets to alter network performance. This assumption has been successful in several examples from biological networks in yeast (He and Zhang, 2006; Jeong et al., 2001) and mammalian cells, including cancer (Dutta et al., 2012; Horváth et al., 2006).

Each module can be further filtered to identify the top hubs relative to desired criteria using measures such as intramodular connectivity (kME) and gene significance (GS) (Langfelder and Horvath, 2008). We filtered the blue module genes to obtain hubs that ranked high in each of the following criteria: a) high positive correlation with SCLC phenotype given by gene significance (GS.SCLC >0.5); b) high intramodular connectivity (blue module kME >0.5); and c) high T-test statistic (overexpression in SCLC versus normal lung > 5) and a p-value less than 0.01. This filtering approach produced 287 hub genes, which are not only overexpressed in SCLC, but also highly connected within SCLC. We refer to this network of 287 hubs as SCLC-specific hub network (SSHN) (Appendix B-Download file).

To validate the robustness of SSHN as a SCLC-specific classifier, it was first applied by unsupervised hierarchical clustering bootstrap analysis to patient samples in a test public dataset (GSE6044) from which the blue module was derived. The SSHN classified SCLC away from every other lung tumor subtype (ADC and SCC) and normal lung, the area under ROC curve (AUC) was 0.87 with 95% confidence interval (CI) of [0.72, 1] (Figure 3.3A). The performance of the SSHN classifier was reproducible in both an independent validation patient dataset of 163 tumors (GSE11969) (Takeuchi et al., 2006) generated in a different array platform (Agilent) (AUC of 1) (Figure 3.4A), as well as in our own microarray dataset containing 23 SCC and 10 SCLC samples (AUC of 0.94 with 95% CI of [0.85, 1])(Figure 3.4B). In the GSE11969 dataset, the SSHN also proved to be an excellent classifier for distinguishing SCLC from large cell carcinoma (LCC) subtype (Figure 3.4A). Interestingly, large cell neuroendocrine carcinomas

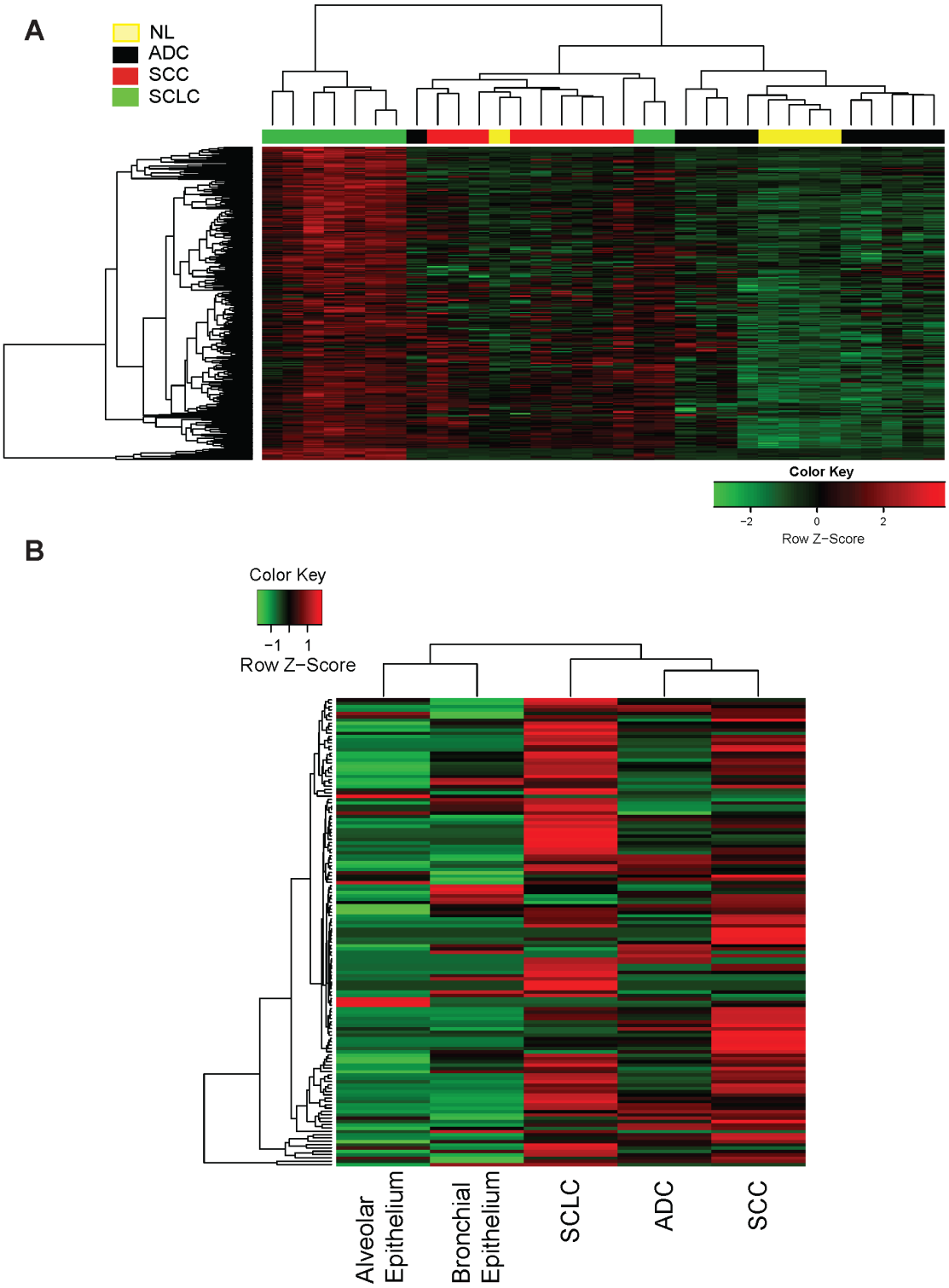


Figure 3.3: Validation of SSHN as a robust classifier for SCLC in two independent datasets from (A) high-throughput gene expression and (B) shotgun proteomic analysis.

Figure 3.3: Validation of SSHN as a robust classifier for SCLC in two independent datasets from (A) high-throughput gene expression and (B) shotgun proteomic analysis.

(A) Unsupervised clustering heatmap based on 287 SSHN genes (rows) of lung cancer patients (columns) in GSE6044 dataset. Red and green indicate high and low expression, respectively. The majority of SCLCs cluster by themselves on the far left of the dendrogram. Two SCLC specimens are excluded from this cluster, a trend to be investigated in more depth if confirmed in larger datasets (see Discussion). (B) SSHN-based unsupervised clustering heatmap of an in-house generated shotgun proteomic dataset comprised of control alveolar and bronchial epithelium, ADC, SCC and SCLC tissue specimens (for each tissue type, specimens from multiple patients, five in this case, were pooled as it is customary for shotgun proteomic analysis). Red and green as denoted in (A). Analysis is limited to 141 out of 287 SSHN proteins (rows), since the remainder proteins were not detect by shotgun proteomics. The 3 tumor specimens segregate together from normal tissue. Within the 3 tumor specimens, ADC and SCC are more similar to each other than to SCLC.

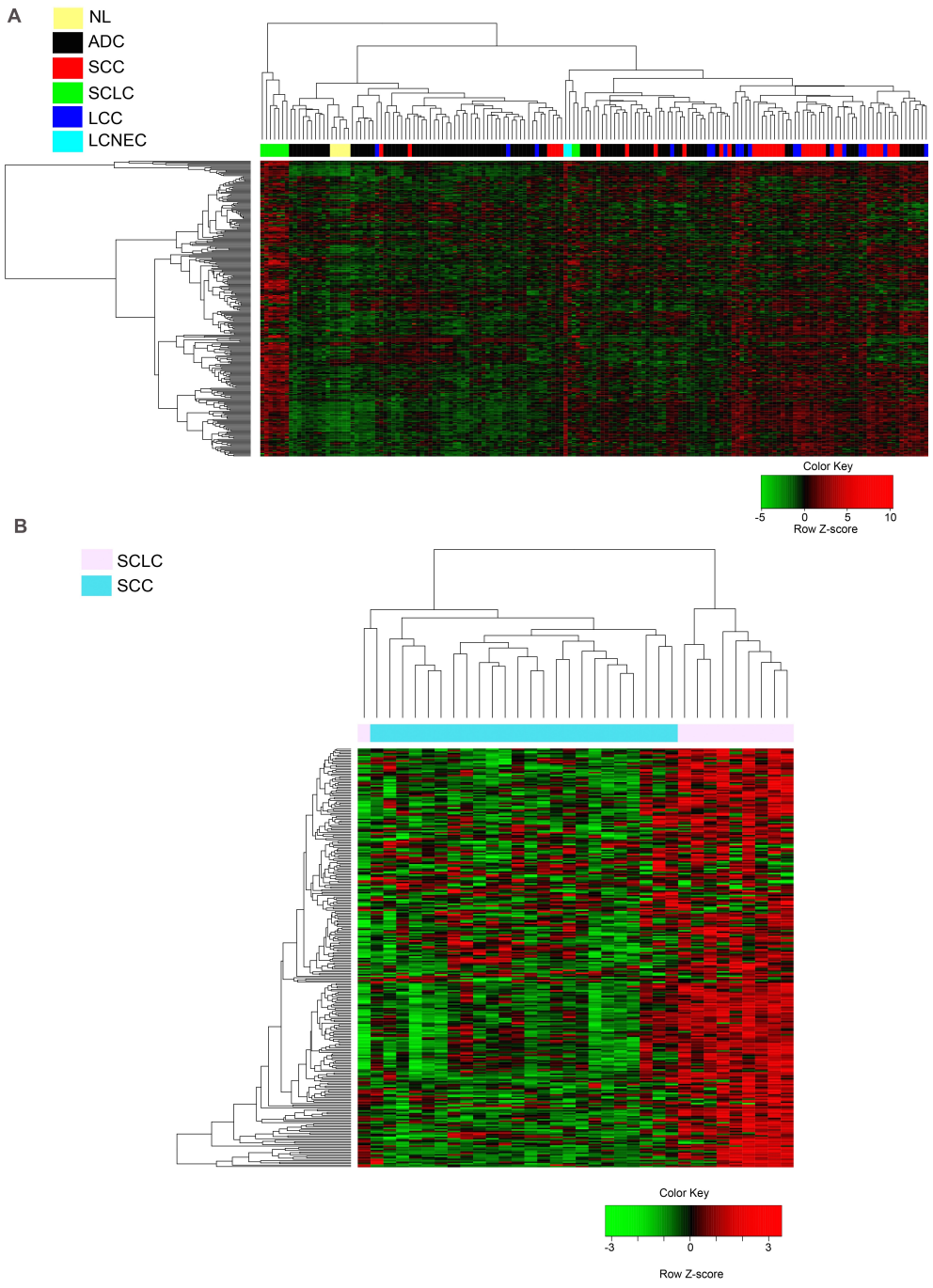


Figure 3.4: SSHN as a reproducible classifier in GSE11969 and in-house Agilent datasets.

Unsupervised clustering heatmap based on SSHN genes (rows) of (A) 163 lung cancer patients (columns) in GSE11969 dataset (Takeuchi et al., 2006), and (B) our own Agilent microarray dataset containing 23 SCC and 10 SCLC samples. Red and green colors in rows of the heatmap indicate high and low expression respectively. LCC- large cell lung carcinoma, LCNEC- large cell neuroendocrine carcinoma.

(LCNC), another high-grade neuroendocrine tumor (NET) of the lung, co-clustered with SCLC, confirming similarities between the 2 tumor types as reported previously (Jones et al., 2004). On all the three patient datasets, the SSHN genes are highly predictive of SCLC against other tissue types with statistically significant p-values less than 0.0001.

To further validate the SSHN as a classifier, we used next-generation sequencing to produce genome-wide RNA-seq data on an independent set of tissues including 10 SCLCs, 5 SCCs, and 5 normal lung tissue specimens. We detected overexpression of 206 genes out of 287 SSHN genes that differentiate SCLC (71.8%) from normal lung alone (at 5% FDR) while 106 genes out of 287 SSHN genes differentiate SCLC (71.8%) from normal lung and SQCC (at 5% FDR) (Appendix B-Download file), indicating that SSHN is a robust classifier in another data type (RNA-seq).

Finally, the SSHN gene expression classifier was further validated at the protein level in yet another in-house, independent set of formalin fixed paraffin embedded patient tissue samples analyzed by shotgun proteomics and comprised of 5 samples each of SCLC, SCC, ADC and age- and smoking history-matched normal lung tissues specimens, pooled by histologic type. Out of 287 SSHN genes, 141 gene products were detected at the proteomic level and also classified the SCLCs apart from the other tissues (Figure 3.3B). To our knowledge, this is a first report of an entire SCLC genomic signature validated at the proteomic level.

In each of the 4 datasets, there were 1-2 specimens that did not segregate with the SSHN-defined SCLC cluster, but were clinically diagnosed as SCLC (Figure 3.3A; Figure 3.4). This could be due to mis-diagnosis as is fairly common in SCLC due to mixed SCLC-NSCLC histology (Nicholson et al., 2002), or possibly a small subset of

patients whose tumors have different biology. Overall, we conclude that the SSHN is a robust molecular classifier to distinguish SCLC from other lung tumor types and normal lung across multiple gene and protein expression platforms.

Biological insights from the SSHN: Network enrichment analysis and target identification

To gain biological insights in SCLC biology, the SSHN component genes were further categorized into functional pathways based on the assumption that they are co-upregulated because of shared cellular functions. Analysis of SSHN by Webgestalt (Zhang et al., 2005) revealed that SSHN is enriched for functional pathways summarized in Appendix C (Download file) and D (Download file) and Figure 3.5, and include cell cycle and checkpoint response (total of 25 genes), cellular stress response (41 genes of which 21 genes related to oxidative stress), and DNA damage response and repair pathways. All p-values were adjusted for multiple comparisons in Webgestalt and therefore effectively rank the significance of these functional pathways in SCLC phenotype.

As a proof-of-concept that connected hubs identified by WGCNA are of biological relevance, we further refined the pathway analysis by focusing on kinases, since these tend to be of the greatest translational value. There were 20 kinases contained in the SSHN (Table 3.1), all worth investigating in the context of SCLC. However, shotgun proteomics data (available for 4 kinases, Table 3.1) indicated that SYK is strongly overexpressed within the SCLC phenotype compared to normal tissue (high “SCLC vs. Bronchial epithelium Rate ratio” and “SCLC vs. Alveolar epithelium Rate ratio”).

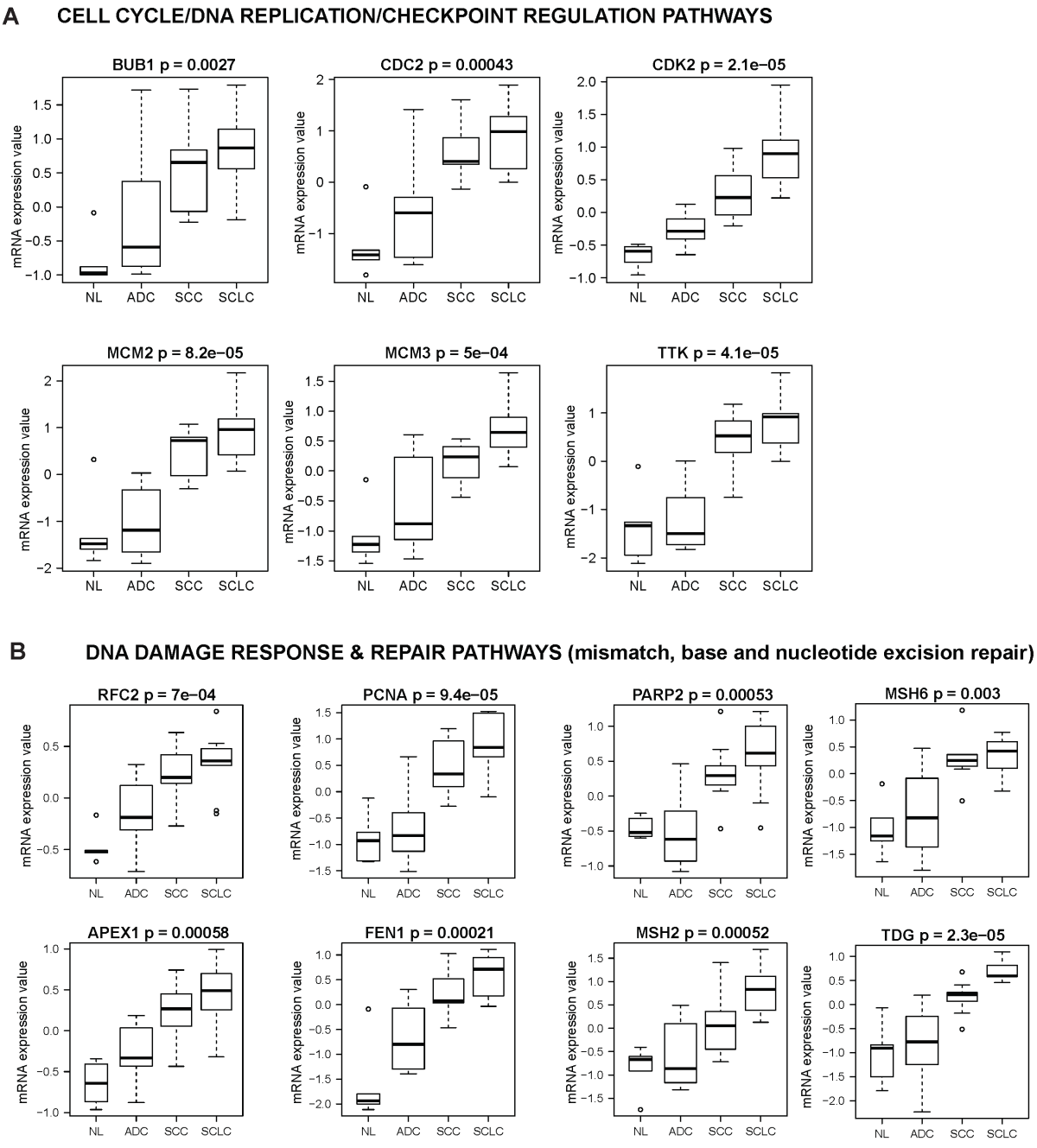


Figure 3.5: mRNA expression of SSHN genes for the top representative canonical pathways from network enrichment analysis.

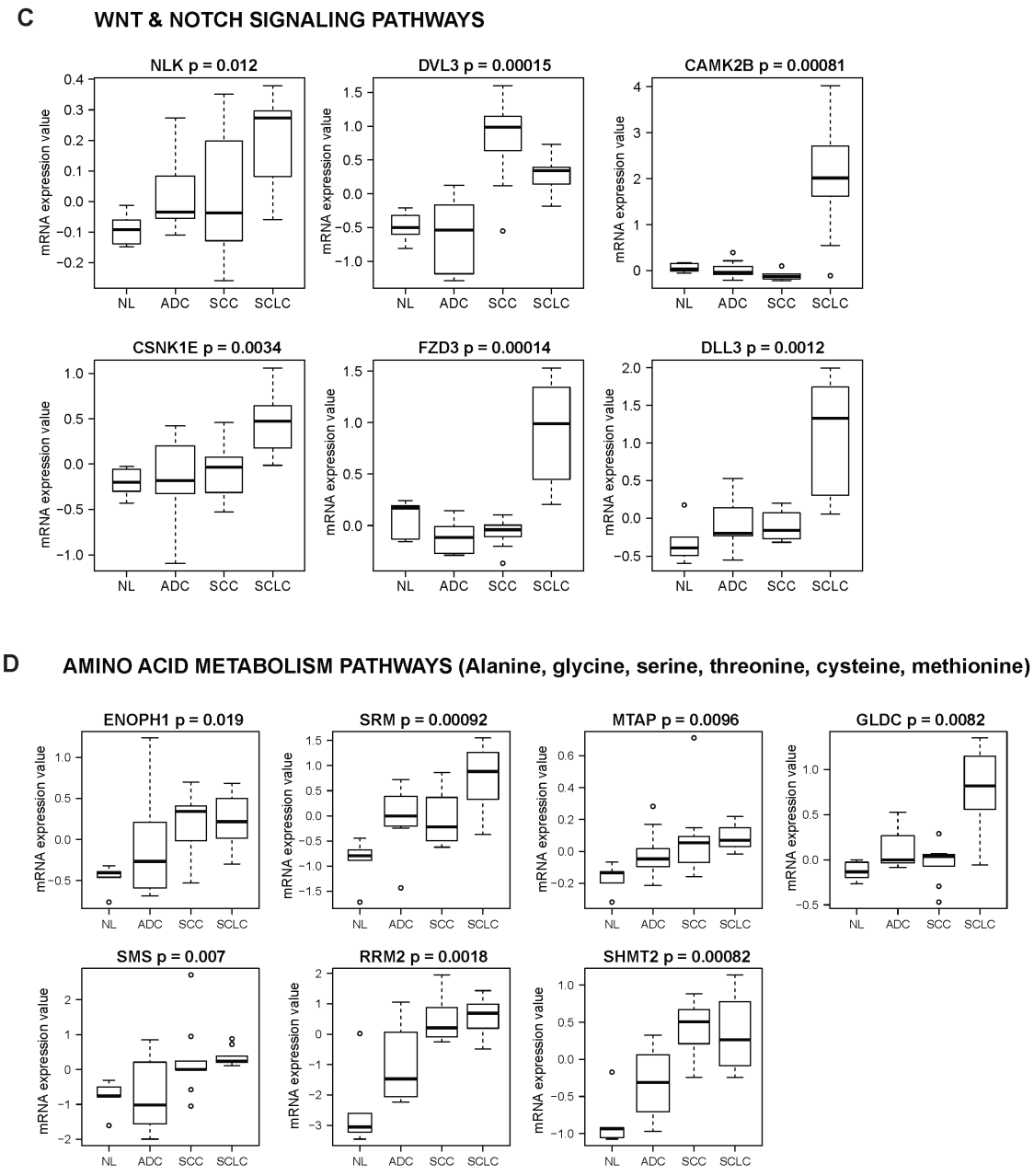


Figure 3.5 continued: mRNA expression of SSHN genes for the top representative canonical pathways from network enrichment analysis.

Functional enrichment analysis was carried out using Webgestalt [33]. Boxplots of mRNA expression of representative SSHN hubs functioning in various pathways (A) Cell cycle checkpoint control and DNA replication; (B) DNA damage response and repair; (C) Wnt and Notch signaling pathways (D) Amino acid metabolism pathways. The outliers are denoted by dots. P-value shows statistical significance by Kruskal-Wallis nonparametric test (Kruskal and Wallis, 1952).

Gene information			WGCNA metrics for intramodular connecitivity, Gene significance and T-test Statistic						Shotgun proteomic data						RNAseq data (25-bp)			
ID	Location	Type(s)	MODULE	KMEBLUE	GS.SCLC	T-TEST STATISTIC-GENEXP	p-value	FDR	SCLCvsBronchial epithelium Rate ratio	Poisson_Pvalue	Poisson_FDR	SCLCvs Alveolar epithelium Rate ratio	Poisson_Pvalue2	Poisson_FDR3	Mean (SCLC)	Mean (Normal)	P-value (SCLC)	FDR adjusted P-value (SCLC)
CDK4	Nucleus	kinase	yellow	0.40	0.88	6.70	8.90E-05	0.003	-0.05	1	1	-0.75	1	1	208.02	26.2824	1.90302E-07	5.46166E-05
DTYMK	Cytoplasm	kinase	blue	0.73	0.84	7.30	4.44E-05	0.003	1.07	0.1197	0.2084	2.96	0.0085	0.0266	3.44113	0.954753	0.0047932	0.00893278
SYK	Cytoplasm	kinase	blue	0.64	0.83	4.67	2.66E-03	0.017	26.79	0.0998	0.1781	26.10	0.1712	0.239	0.770973	0.285704	0.00696815	0.011844
CDC2	Nucleus	kinase	blue	0.62	0.84	5.43	4.89E-04	0.007	28.24	0.0998	0.1781	27.54	0.1712	0.239	#N/A	#N/A	#N/A	#N/A
CDK2	Nucleus	kinase	blue	0.81	0.89	8.28	4.02E-06	0.001	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	5.42382	0.444168	0.000239147	0.00132869
CDC7	Nucleus	kinase	blue	0.92	0.84	6.35	3.68E-05	0.002	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	1.00922	0.0830575	0.00057998	0.00211053
FYN	Plasma Membrane	kinase	blue	0.77	0.81	6.10	8.30E-05	0.003	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	10.7748	0.591148	0.000665613	0.00231662
BUB1	Nucleus	kinase	blue	0.53	0.84	6.09	7.44E-05	0.003	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	36.8445	1.1555	0.000716467	0.00241259
PBK	Cytoplasm	kinase	blue	0.63	0.87	5.74	6.20E-04	0.008	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	44.3023	0.984071	0.000343411	0.00164265
NEK2	Cytoplasm	kinase	blue	0.67	0.85	5.52	6.24E-04	0.008	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	15.8659	1.47427	3.44664E-05	0.000989186
TTK	Nucleus	kinase	blue	0.70	0.87	5.39	1.50E-03	0.013	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	0.926709	0.0409683	0.000208963	0.00132869
STK25	Cytoplasm	kinase	blue	0.74	0.75	5.02	4.02E-04	0.007	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	91.1644	31.0777	0.00588725	0.0103659
CDK5R1	Nucleus	kinase	blue	0.89	0.74	5.01	6.90E-04	0.008	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	20.5632	1.13017	0.000060467	0.00103531
NLK	Nucleus	kinase	blue	0.54	0.75	5.01	4.09E-04	0.007	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	18.2468	8.33942	0.0368493	0.0457825
CSNK1E	Cytoplasm	kinase	blue	0.69	0.76	4.87	3.87E-04	0.006	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	2.83195	1.25658	0.0168093	0.0236484
PRKD3	unknown	kinase	blue	0.53	0.76	4.74	5.09E-04	0.007	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	18.6005	4.31658	0.00104703	0.00294606
BRD3	Nucleus	kinase	blue	0.77	0.74	4.64	5.84E-04	0.008	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	24.3405	5.1199	0.00146226	0.00374705
BUB1B	Nucleus	kinase	blue	0.61	0.81	4.63	2.15E-03	0.015	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	33.0484	0.963844	0.000290493	0.00146266
CAMK2B	Cytoplasm	kinase	blue	0.87	0.70	4.61	1.64E-03	0.013	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	1.51183	0.0328172	0.0896061	0.105397
TLK2	Cytoplasm	kinase	blue	0.67	0.76	4.59	6.89E-04	0.008	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	0.315878	0.602432	0.452278	0.493551

Table 3.1. Kinase hubs of SSHN.

This table shows expression values for twenty kinase genes (identified by WGCNA analysis) enriched in SSHN across various datasets. RNAseq data shows comparisons of differential expression of SCLC versus normal lung and associated statistics such as p-value and False discovery rates (FDR). Shotgun proteomic data denotes the comparison of rate ratios (obtained from Shotgun data, see Materials and methods) of SCLC versus normal bronchiolar epithelium and SCLC versus normal alveolar epithelium.

SYK is an oncogenic non-receptor tyrosine kinase involved in hematologic malignancies (Hahn et al., 2009; Cheng et al., 2011; Friedberg et al., 2010). Another oncogene, the SRC-family kinase FYN, was also part of this SSHN kinase set. SYK is an intracellular signal transducer downstream of growth factor/T-cell/B-cell receptors well known to work in concert with SRC-family kinases (Mócsai et al., 2010). Specific overexpression of SYK and FYN in SCLC, compared to other lung tumor types, has not been previously reported, to the best of our knowledge (Figure 3.6). Together, these clues prompted us to select SYK and FYN for further investigation in the context of SCLC tumors.

To verify co-expression at the protein level, we immunostained for SYK and FYN in a panel of SCLCs assembled in tissue microarrays (TMAs). All specimens were tested in duplicate, and the expression of SYK and FYN consistently co-varied (Figure 3.7A), with a correlation of 0.28 across SCLC specimens. Clustering analysis of the staining scores of SYK/FYN expression separated the TMA specimens into 2 groups, SYK/FYN-positive and -negative tumors (Figure 3.7B).

Preservation of SSHN and differential SYK/FYN expression in SCLC cell lines

SYK and FYN are attractive candidates for targeted therapy (Riccaboni et al., 2010; Saito et al., 2010). To test their functional relevance in SCLC, we turned to SCLC cultured cell lines. The SSHN classifier was conserved in a large panel of lung cell lines (Lockwood et al., 2008). As indicated by clustering analysis (Figure 3.8A), 21 out of 23 SCLC cell lines separated nicely from the other 36 lung cancer cell lines tested (AUC of 0.97 with 95% CI of [0.94, 1]). Note that 2 SCLC cell lines did not follow this pattern, an

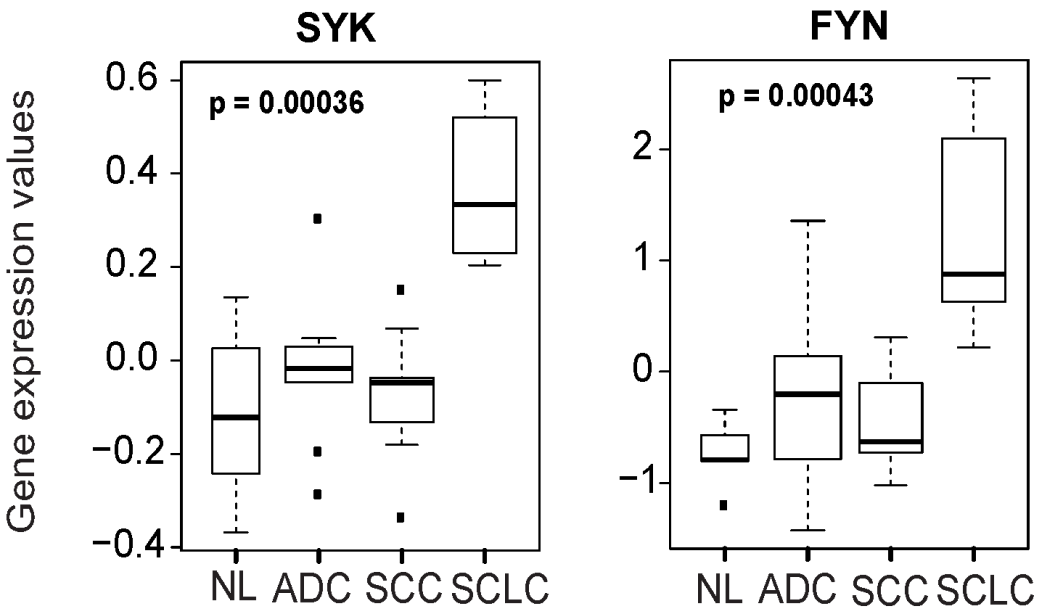


Figure 3.6: Co-expression of 2 SSHN kinases FYN and SYK in SCLC patients.

Log2 expression values are indicated in the boxplots for each individual hub within SSHN across various patient lung tissues from the GSE6044 test dataset (Rohrbeck et al., 2008). The outliers are denoted by dots. P-value shows statistical significance by Kruskal-Wallis nonparametric test (Kruskal and Wallis, 1952). FYN and SYK are co-overexpressed in SCLC patients versus NSCLC (ADC, SCC) and normal lung.

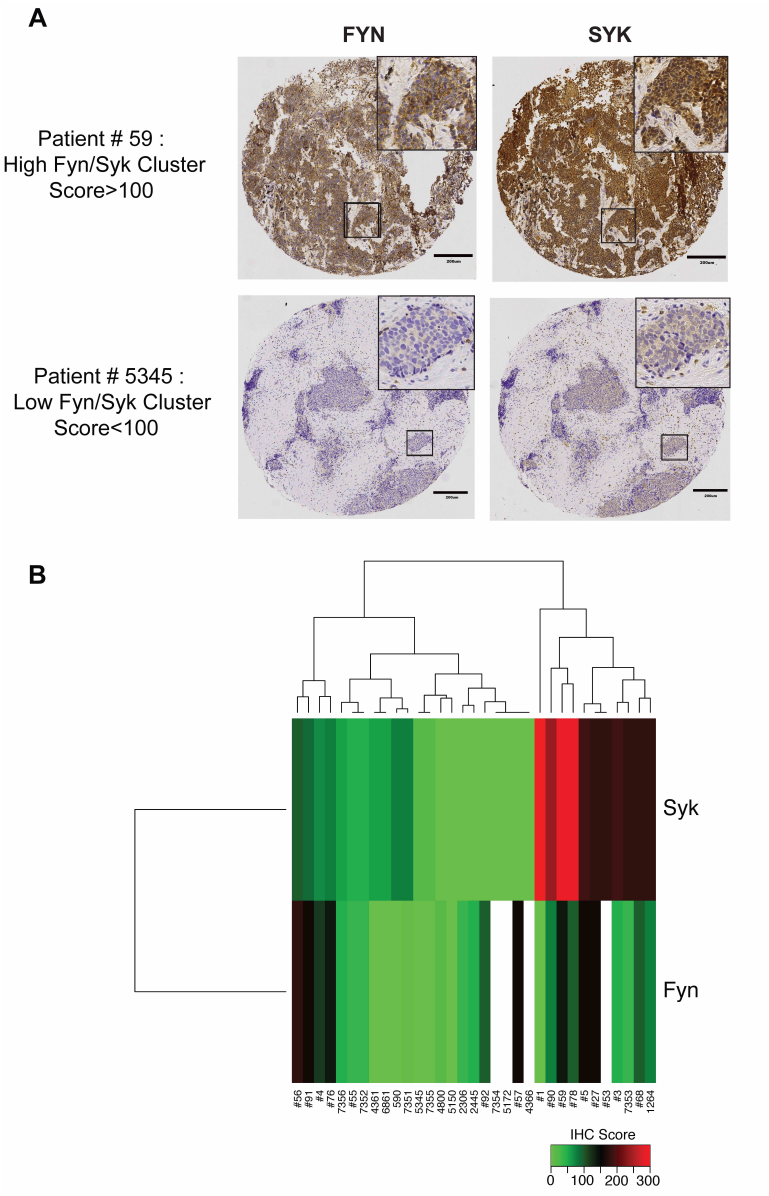


Figure 3.7: Co-expression of SYK and FYN in a subset of SCLC tumors.

Contiguous sections of TMAs from 39 SCLC patient specimens were stained with antibodies to SYK and FYN, respectively. Stained sections were scored by a pathologist as described in Methods. (A) Representative stained sections showing positive (upper) or negative (lower) results. See text for additional details. Tumor spot images were captured by brightfield microscopy at 20X magnification. (B) Unsupervised hierarchical clustering heatmap of SYK and FYN immunostaining intensity scores across SCLC patients distinguished positive from negative tumors as described in Methods. Red and green indicate high and low expression, respectively. Specimens that are positive for both SYK and FYN segregate in one cluster, on the right. Patient ID shown below the heatmap.

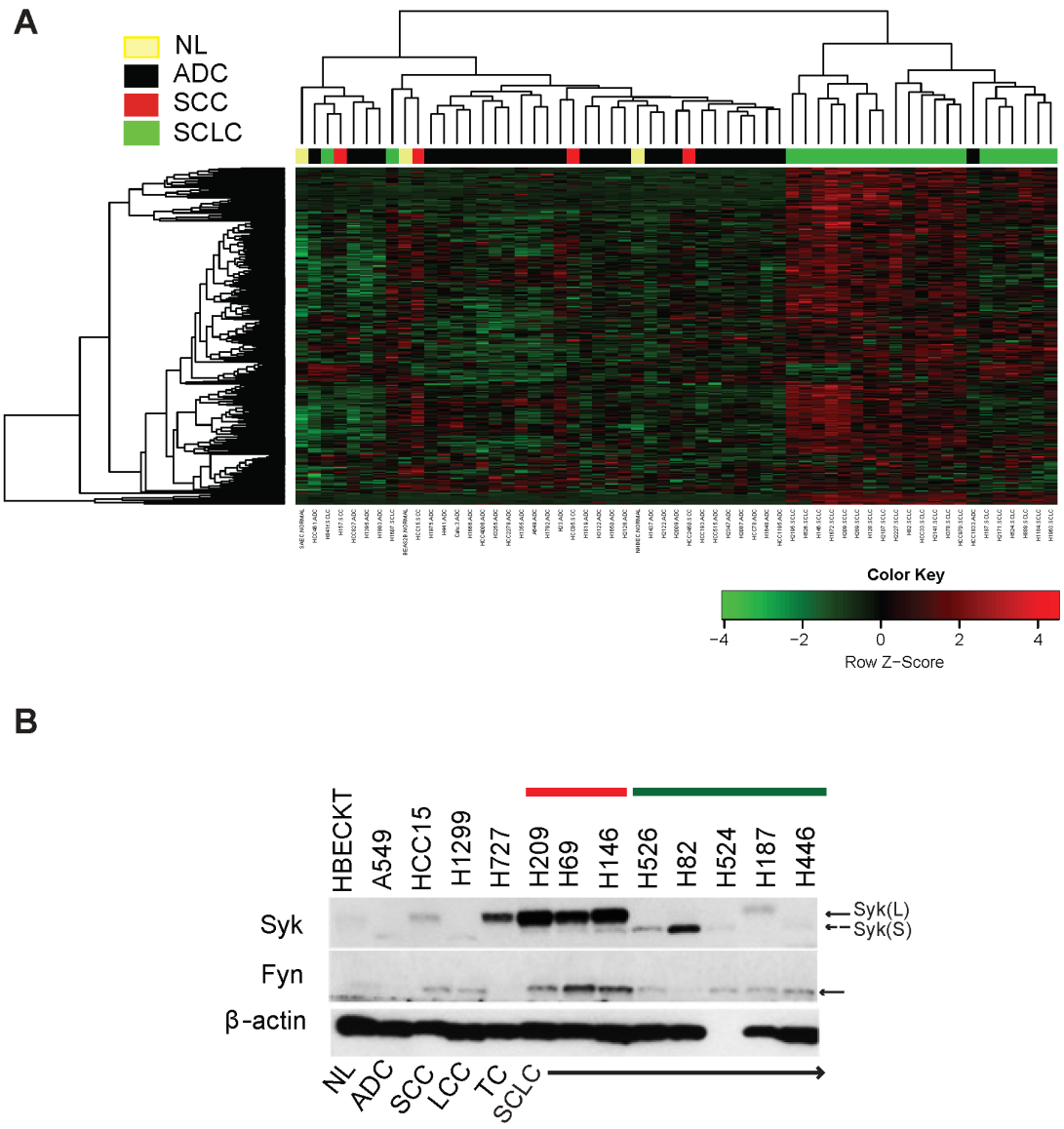


Figure 3.8: SSHN is preserved in SCLC cell lines.

(A) Unsupervised clustering heatmap based on SSHN genes (rows) of lung cancer cell lines (columns) in GSE4824 dataset [36]. Red and green colors in rows of the heatmap indicate high and low expression respectively. This analysis shows SSHN conservation across SCLC cell lines. (B) Representative western blot of SYK and FYN in various lung cancer cell lines. FYN and SYK are selectively overexpressed in SCLC cell lines. Within the SCLC cell lines, the red and green bars indicate FYN/SYK-positive and -negative SCLC cell lines, respectively. Arrows point to bands corresponding to the expected molecular weight for SYK and FYN. The dotted arrow indicate the position of a shorter form of SYK protein (SYKB or S) that lacks 23 amino acids (Sada et al., 2001).

observation mirrored in tumor specimens (Figure 3.3; Figure 3.4) that warrants further studies.

We investigated co-expression of SYK and FYN in SCLC cell lines by western blotting of whole-cell lysates with appropriate antibodies (Figure 3.8B). Similar to our protein expression shown by immunostaining of our TMAs, SYK and FYN exhibited a trend to co-vary in SCLC cell lines (Figure 3.8B), opening an avenue to biochemical analyses of the functional value of this differential expression. Note that SYK has two splice-variant isoforms - long (L or $p72^{SYK}$) and short (S or B) that lacks 23 amino acids (Sada et al., 2001). The SYK positive cell lines overexpress SYK (L) form while other cell lines express low or no SYK (S) (Figure 3.8B).

Inhibiting SCLC cell line viability by SYK knock-down

To assess the validity of SYK and/or FYN as targets in SCLC, we down-regulated the expression of these proteins using siRNA in the H69 and H146 cell lines (Figure 3.8B). siRNA induced 80-90 percent reduction in total protein expression for each of these molecules in both H69 and H146 (Figure 3.9A and D; Figure 3.10A and D). We assessed viability with automated microscopy, imaging-based methods (Live-dead assay, see Methods; images and segmentation for obtaining cell counts shown in Figure 3.11). SYK knock-down caused a significant decrease in proliferation rates compared to scrambled control in both H69 and H146 (Figure 3.9B and E), while FYN knock-down showed little effect (Figure 3.10B and E). The decrease in proliferation was in part due to a loss of cell viability, as indicated by increased cell death by Day 5 in SYK knock-down cells assessed by ethidium homodimer positivity (Figure 3.9C and F;

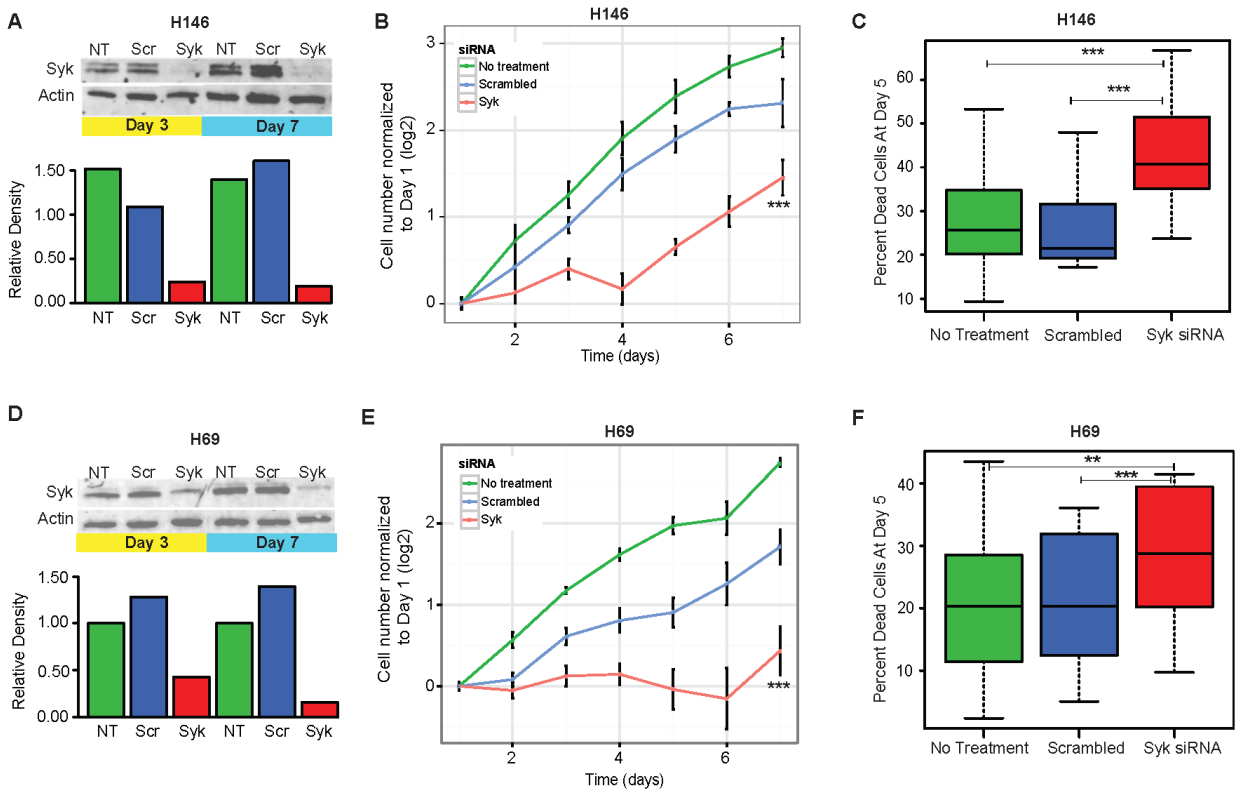


Figure 3.9: Effect of Syk knock-down in Syk/Fyn positive SCLC cell lines.

The SCLC cell lines H146 (A-C) and H69 (D-F) were treated with Syk-specific and control siRNA as described in Materials and Methods section. (A, D) The efficiency of inhibition was measured by Western blotting on day 3 and 7 post transfection. Syk resolves as two bands, of which the lower is a less-functional splice variant that lacks 23 amino acids [37]. Band intensity (lower panels) was quantified by densitometry in ImageJ. (B, E) Cell proliferation, measured by cell counts as described in Materials and Methods section, shows that Syk-siRNA treatment induces statistically significant growth inhibition compared to untreated cells and to scrambled siRNA treatment. Asterisks denote overall statistical significance of slope as compared to control across the siRNA conditions, as follows: <0.0005 '***' 0.001 '**' 0.01 '*' 0.05 '.'. The viability growth curves (from N=4 experiments) statistics were generated from slopes of a linear regression model. Multiple comparison of treatments were derived using ANOVA and Tukey's method [80]. (C and F) Percentage of dead cells (percent of ethidium homodimer positive cells normalized to total cell counts, see Materials and Methods) is significantly higher (H69 p-value < 2.2e-16; H146 p-value < 2.2e-16) in Syk siRNA treated cells at day 5, compared to controls. Asterisks denote statistical significance measured by paired t-test as compared to control across the siRNA conditions, as follows: <0.0005 '***' 0.001 '**' 0.01 '*' 0.05 '.'.

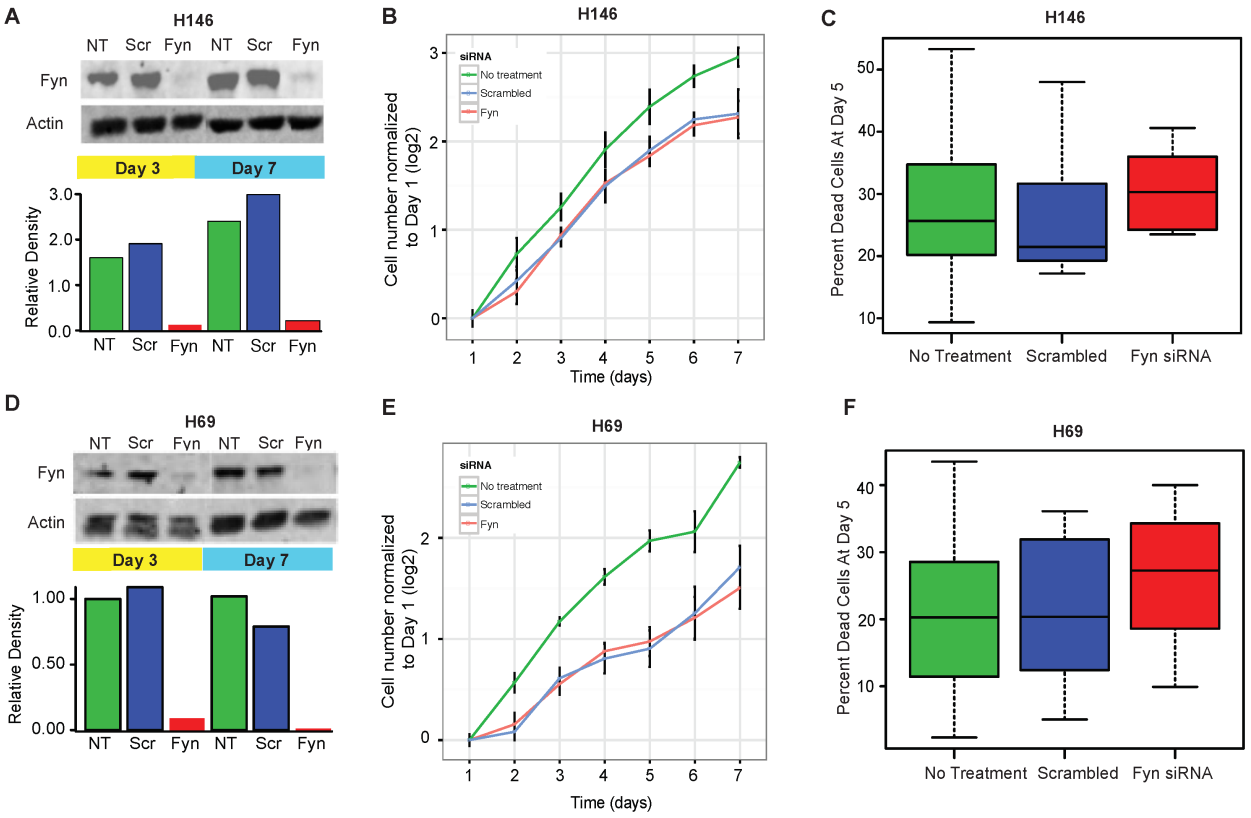


Figure 3.10: Fyn KD has no effect on Fyn and Syk positive SCLC cell lines.

The SCLC cell lines H146 (A-C) and H69 (D-F) were treated with Syk-specific and control siRNA as described in Materials and Methods section. (A, D) The efficiency of Fyn siRNA inhibition was measured by Western blotting on day 3 and 7 post transfection. Band intensity (lower panels) was quantified by densitometry in ImageJ. (B, E) Cell proliferation, measured by cell counts as described in Materials and Methods section, shows that Fyn-siRNA treatment shows no growth inhibition compared to untreated cells and to scrambled siRNA treatment. Asterisks denote overall statistical significance of the slope as compared to control across the siRNA conditions, as follows: <0.0005 '***' 0.001 '**' 0.01 '*' 0.05 '.'. The viability growth curves (from N=4 experiments) statistics were generated from slopes of a linear regression model. Multiple comparison of treatments were derived using ANOVA and Tukey's method (Kuehl, 2000). (C and F) Percentage of dead cells (percent of ethidium homodimer positive cells normalized to total cell counts, see Materials and Methods) in Fyn siRNA treated cells at day 5, compared to controls. Asterisks denote statistical significance measured by paired t-test as compared to control across the siRNA conditions, as follows: <0.0005 '***' 0.001 '**' 0.01 '*' 0.05 '.'.

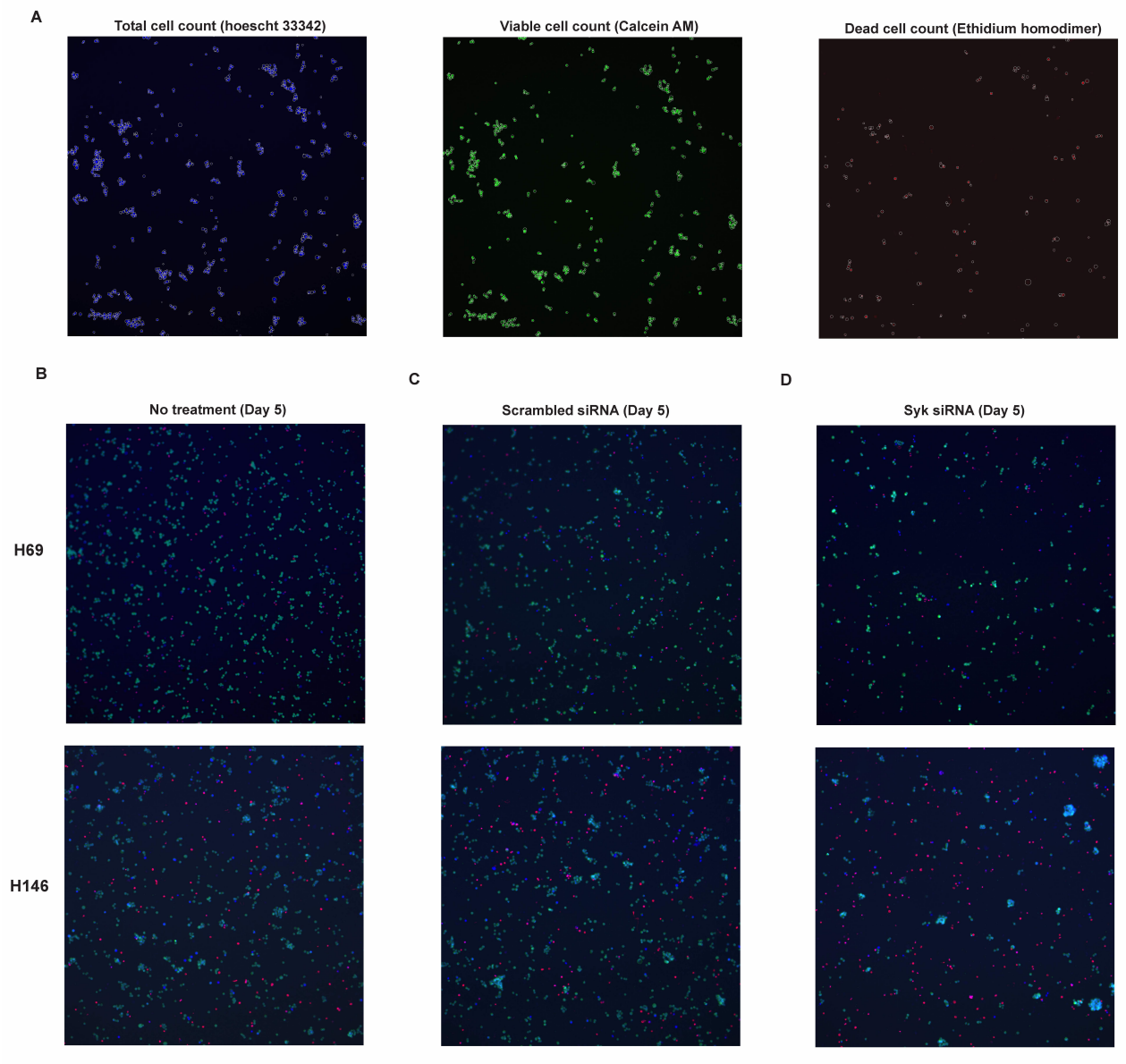


Figure 3.11: Viability assay measurements using Cellavista high-throughput imaging microscope. (A) Individual cell populations and segmentation performed by Cellavista Roche viability kit algorithm. The colors denote the different dyes used for measurement of total cell count (blue, Hoescht 33342 – left image), viable cell count (green, calcein AM – center) and dead cell count (red, ethidium homodimer - right). Representative viability assay images of H146 (top panel) and H69 (bottom panel) - (B) No treatment, (C) Scrambled and (D) SYK siRNA. SYK knock-down decreases cellular viability via increased death in both H69 and H146.

Figure 3.10C and F). Together, these data suggest that SYK is a candidate therapeutic in SYK/FYN-expressing SCLCs.

Discussion

We report several findings of immediate translational value for SCLC: 1) derivation of an SCLC-specific hub network (SSHN) that classifies SCLC from other lung cancers, including the closely related neuroendocrine tumors; 2) validation of the SSHN classifier across many data types, including expression microarrays from multiple platforms, RNAseq and shotgun proteomics; 3) co-varied expression of 2 oncogenes, SYK and FYN, in a subset of SCLC tumors and cell lines; and 4) identification of SYK as a candidate biomarker and therapeutic target for SCLC.

The increasing availability of large gene expression cancer datasets presents unprecedented opportunities for translational advances. Challenges in data analytics, however, must be met. For instance, the predominant metric of differential gene expression is silent on disease relevance of identified gene products, since it provides no measure of their functional relatedness (Khatri et al., 2012), and its resulting signatures do not replicate well across datasets (Shi et al., 2010; Subramanian and Simon, 2010). The number of potential therapeutic targets (e.g., ranked by differential expression scores) is large and expanding, but target prioritization is hampered by lack of functional insight. In contrast, analyses based on gene co-expression algorithms perform well across data types (Shi et al., 2010) and inspire working hypotheses since their results resemble hierarchical signaling networks. Accordingly, the SCLC-specific co-expressed gene classifier network SSHN we report here is robust across datasets

encompassing different types of lung cancer (Figure 3.3 and 3.4). In particular, despite being derived from gene expression microarray data, the SSHN performed well on proteomic lung cancer specimens. Note that each of the datasets tested were obtained from independent SCLC patient cohorts. To our knowledge, this is the first report of signature preservation on a shotgun proteomic SCLC dataset. Other co-expression based approaches have also been successfully applied in other cancers such as breast cancer (Shi et al., 2010).

Neuroendocrine lung tumors, to which SCLC belong, are sometimes difficult to sort out based solely on the current World Health Organization (WHO) criteria of morphology and mitotic rate, warranting searches for additional biomarkers (Nicholson et al., 2002; Rekhtman, 2010; Renshaw et al., 2005). The SSHN signature begins to address this need, e.g., distinguishing SCLC that stain negative for neuroendocrine markers such as synaptophysin and chromogranin A (~25%) (Hiroshima et al., 2006; Nicholson et al., 2002) from NSCLC, and mixed SCLC-NSCLC from NSCLC. However, because of the very small number of LCNEC samples studied by gene expression analysis, we cannot exclude the possibility that other LCNEC tumors would co-cluster with SCLC. In addition, due to the lack of larger SCLC datasets and the limited clinical information on the available SCLC datasets, careful validation of our results, including outcome associations, is definitely warranted.

While SSHN as a whole is an effective SCLC classifier, its individual component genes (or gene products) may or may not be expressed in a particular tumor. This is not at all surprising, due to the expected inter-tumor heterogeneity within a particular histological type (Nicholson et al., 2002). Our data suggest that within the SCLC cluster

defined by SSHN, a further subdivision between SYK/FYN-positive and –negative may be informative. A few specimens classified as SCLC by pathological and clinical criteria, did not cluster with SSHN-defined SCLC (Figure 3.3 and 3.4A). Whether these are misdiagnosed or represent disease heterogeneity or different stage of tumor progression remains to be tested.

Receptor and non-receptor tyrosine and serine-threonine kinases are effective actionable targets in cancer. SSHN contains twenty kinases and growth factor receptors, including TTK, TLK2, NEK2, CDK4, FYN, PLCG1, SYK (Table 3.1). None of these were previously reported in SCLC; thus, prioritization strategies are called for. The kinases SYK and FYN stand out as potential SCLC targets for several reasons. Besides being tightly associated with the SCLC phenotype, they are already proven as candidate targets in other cancers, such as CML (Ban et al., 2008; Buchner et al., 2009), AML (Hahn et al., 2009), retinoblastoma (Zhang et al., 2012), glioblastoma (Lu et al., 2009) and prostate cancer (Cai et al., 2011; Posadas et al., 2009). They also activate Focal adhesion kinase (FAK) (Parsons and Parsons, 2004; SADA et al., 1997), previously shown by our group to be amplified, overexpressed and constitutively activated in SCLC (Ocak et al., 2010; 2011). They play key roles in anchorage independence, survival and oxidative stress response by activating multiple downstream pathways including AKT and ERK kinases (Mócsai et al., 2010; Saito et al., 2010).

We found that SYK knock-down significantly decreased viability and growth rates in SYK/FYN-positive SCLC via increased cell death (Figure 3.9), suggesting that SYK plays an oncogenic driver role and that inhibitors could potentially be used in SYK-

positive SCLC, alone or in combination with chemotherapy. Increased cell death was also observed in AML via knock-down of SYK (Hahn et al., 2009). Further studies are needed to discriminate between overexpression versus activation of SYK in SCLC.

Our findings unveil an unsuspected link between SCLC and the biology of B-cell leukemias/lymphomas that is worth exploring. The role of SYK in B-cell receptor (BCR) initiated tonic signaling both in normal B-cells and lymphomas is well established (Chen et al., 2008; Mócsai et al., 2010). Tonic signaling promotes proliferation and survival of B-cells. Mice lacking SYK exhibit profound B-cell development deficits, and die embryonically from severe hemorrhages, also pointing to indispensable SYK signaling in cell types other than B-cells (Cheng et al., 1995). Targeted SYK therapy has been advocated in various types of B-cell lymphomas, and specific inhibitors for its kinase activity are already approved such as R406, fostamatinib (Chen et al., 2008; Cheng et al., 2011; Friedberg et al., 2010; Riccaboni et al., 2010), opening avenues for testing targeted treatment in SCLC. SYK signaling in NE (and possibly SCLC) may be associated with oxygen sensing (Buttigieg et al., 2012), but SYK-associated receptor(s) in NE or SCLC cells remain to be defined.

There are several reports of tumor suppressor functions for SYK in several solid tumor types, including breast cancer (Coopman et al., 2000), gastric cancer, and melanoma (Coopman and Mueller, 2006). Additional data are needed to reconcile these seemingly conflicting roles of SYK as oncogene or tumor suppressor. In this regard, it is worth noting that in B-cells effects of SYK on survival and proliferation are modulated by associated SRC-family kinase members (Woyach et al., 2012). Differential interactions of SYK with such kinases in a tumor-specific manner are a possible explanation for the

dual role of SYK as a tumor suppressor in some cancers (Coopman et al., 2000; Coopman and Mueller, 2006), and an oncogene in hematologic malignancies (Buchner et al., 2009; Hahn et al., 2009) and SCLC. Therefore, an immediate priority is to determine the type of receptor SYK is associated with in SCLC, and its possible regulation by SRC-family kinases such as FYN.

In agreement with our results, in the Cancer Cell Line Encyclopedia (Barretina et al., 2013), 35 out of 49 SCLC cell lines tested overexpress SYK (> 2 fold of the median centered intensity values). In another recent large dataset 33 of 53 SCLC cell lines overexpress SYK (Garnett et al., 2012). We confined our experimentation to SCLC cultured cell lines and knock-down of SYK expression. While our data are encouraging, future studies should address applicability to spontaneous (Sutherland et al., 2011) or human xenotransplant mouse models of SCLC (Daniel et al., 2009). Furthermore, it remains to be seen whether inhibition of SYK-kinase activity, in addition to expression, elicits a death response in SCLC.

It is worth noting that to date no SYK mutations have been reported in any tumor type. SYK gene fusions or translocations have been reported in hematologic malignancies, in which a driver function for overexpressed SYK has also been postulated (Kuno, 2001; Mócsai et al., 2010; Rigby et al., 2009). On the other hand, SYK negative tumors have hypermethylation and loss of function of the SYK gene (Yuan et al., 2001). Thus, the biology of SYK-positive SCLC tumors may be potentially distinct from SYK-negative SCLC tumors, with differences due to stages of progression, or divergence of transforming mechanisms.

SYK signaling functions are mediated in concert with SRC-family kinases (Mócsai et al., 2010). This subject is not fully understood and, in particular it is not clear to what extent various SRC-family kinases are interchangeable in this role within a given cell type. It is perhaps not coincidental that a SRC-family kinase, FYN, was identified in the blue module by WGCNA and that a strong co-expression correlation was found in SCLC TMAs and cell lines (Figure 3.7 and 3.8). Byers.et.al also reported activation of SRC-family kinases in SCLC assessed via reverse phase protein arrays (RPPA) (Byers et al., 2012). On the other hand, FYN kinase inhibition had no effect on SCLC cell line survival (Figure 3.10). Clarifying the SYK-FYN signaling connection in SCLC, and the possible redundancy of SRC-family kinases may open avenues to productively deploy inhibitory combination of SYK and FYN targeted therapy.

In the TMA patient dataset, we detected 2 groups of SCLC based on SYK/FYN expression alone (Figure 3.7). Admittedly, this dataset is too small to reach conclusions, highlighting the need for larger patient populations. Nonetheless, our observations raise the possibility of distinct treatment strategies in SYK-positive SCLC tumors, by analogy to lung tumors overexpressing EGFR, or HER2+ breast cancers, whose response to targeted therapy dramatically improves the outcome (Hirsch, 2003; Hirsch et al., 2006).

Here we have implemented an alternative strategy to large scale sequencing, based on a systems view of signaling networks provided by gene co-expression analysis. We respectfully submit that this approach can provide useful translational insights in the biology of specific cancer types.

Acknowledgements

We thank Dr. Steve Horvath (UCLA), Dr. Darren Tyson and Shawn Garbett for valuable input on statistical data analysis and visualization. We also thank the Vanderbilt Epithelial Biology Center and Vanderbilt Translational Pathology Shared Resource for imaging and staining of the patient TMA respectively, and the Vanderbilt Genome Science Resources for generating the RNAseq dataset.

AS: Grant 1UL1 RR029893 from the National Center for Research Resources and 1 R01 LM011179-01A1 from the National Library of Medicine, National Institutes of Health.

AU, VQ, LE: Grant 5 U54 CA113007-09 from the National Cancer Institute, Integrative Cancer Biology Program (NCI ICBP).

PPM, AS: Grant 1I01CX000242 from the Department of Veterans Affairs and CA90949 from the NCI SPORE program.

CHAPTER IV

DISTINCT TRANSCRIPTIONAL PROGRAMS DRIVE PHENOTYPIC HETEROGENEITY IN SMALL-CELL LUNG CANCER

Abstract

While small-cell lung cancer (SCLC) is the most lethal amongst lung cancers, it lacks targetable single mutant/amplified oncogenic drivers, and is monolithically treated with standard combination chemotherapy (cisplatin-etoposide). Here, we have identified two anti-correlated modules of co-expressed genes that together provide a framework for a heterogeneous phenotypic state space in SCLC. Mathematical modeling of a common transcriptional network regulating these modules predicts a discretization of the seemingly continuous phenotypic state space of SCLC into two distinct attractor basin clusters – neuroendocrine and mesenchymal. Each cluster of attractors is defined by specific stable activation states of neuroendocrine, epithelial and mesenchymal transcription factors (TFs). These attractor states were experimentally validated via differences in protein expression of TFs, surface markers, kinases, and signaling in SCLC cell lines and patient samples. At a single-cell level, multidimensional flow cytometry analysis reconfirms phenotypic heterogeneity in SCLC as two discrete stable attractors. Collectively, a mixed bioinformatic-modeling-experimental approach defines heterogeneity in human SCLC as distinct phenotypic attractor states - neuroendocrine and mesenchymal, and provides a foundation for guiding personalized treatment strategies for SCLC patients in the future.

Introduction

Small-cell lung cancer (SCLC) comprises 15-20% of all lung cancers, but exhibits the worst survival statistic (less than 5% at 5 years). SCLC occurs almost exclusively in heavy smokers and close to 100% of patients present inactivation or chromosomal loss of tumor suppressors p53 and Rb (Peifer et al., 2012). Genetic mouse models of SCLC have been generated by Cre-mediated knockout of p53 and Rb in the Pulmonary NeuroEndocrine Cells (PNEC), supporting the hypothesis that PNEC is the cell of origin for SCLC (Calbo et al., 2011; Park et al., 2011a; Sutherland et al., 2011). Standard of care of SCLC has remained unchanged for decades and consists of combination chemotherapy comprising of cisplatin and etoposide and prophylactic cranial irradiation, which increases overall survival from 6 weeks to 6-8 months (Fischer et al., 2007; Hann and Rudin, 2007). Typically, SCLC patients do not undergo surgical resection leading to a lack of availability of tissue specimens for 'omics' studies. Nonetheless, based on available datasets, several attempts have been made to make inroads into actionable molecular drivers of SCLC such as PARP1 (Byers et al., 2012), SOX2 (Rudin et al., 2012), FAK (Ocak et al., 2010) and SYK (Udyavar et al., 2013). However, thus far, clinical trials of single oncogenic driver targeted therapeutics in SCLC have been disappointing (Rossi et al., 2008). Clearly there is an urgent need for fundamental insights into the biology of this disease that would lead to novel treatment strategies.

We previously used a gene co-expression network-based approach to begin identifying subsets of SCLC with the hope of uncovering personalized targeting strategies (Udyavar et al., 2013). Here we define these subsets in terms of transcription factor networks that regulate them and then build a theoretical framework that describes

a phenotypic landscape for SCLC. This landscape is characterized by stable attractors, each defined by alternative steady states of a SCLC-specific core transcriptional regulatory network. This approach is inspired by recently proposed concepts in theoretical systems biology (Huang et al., 2012; Huang et al., 2013; Wang et al., 2010; Wang et al., 2011; Bhattacharya et al., 2011) that merge dynamical systems theory of attractors (whose origin can be traced to Poincaré and Lyapunov), with Waddington's epigenetic landscape whereby differentiating cells roll down an energy landscape into distinct basins. Here we show that the attractor theory can usefully describe phenotypic heterogeneity in human SCLC, in ways that may lead to actionable strategies for treatment.

Results

Anti-correlated gene co-expression networks suggest a heterogeneous phenotypic state space in human SCLC.

Weighted gene co-expression network analysis (WGCNA) identifies networks of co-expressed genes (termed modules by WGCNA) specific to a phenotype of interest from a gene expression dataset (Langfelder and Horvath, 2008; Parikshak et al., 2013; Voineagu et al., 2012; Wang et al., 2011a). Using WGCNA, we previously defined a SCLC-specific network signature conserved at the genomic and proteomic level in both cell lines and patients (Udyavar et al., 2013). This SCLC-specific signature unequivocally distinguishes SCLC from all other lung cancer types and normal lung. However, it also provided indications of SCLC inter-tumor heterogeneity as judged by graded expression of signature genes from specimen-to-specimen. Since available

SCLC patient specimens are too limited to support a statistically significant clustering analysis, to further investigate SCLC heterogeneity we applied WGCNA to a well characterized collection of 53 SCLC cell lines from the Cancer Cell Line Encyclopedia (CCLE) database (Barretina et al., 2013). WGCNA identified 14 modules, each labeled by color and containing an exclusive set of co-expressed genes (Figure 4.1A). The Blue module contains genes that overlap by 58% with the SCLC-specific modules previously established from tumor specimens (Chapter III) (Udyavar et al., 2013), supporting its translational relevance.

A heat-map (Figure 4.1B, left panel) of the Blue module genes shows differential expression across the panel of 53 SCLC cell lines, suggesting phenotypic heterogeneity. Furthermore, expression of genes from the Turquoise module appears to be anti-correlated (Figure 4.1B, right panel). The anti-correlated pattern of Blue and Turquoise modules was quantified by comparing their first principal components, or module eigengenes (MEs) shown as barplots (cor: -0.86, p-value: 1.6e-16) (Figure 4.1B).

Gene co-expression networks are derived from WGCNA using correlations between genes, agnostic of the underlying biology. Co-expressed genes are highly correlated with one another, possibly participate in similar pathways. EnrichmentMap and Gene Ontology (Merico et al., 2010) analysis on the Blue and Turquoise module (See Chapter II) revealed that the Blue module is enriched in neuroendocrine and epithelial differentiation processes (namely, neuronal differentiation/development, synaptic assembly, neurotransmission), while the Turquoise module contains pathways

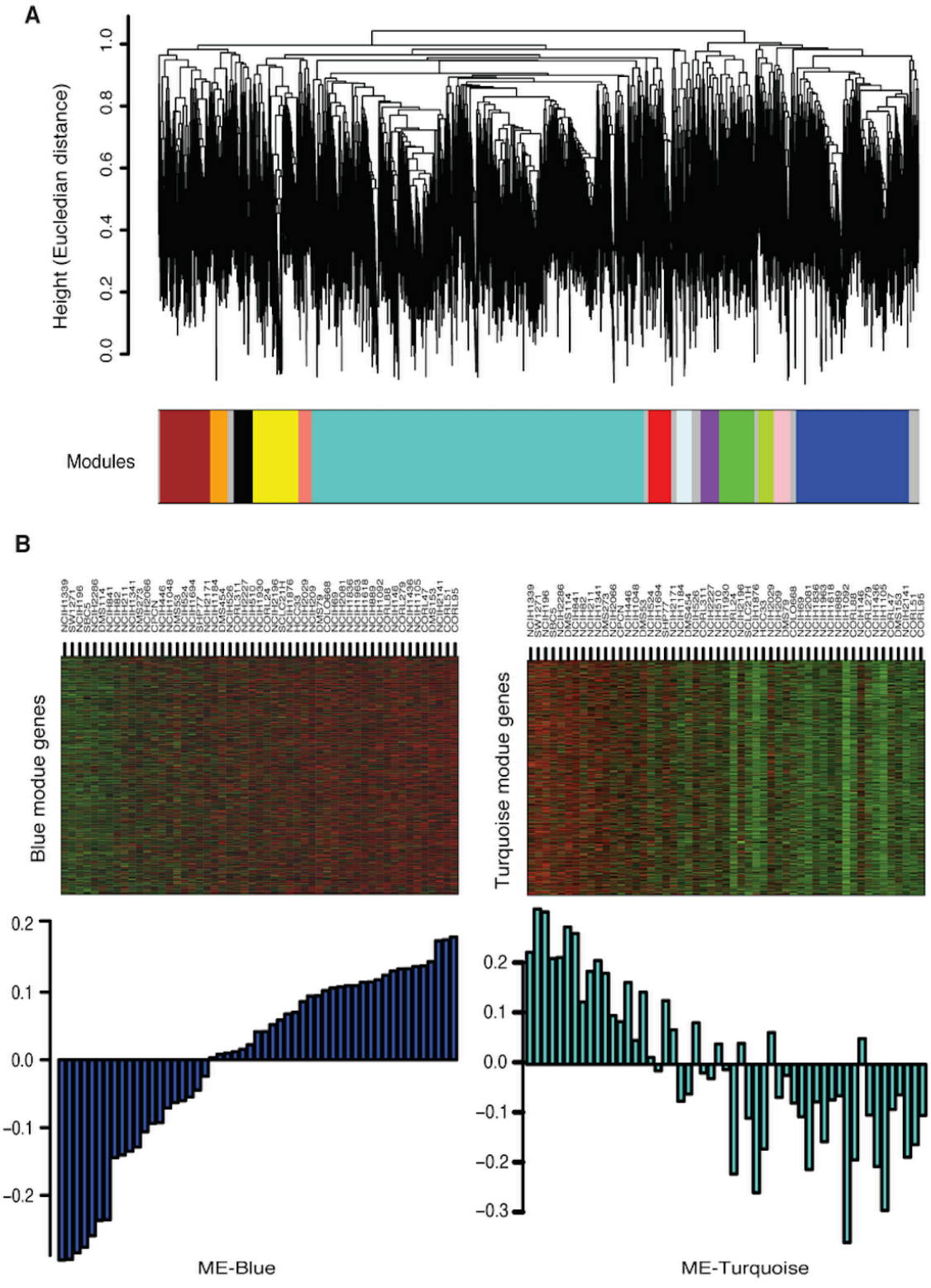


Figure 4.1: Identification of anti-correlated gene co-expression networks that delineate a phenotypic state space in human SCLC.

Figure 4.1: Identification of anti-correlated gene co-expression networks that delineate a phenotypic state space in human SCLC.

(A) Unsupervised hierarchical clustering analysis of the 53 SCLC cell lines from CCLE dataset identifies 14 modules of co-expressed genes (indicated by the black lines), given by the various colors below the dendrogram as defined by WGCNA method. (B) Heatmap view of the Blue and Turquoise module genes (rows) across 53 SCLC cell lines (columns). Eigengenes of the two modules (ME-Blue and ME-Turquoise) effectively summarize the expression of the modules as a whole in each SCLC cell line shown as a barplot below the heatmaps for the respective modules. Linear projection of the modules given by ME-Blue and ME-Turquoise shows high expression of ME-Blue on 1 end and ME-Turquoise on the other indicating anti-correlation in each sample (cor: -0.86, p-value: 1.6e-16).

involved in mesenchymal phenotype and epithelial-mesenchymal transition (TGFbeta, NFKB, cytokine signaling, cell-matrix adhesion) (Figure 4.2 and 4.3A). To compare the Blue and Turquoise module pathways to published signatures of established neuronal and mesenchymal tumor subsets, we performed gene-set enrichment analysis (GSEA) with published signatures of proneural, proliferative and mesenchymal subtypes of gliomas (Verhaak et al., 2010). The proneural and mesenchymal glioma subtype signatures were enriched in the ME-Blue high cell lines and the ME-Turquoise high SCLC cell lines respectively, thus validating the neuroendocrine (NE) and mesenchymal-like (ML) phenotypic state enrichment in the 2 networks (Figure 4.3B). Interestingly the proliferative glioma subtype signature was not significantly enriched in either of the 2 modules (p-value >0.1, data not shown). Network view of Blue (Figure 4.4A) and Turquoise (Figure 4.4B) networks also highlights the well-known neuroendocrine/epithelial and mesenchymal/EMT phenotype hubs (highly connected genes) respectively enriched in these networks. Based on these results, we hypothesized that the Blue-Turquoise module anti-correlated gene expression is reflective of a heterogeneous SCLC phenotypic state space composed of NE and ML states, and set out to test this possibility by a combined theoretical and experimental approach.

State-space analysis (Boolean model simulations) of transcription factor network dynamics predict phenotypic attractor states in SCLC.

Differentiated phenotypic states are generally controlled by regulatory networks of transcription factors (TFs) that drive expression of co-regulated target genes (Huang,

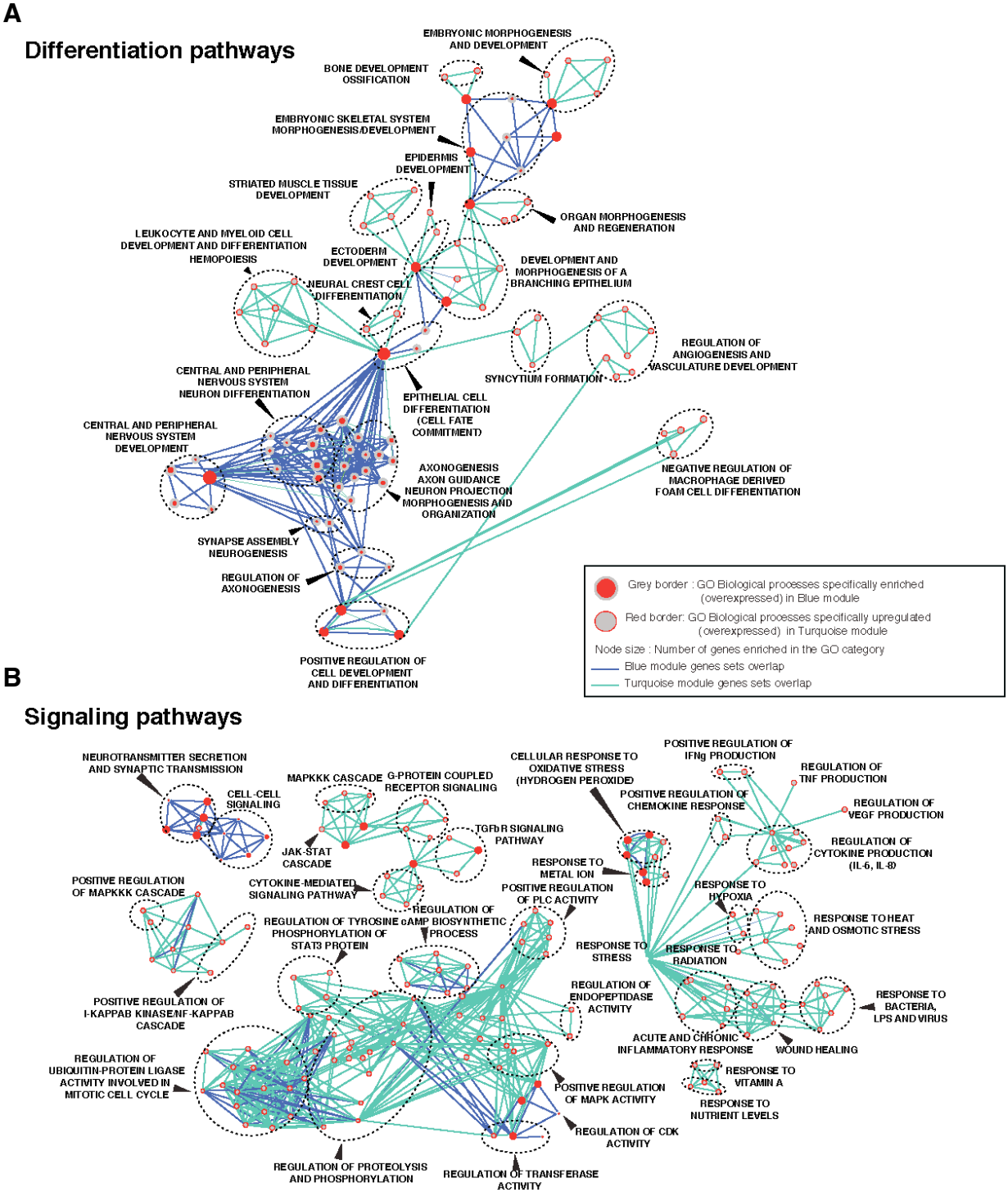


Figure 4.2: Differentiation pathway differences between the Blue and Turquoise modules given by comparative pathway enrichment analysis.

Figure 4.2: Differentiation pathway differences between the Blue and Turquoise modules given by comparative pathway enrichment analysis.

This figure describes comparative enrichment analysis of Gene Ontology (GO) pathways enriched in Blue versus Turquoise networks using BINGO and EnrichmentMap in Cytoscape®. Blue and turquoise modules show statistically significant differences in metabolism, signaling, stress response, adhesion, differentiation, transcription, proliferation and apoptosis. Nodes denote the enriched GO categories. Edges denote the connections between the pathways, Blue and Turquoise colors of the edges indicate pathway groups enriched specifically in Blue or Turquoise modules respectively. (A) and (B) show differences in differentiation and signaling pathways respectively in the 2 modules. Blue module shows enrichment for epithelial and neuronal development and differentiation, neuronal signaling, axon guidance, neurotransmitter secretion and cell-cell signaling. Turquoise module shows enrichment for myeloid and neural crest differentiation, MAPKK, JAK-STAT, NFKappaB, TGFBeta, cytokine signaling cascades (TNF, VEGF, IL-6, IL-8) that are known to be associated with a mesenchymal/EMT phenotype.

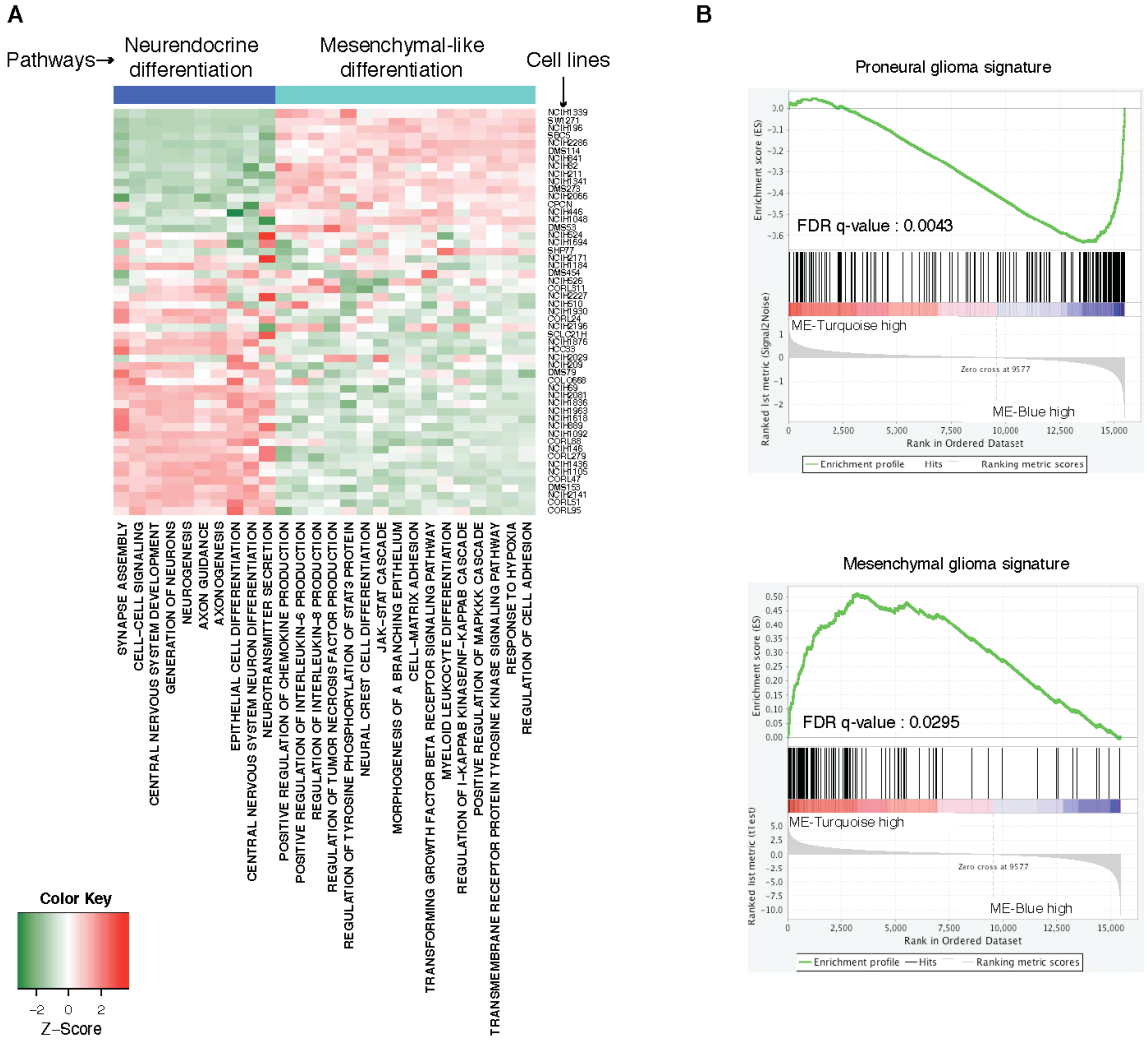


Figure 4.3: Pathway expression of the Blue and Turquoise modules given by comparative pathway enrichment analysis.

Figure 4.3: Pathway expression of the Blue and Turquoise modules given by comparative pathway enrichment analysis.

(A) Summarized gene expression of statistically significant differentiation and signaling pathways (obtained by mean of the expression values of the genes within a pathway, in columns) across the 53 SCLC cell lines (ordered by ME-Blue, in rows) in the 2 modules that were identified in Figure 4.2. Blue module shows enrichment for neuronal signaling, axon guidance, neurotransmitter secretion and cell-cell signaling. Turquoise module shows enrichment for MAPKK, JAK-STAT, NFKappaB, TGFBeta, cytokine signaling cascades (TNF, VEGF, IL-6, IL-8) that are known to be associated with a mesenchymal/EMT phenotype. (B) GSEA enrichment analysis of gene signatures of proneural and mesenchymal subtypes of glioma in ME-Blue high vs ME-Turquoise high SCLC cell lines.

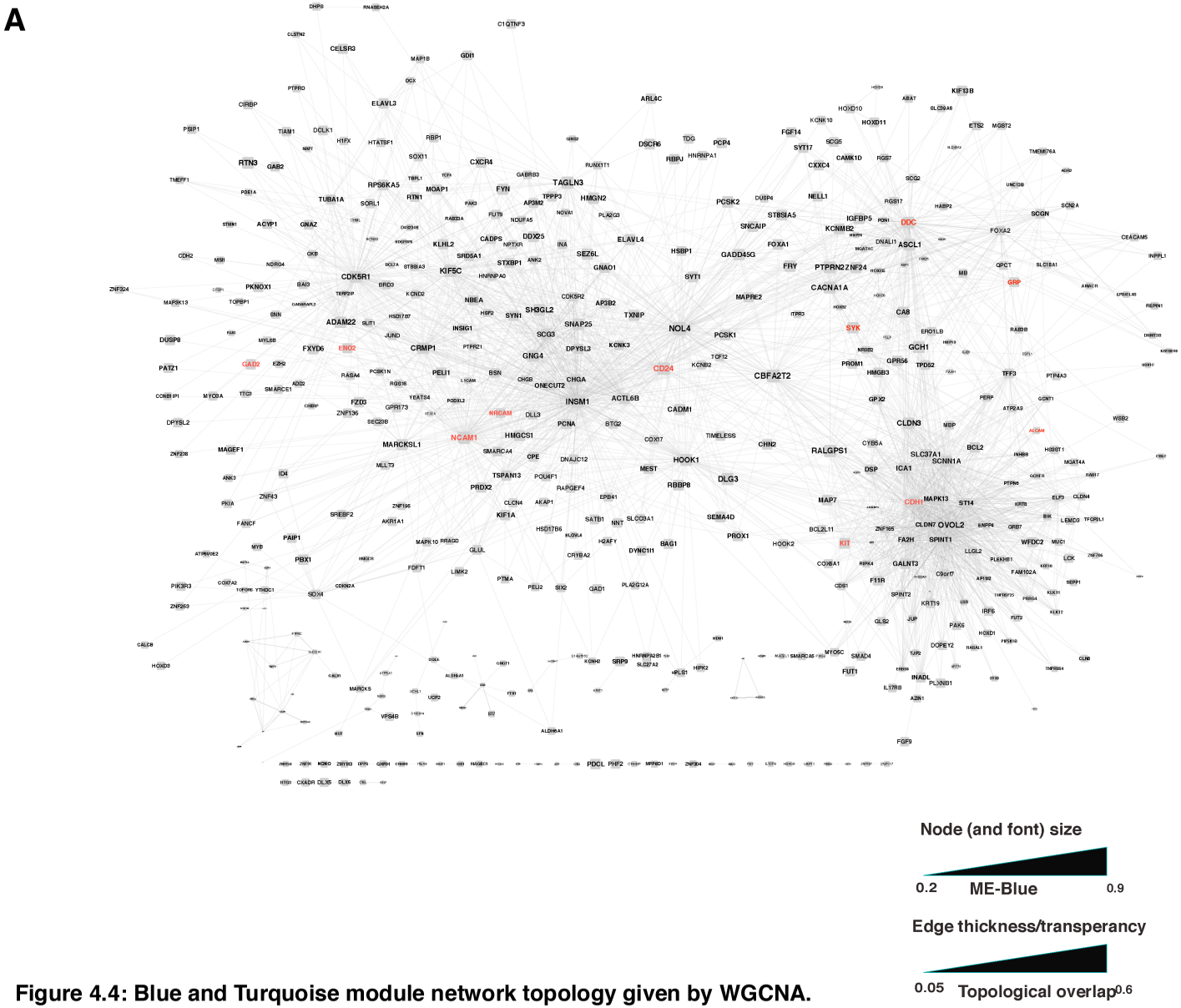


Figure 4.4: Blue and Turquoise module network topology given by WGCNA.

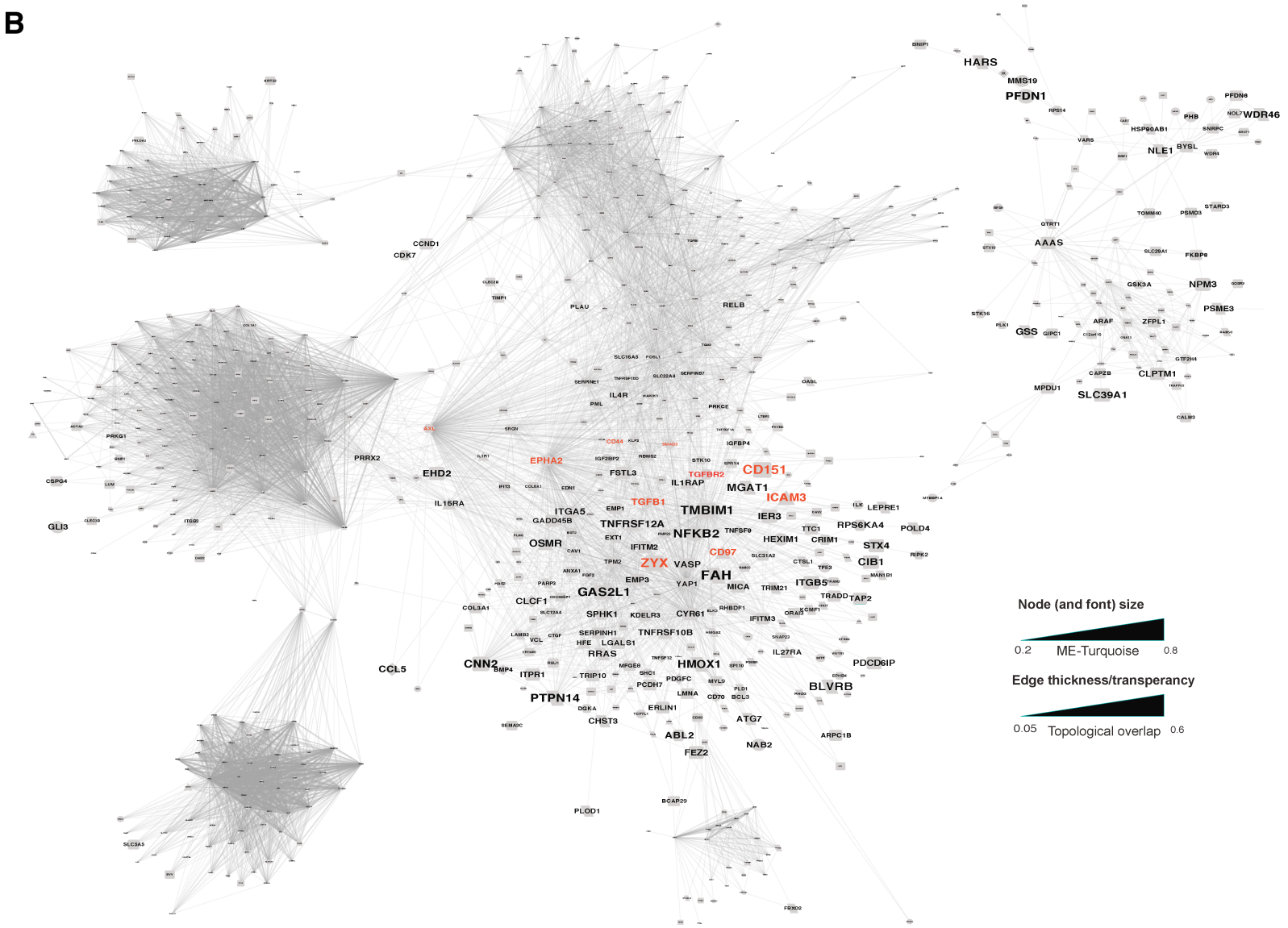


Figure 4.4 continued: Blue and Turquoise module network topology given by WGCNA.

Figure 4.4: Blue and Turquoise module network topology given by WGCNA.

This figure shows nodes within the (A) Blue and (B) Turquoise modules defined via WGCNA while edges denote a topological overlap measure (TOM). TOM is a metric for the degree of co-expression/correlation between a pair of genes (Langfelder and Horvath, 2008). If the TOM is significant, an edge is drawn between a pair of genes. The thickness of the edges denotes high TOM (> 0.1). The size of the node and its font denotes module eigengene values for the respective modules, higher the value – larger the size of the node and its font. The nodes in red denote well known and novel hub biomarkers of neuroendocrine, epithelial and mesenchymal differentiation.

2011). An appealing theoretical framework is that a differentiated phenotypic state may be a stable 'attractor' arising from TF network dynamics (Huang, 2012). Specifically, within this framework the SCLC phenotypic states would arise as stable attractors from the dynamics of a network comprised of TFs that are hubs of the Blue and Turquoise networks.

Therefore, we used the ARACNe algorithm to predict the TFs that correlate with genes expressed in the Blue and Turquoise modules, respectively. We first generated a SCLC-specific TF regulatory network using ARACNe where TFs-nonTF interactions are inferred using a mutual-information based method (Figure 4.5A). Overlap of this transcriptional regulatory network with the previously established SCLC NE and ML signatures (Figures 4.2, 4.3 and 4.4) was tested using global master regulator analysis or Fischer's exact test (Lefebvre et al., 2012) of the Blue and Turquoise genes, revealing 73 significant (FDR < 0.05) master regulators of the Blue network, 184 TFs (FDR < 0.05) of the Turquoise network and 23 common regulators (FDR < 0.05) of the 2 networks (Figure 4.5B). Interestingly, most of the master regulators of the Blue aka NE network showed positive and negative correlation with its target genes from the Blue/NE and Turquoise/ML modules respectively, while the opposite trend was observed in the Turquoise module master regulators (Figure 4.5B and C). This suggested that the master regulators differentially regulate the NE and ML networks, underscoring the anti-correlation between these 2 networks.

Furthermore, we independently validated the ARACNe predicted regulatory TFs with CHIP-Seq and TF-target binding site prediction and literature databases (Chapter II). With these filtration steps, we identified a list of 76

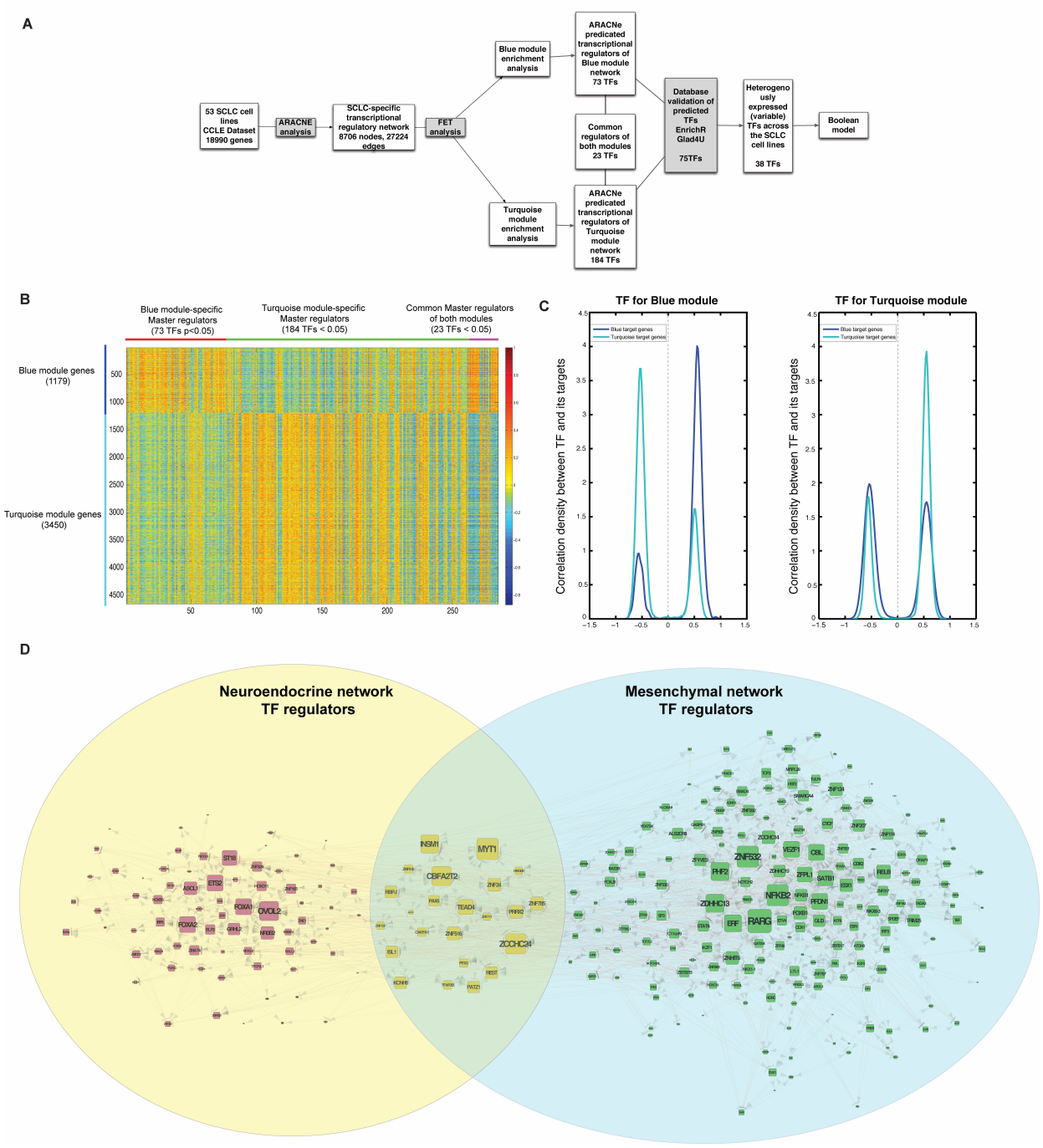


Figure 4.5: Identification of transcription factors that regulate SCLC phenotypic states.

Figure 4.5: Identification of transcription factors that regulate SCLC phenotypic states.
(A) Overview of ARACNe analysis. To identify a global SCLC transcriptional regulatory network, ARACNE analysis (based on mutual information between genes) was performed on 53 SCLC cell lines and 18990 genes in the CCLE dataset. The analysis yielded a network of 8706 nodes (genes) and 27224 edges (see Supplementary section). This core SCLC network was analyzed using master regulator analysis to identify top transcription factors (TF) that act as master regulators of either the neuroendocrine or mesenchymal networks (identified via WGCNA). These TFs were independently validated using literature and transcription factor ChIP-Seq and TF-binding site prediction databases, leading to a final list of 76 TFs. Only the most variant TFs across the SCLC cell lines were selected for building the boolean model network. (B) Correlation heatmap plot of individual 73 Blue and 184 Turquoise module and 23 common TF regulators (columns) with 1179 Blue and 3471 Turquoise module genes (rows). Yellow-orange-red indicates positive correlation suggesting positive target gene regulation while green-blue indicates negative correlation suggesting negative target gene regulation. (C) Density histogram of the Blue/ Turquoise TF regulators and correlation with its targets in the Blue or Turquoise modules. This suggests that a particular TF differentially regulates the 2 modules. (D) ARACNE network view of the top TFs shown in B and C (identified via master regulator analysis) that regulate the Blue, Turquoise or both modules. The node connectivity of a TF is given by its bigger size indicative of the number of targets regulated by the TF. Edges are derived from ARACNE mutual information between the nodes given its co-expression.

likely TF regulators of NE and/or ML differentiation. Finally, to identify TFs which play an active role in maintaining the phenotypic differences between the SCLC cell lines, we considered only heterogeneously expressed TFs (median absolute deviation above the 50th percentile) yielding a final list of 38 TFs that we used to build a boolean network model for SCLC (Figure 4.6). We then extracted interactions between these TFs from manual curation of literature, ChIP-X databases (ChEA (Lachmann et al., 2010), ENCODE (Landt et al., 2012) and TF-binding motif predictions (TRANSFAC (Matys et al., 2003), JASPAR (Mathelier et al., 2014)).

With this information, we constructed a directed network (see methods) consisting of 38 SCLC TFs (Figure 4.6), 33 of which contribute to a complex feedback structure, while the remaining 5 serve as sink nodes (only incoming and no outgoing interactions) . Because the 5 sinks are incapable of contributing to dynamic regulation of the network, they are excluded from remaining analyses. Direct positive and negative feedback loops may be visualized by splitting the network into two subsets of nodes such that each side has a minimum number of negative feedback loops within it, and a maximum number between the two sides (Figure 4.7A). The left subset was enriched for TFs known to control neuroendocrine and epithelial phenotype, the other for mesenchymal/EMT TFs. The dynamics of this TF network can be simulated using Boolean logic (Saadatpour et al., 2010; Wang et al., 2012), which represents a coarse-grained approximation wherein TFs are in one of 2 discrete states – ON or OFF. This approach has been successfully applied to study TF network dynamics and identifies stable attractors (Choi et al., 2012; Davidich and Bornholdt, 2008; Krumsiek et al., 2011; Villani et al., 2011).

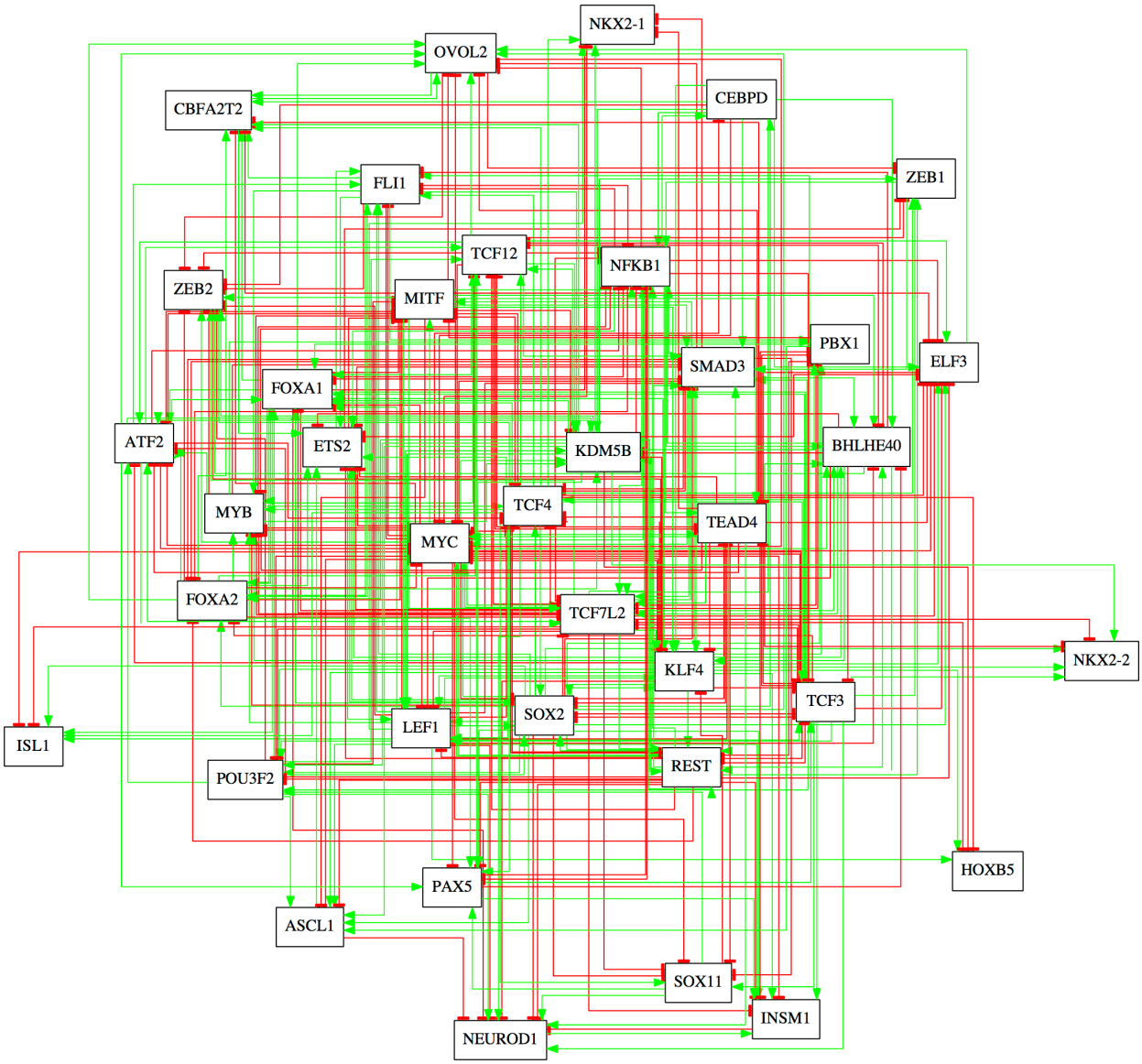


Figure 4.6: Boolean network model of SCLC-specific transcription factors.

After ARACNE analysis and filtration with databases, there were 76 TF regulators of the Blue and Turquoise modules. Selection of the most variant genes across the 53 SCLC cell lines using the median absolute deviation metric, the final list of TFs used for Boolean modeling was 38 TFs shown in this figure. Positive and negative interactions between the 38 TFs are indicated by the green and red colored lines respectively, derived from literature and publicly available TF-TF interaction databases (see Materials and Methods).

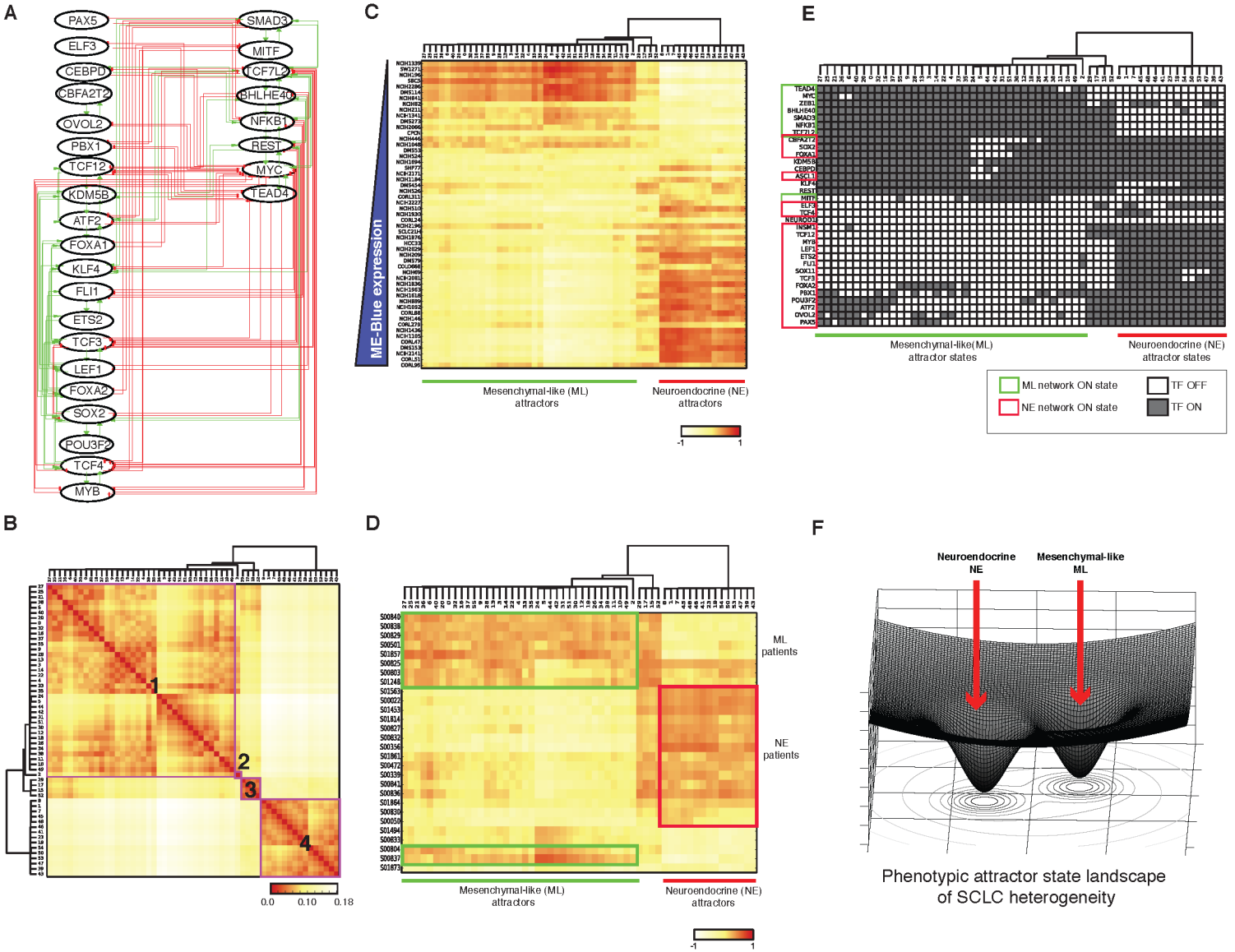


Figure 4.7: Establishment of dynamic transcription factor network attractor states corresponding to distinct SCLC differentiation states.

Figure 4.7: Establishment of dynamic transcription factor network attractor states corresponding to distinct SCLC differentiation states.

(A) Shows a Boolean network containing feedback loops between 28 core transcription factor (TF) regulators identified via TF enrichment analysis of the Blue and Turquoise modules as described in text and methods. Edges denote a TF binding to the promoter region of its target TF. Green edges denote activation while red edges denote inhibition. (B) Boolean TF network dynamics simulations identified 56 stable phenotypic attractor states (columns). At these states, the network has reached a balance point where none of the transcription factors are being induced switch on or off, given the active set of TFs. Hierarchical clustering suggests there are 4 distinct groups of attractor states indicated by the purple squares along the diagonal and numbers 1-4. (C) A score is computed to quantify how well each attractor (column) correlates with each cell line (row). High correlation of the left- and right-most attractor clusters with the mesenchymal (ML) and neuroendocrine (NE) cell lines suggests that these theoretically predicted attractors may represent these distinct phenotypic states. (D) Correlation between attractors and SCLC patient samples, (E) TF status (ON/OFF) for each of the 56 stable attractor states. Grey denotes TF is ON (1), while white denotes OFF (0) in a particular attractor state. (F) Conceptual illustration diagram of attractor state landscape in human SCLC.

Using a random order asynchronous update scheme (see methods section), we found 56 stable fixed-point but no oscillating attractors (Figure 4.7B and E), consistent with the expectation that cell phenotypic states should be steady. Unsupervised hierarchical clustering segregates the 56 attractors into four distinct clusters (Figure 4.7B), each characterized by a common ON-OFF configuration of the 33 TFs (Figure 4.7E). The leftmost (attractors 27 to 49) and rightmost (attractors 8 to 43) clusters are distinguished by mutually exclusive expression of mesenchymal/EMT (MYC, NFKB1, SMAD3), and neuroendocrine (INSM1, POU3F2, SOX2, SOX11) and epithelial (FOXA2, OVOL2) TFs respectively. This suggests that the left and right clusters might represent neuroendocrine (NE) and mesenchymal (ML) attractors, respectively. Interestingly, we also identified novel TFs such as MITF, TEAD4, and TCF12, MYB, LEF1, ETS2 to be specifically associated with the left and right clusters respectively.

To score how well these predicted attractors describe the experimentally measured cell line TF expression we calculated Pearson's r correlation coefficient for each attractor with each cell line. Because the model prediction limits TF expression values to 0 or 1, the expression data was linearly scaled such that each transcription factor ranged from 0 in the cell line with minimum expression to 1 in the cell line with maximum expression. Similar results were obtained when we performed this scaling across only the SCLC datasets (which we report here) and across all of the CCLE cell lines (data not shown). The high correlation of the left cluster of attractors with the mesenchymal cell lines, and the right cluster with the neuroendocrine cell lines confirms that the simulated attractors accurately reflect the biological data (Figure 4.7C). Similar

results were obtained by comparing attractors with SCLC patient data, suggesting that the insights generated by the model may have clinical significance (Figure 4.7D).

The Boolean model simulations effectively distributes SCLC cell lines and patients into two discrete phenotypic states, schematically visualized in a conceptual attractor landscape (Figure 4.7F). These two phenotypic states are driven by the dynamic feedback regulation of a key TF network existing in distinct network states summarized in Figure 4.6E.

Consensus clustering analysis orthogonally/independently validates two distinct transcriptional subtypes in SCLC cell lines and patients.

Using WGCNA, we identified clusters of genes whose anti-correlated expression described a spectrum across SCLC cell lines that spanned from high neuroendocrine (ME-Blue high) to high mesenchymal-like (ME-Turquoise high) expression patterns. Boolean network analysis identified two discrete clusters of phenotypic attractor states in SCLC, distinct from the continuous spectrum of the WGCNA module eigengenes. To further verify the existence of these SCLC phenotypic states by an orthogonal approach, we performed consensus clustering analysis (Perez-Moreno et al., 2012; Zhu et al., 2013) on the 53 SCLC cell lines (Barretina et al., 2013). to test the continuous/ discrete nature of these phenotypic states. We limited the analysis to genes from the Blue and Turquoise modules to reduce noise unrelated to the phenotypic heterogeneity.

Unsupervised consensus clustering revealed that the SCLC cell lines can be naturally subdivided into two clusters based on the cumulative distribution function

(CDF) curves (Figure 4.8A). This is also illustrated by the heatmaps (Figure 4.8B) where k=2 shows the least amount of inter-cluster instability. To verify that these results were present in patient samples as well, the analysis was repeated for an independent dataset of SCLC patients (Clinical Lung Cancer Genome Project (CLCGP)Network Genomic Medicine NGM, 2013) where two clusters were robustly identified given by the CDF and heatmaps (Figure 4.9).

These data strongly suggest the presence of two distinct NE and ML transcriptional subtypes in human SCLC. To our knowledge, this is the first report of phenotypically distinct transcriptional subtypes in human SCLC. We next proceeded to seek experimental validation for these theoretical and bioinformatic predictions that SCLC is comprised of two phenotypic states, NE and ML.

Experimental validation of heterogeneous phenotypic state space of SCLC.

To validate the boolean model-predicted TF ON/OFF configurations characteristic of the two discrete phenotypic states (Figure 4.10A), we first compared the gene expression status of each TF in the NE and ML transcriptional subtypes defined by consensus clustering (Figure 4.10B and C). As predicted by the model, the NE TFs were expressed at a higher level in NE cell lines than ML cell lines and vice versa (Figure 4.10B and C). Next we experimentally measured expression of 10 out of 33 TFs in SCLC cell lines (Figure 4.10D and E) that were representative of the network state configurations driving the two attractor states (Figure 4.10A). Patterns of TF protein expression were consistent with the model predictions in cell lines (Figure 4.7C and E). That is, all of the cell lines that correlated with the NE cluster, did in fact express the TFs

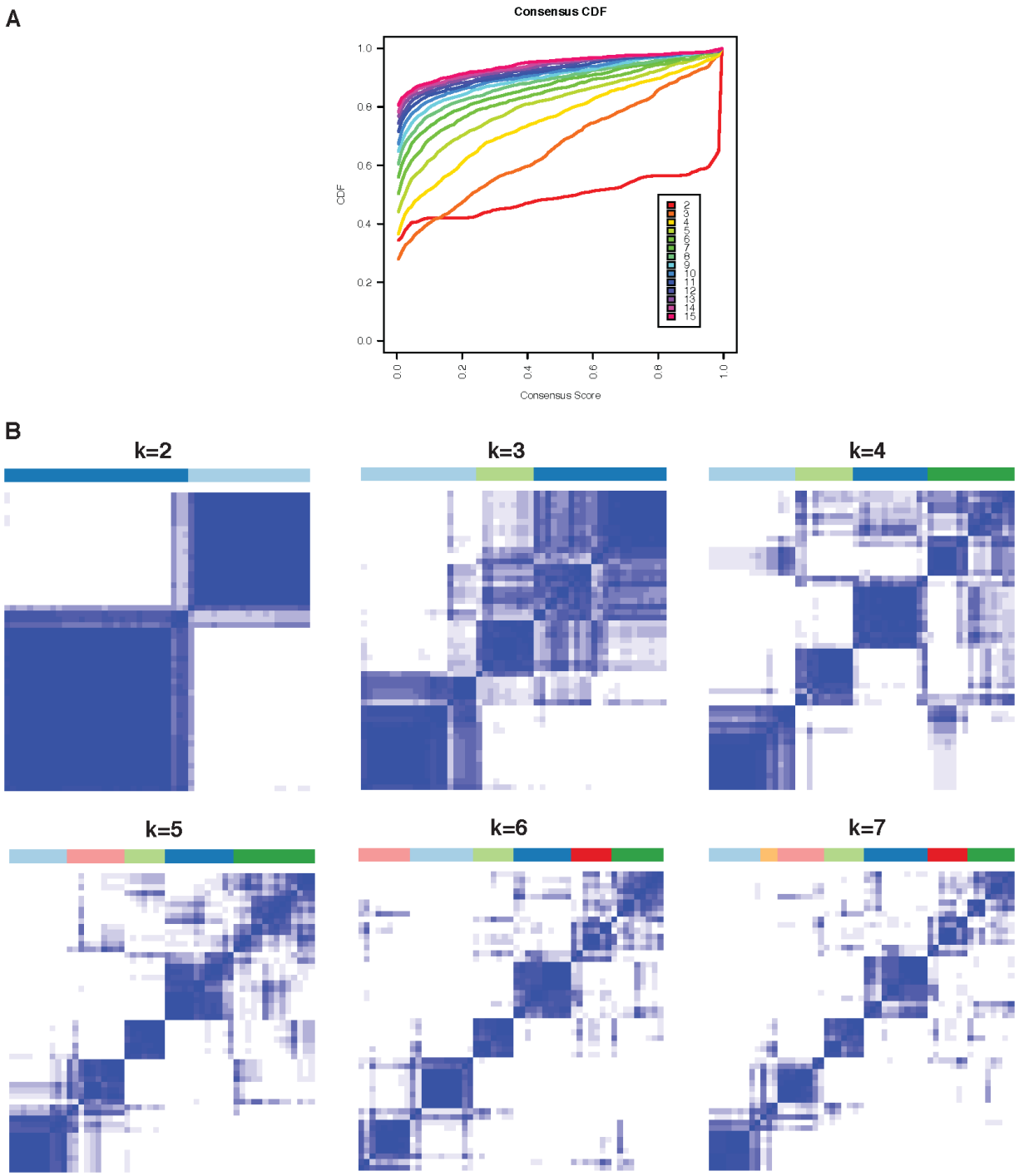


Figure 4.8: Validation of the SCLC phenotypic states in cell lines using independent consensus analysis.

Figure 4.8: Validation of the SCLC phenotypic states in cell lines using independent consensus analysis.

Consensus clustering analysis of the 53 SCLC cell lines using the Blue and Turquoise module genes. Clustering was performed using k-means with 1 - spearman correlation as the distance metric. Consensus clustering was computed using 80% sample and feature resampling. (A) shows the cumulative distribution curves for various values of k (k=2 to k=15). A good fit is identified by having a highly horizontal CDF. (B) Consensus cluster matrix heatmaps are shown for k=2 to k=15. Across 1000 iterations, the frequency that any pair of two cell lines are allocated to the same cluster is calculated and scaled from 0 (white) if they never cluster together, to 1 (blue) if they always cluster together. Since higher numbers of clusters will trivially produce a horizontal CDF, these data most strongly support 2 subtypes in SCLC.

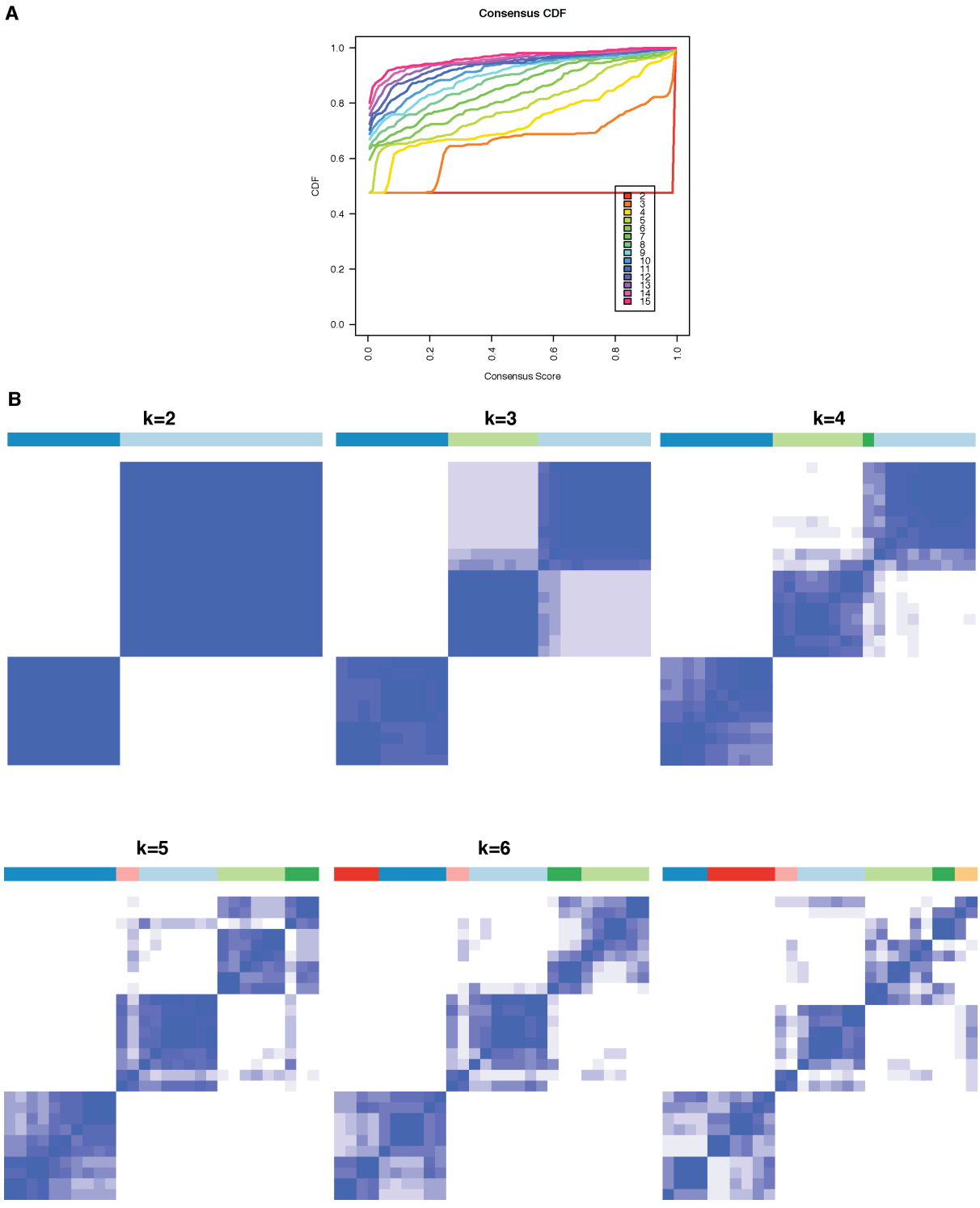


Figure 4.9: Patient dataset verification of SCLC phenotypic states using unsupervised consensus clustering analysis.

Figure 4.9: Patient dataset verification of SCLC phenotypic states using unsupervised consensus clustering analysis.

Consensus clustering analysis of the 53 SCLC cell lines using the Blue and Turquoise module genes. Clustering was performed sampling 80% of the samples and genes using k-means and (1-spearman correlation) as the distance metric. (A) shows the cumulative distance function curves at various k values (k=2 to k=15). (B) Consensus cluster matrix heatmaps identified at k=2 to k=15. During 1000 iterations of this algorithm, prediction of the patients allocated to various clusters is calculated. Each consensus cluster indicates the frequency of samples being in the same group at a particular k value. Blue indicates that the cell lines were always in the same cluster in every iteration, while white denotes that the cell lines were never in the same clusters. These data suggest that 2 subtypes exist in SCLC patients as observed in cell lines.

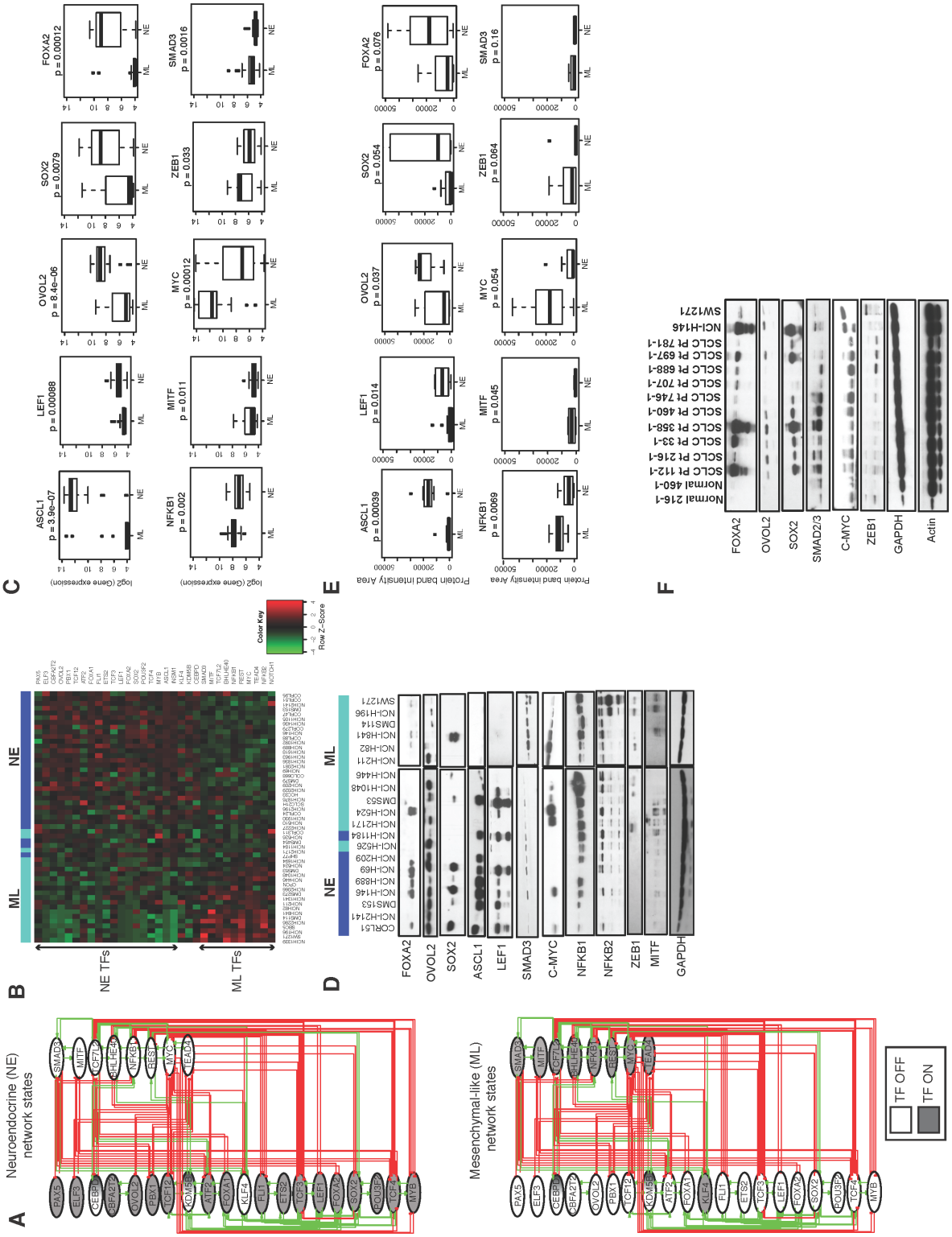


Figure 4.10: Experimental validation of TF network states in human SCLC.

Figure 4.10: Experimental validation of TF network states in human SCLC.

(A) shows the network states that drive NE (top) or ML (bottom) differentiation. The TFs are labeled grey if they are active in the boolean model attractor state or white if inactive. (B) and (C) show heatmap and boxplots denoting gene expression of the TFs that are part of the boolean network model and in the NE and ML cell lines as defined by the consensus clustering and WGCNA. The top and bottom panels denote NE and ML TFs respectively in Panel C. (D) and (E) Protein validation of the TF network states in SCLC cell lines using western blots and the quantification of the chemiluminescence bands using ImageJ where NE cell lines are characterized by expression of ASCL1, OVOL2, SOX2, LEF1 and FOXA2 while the ML cell lines are characterized by c-MYC, NFKB1, MITF and ZEB1. This validates the network state model predictions of the 2 attractor states. (F) Validation of TF expression in 10 SCLC patient samples and 2 normal samples using western blot, indicating that these TFs are also expressed in SCLC patients.

that were used to define this cluster by Boolean simulations (Figure 4.7C) and not mesenchymal TFs. Likewise, the cell lines correlating with the mesenchymal attractor cluster expressed mesenchymal, but not neuroendocrine TFs. Similar patterns of expression of the NE and ML TFs were observed in SCLC patients at gene expression level from the 28 SCLC patient subtypes defined by consensus clustering and modeling (Figure 4.10F). Independent validation of TF protein expression in 10 SCLC patient whole-tumor lysates were predictive of NE or ML phenotypic states (Figure 4.10G).

We further tested the existence of these phenotypic states experimentally by several independent orthogonal approaches. Since the Turquoise module is enriched for cell-matrix adhesion and migration pathways (Figure 4.2 and 4.3A), it raised the possibility of differential adhesion among the 2 SCLC phenotypes. Accordingly, neuroendocrine (NE) SCLC cell lines grow in culture as multicellular suspension aggregates (CORL51, NCI-H2141, NCI-H146), while the ML cell lines (ME-Turquoise high) grow as adherent monolayers (SW1271, DMS114, NCI-H841) (Figure 4.11A).

We then characterized the cell lines by expression of well-established biomarkers using flow cytometry, focusing on surface receptors, kinases/enzymes and adhesion molecules that are known biomarkers of neuroendocrine, epithelial and mesenchymal/ EMT differentiated states in other cancers. As expected, NE cell lines exhibited high expression of several neuronal markers, including NCAM1/CD56 (Marro et al., 2011), CADM1 (Thomas et al., 2008) and CD24 (Pruszak et al., 2009) (Figure 4.11B).

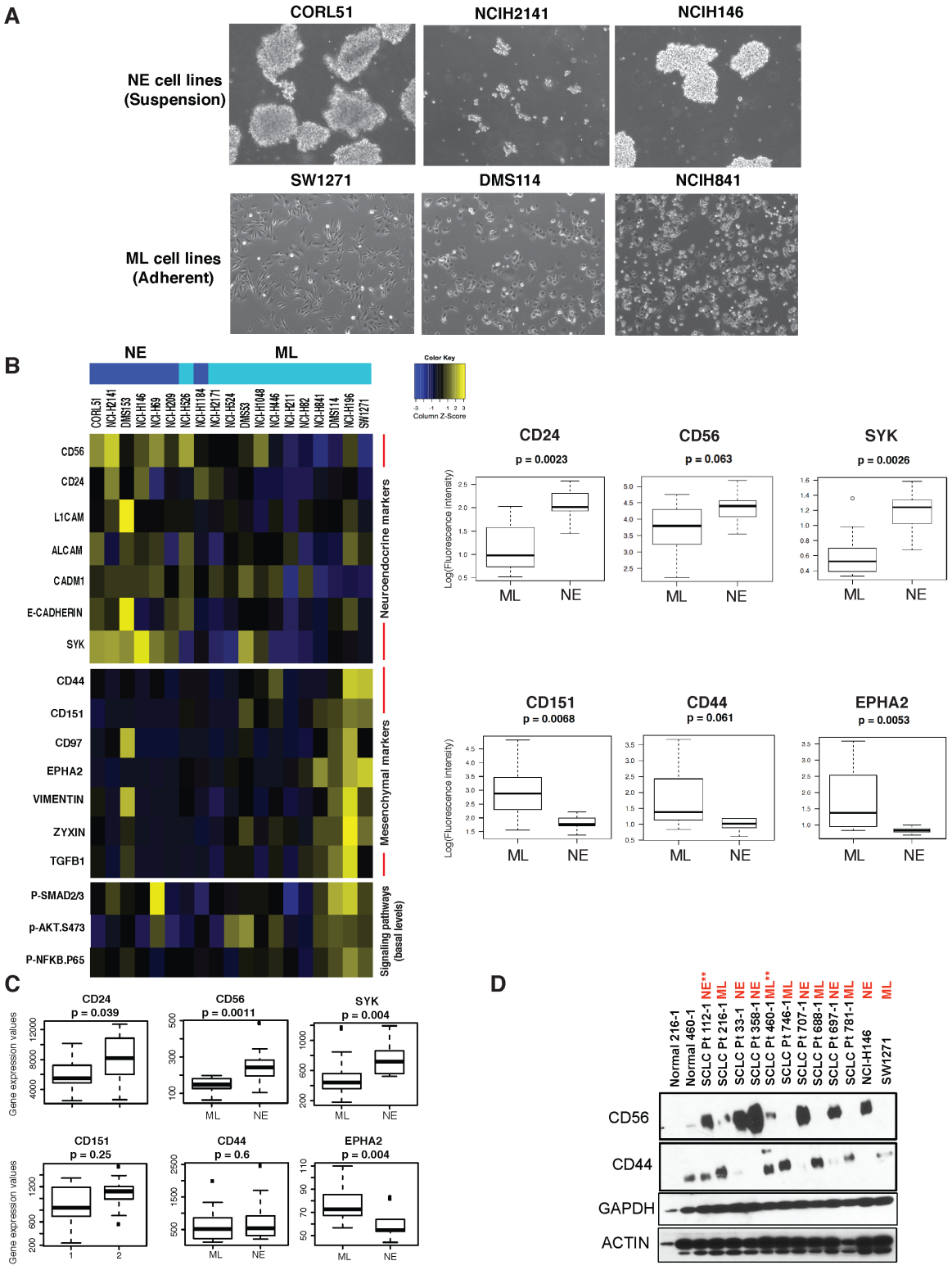


Figure 4.11: Phenotypic characterization of distinct attractor states in human SCLC cell lines and patients.

Figure 4.11: Phenotypic characterization of distinct attractor states in human SCLC cell lines and patients.

(A) shows brightfield images (20X magnification) of SCLC cell lines in *in vitro* cultures. The NE cell lines show classical SCLC suspension aggregates morphology while ML cell lines grow in adherent monolayers. (B) Proteomic validation of NE differentiation markers and kinases (NCAM1/CD56, CADM1, L1CAM, CD24, E-cadherin, SYK) and ML differentiation markers (CD44, CD151, CD97, Ephrin A2 (EPHA2), Zyxin (ZYX) and Vimentin), measured via flow cytometry displayed as heatmap and boxplots across the NE and ML subtypes defined by consensus clustering (Figure 4.8). Flow cytometry values are transformed ratios of median fluorescence intensities normalized to the minimum of each column (marker) to obtain the color scale. TGFbeta, PI3K and NFKB pathway activity is high in MC cell lines given by constitutive activity of pSMAD2/3, pAKT, and pNFKB p65. (C) Boxplot of representative surface marker gene expression in the consensus clustering-defined patient subsets from (Clinical Lung Cancer Genome Project (CLCGP)Network Genomic Medicine NGM, 2013). (D) Protein expression of representative NE and ML markers across an independent in-house set of 10 SCLC patients, 2 matched normal and 2 representative NE (NCI-H146) and ML (SW1271) cell lines. Based on the expression of NE vs ML markers, the 10 patient tumors can be classified as NE or ML.

In contrast, ML cell lines were characterized by high expression of mesenchymal differentiation markers including CD44 (Park et al., 2010; Ponta et al., 2003), CD151 (Ke et al., 2011), vimentin (Micalizzi et al., 2010) , vimentin (Polyak and Weinberg, 2009) and zyxin (Mori et al., 2009; Sperry et al., 2010). Expression of these markers is summarized in Figure 4.11B. Interestingly, Spleen Tyrosine Kinase (SYK), previously shown to be a functional target in a subset of SCLC (Udyavar et al., 2013), and Ephrin A2, a receptor tyrosine kinase important in glioblastoma stem-cells and EMT (Binda et al., 2012; Huang et al., 2013), were expressed in NE and ML SCLC cell line subsets respectively (Figure 4.11B). Expression of CD56 and CD151 was also measured by western blot in human SCLC specimens. Of these, 5 were NE, 4 were ML (Figure 4.11D). Interestingly, two tumors (indicated by asterisks) showed a mixed phenotype with expression of both NE and ML markers.

Next we measured the activation of signaling pathways that have been associated with mesenchymal differentiation and found that TGFbeta, and NFKappa beta pathways were constitutively active in the ML cell lines but not the NE cell lines. PI3Kinase pathway was constitutively active in NM and ML cell lines, not NE (Figure 4.11B). In summary, these orthogonal assays validate the existence of three distinct attractors in SCLC.

Multidimensional flow cytometry analysis captures the existence of heterogeneous phenotypic attractors in SCLC.

Attractor states are generally described as phenotypic state space that can be occupied by a single cell. To investigate if these phenotypic states can define

heterogeneity at a single cell level within SCLC cell lines, we investigated the expression of NE and ML markers at a single-cell level using flow cytometry. Biaxial plots of CD56 and CD151, neuroendocrine and mesenchymal markers, respectively, capture NE and ML heterogeneity at the single-cell level within 6 representative cell lines (Figure 4.12A). In 3 ML cell lines, $CD56^{-}CD151^{+}$ cells represented 85, 83 and 72% of the population respectively whereas $CD56^{+}CD151^{-}$ cells were less than 1%. In contrast, 3 NE cell lines contained 86, 62 and 74% of $CD56^{+}CD151^{-}$ cells and less than 2% of $CD56^{-}CD151^{+}$. These results confirm the NE and ML nature of the predicted respective attractors.The single cell analyses were then extended to 22 SCLC cell lines and 2 SCLC PDXs tested with 3 NE – CD56, CD24 and CADM1, and 3 ML surface markers – CD44, CD151 and EPHA2. To perform unsupervised clustering of this high dimensional flow cytometry dataset, we applied NbClust algorithm to predict the number of k-means clusters that would best capture the heterogeneity in this dataset. Out of 30 metrics given by NbClust for optimal cluster prediction, 12 metrics predicted 2 clusters and 6 metrics predicted 4 clusters (Figure 4.12B). Unsupervised k-means clustering analysis (with k=2) of the 24 SCLC samples showed a high degree of overlap with CD56high cells and k-means cluster 2 while cells with CD151high expression overlapped with k-means cluster 1 (Figure 4.12C, top panel). Cells in cluster 1 showed co-expression of mesenchymal markers CD151, CD44 and EPHA2 while cells in cluster 2 exhibited co-expression of neuroendocrine markers CD56, CD24 and CADM1 (Figure 4.12C and D). Each SCLC cell line and patient derived xenograft (PDX) exhibits presence of both NE and ML phenotypic states at various levels (Figure 4.12E), although the two PDX NJ-H29 and LX-22 display more heterogeneity than cell lines.

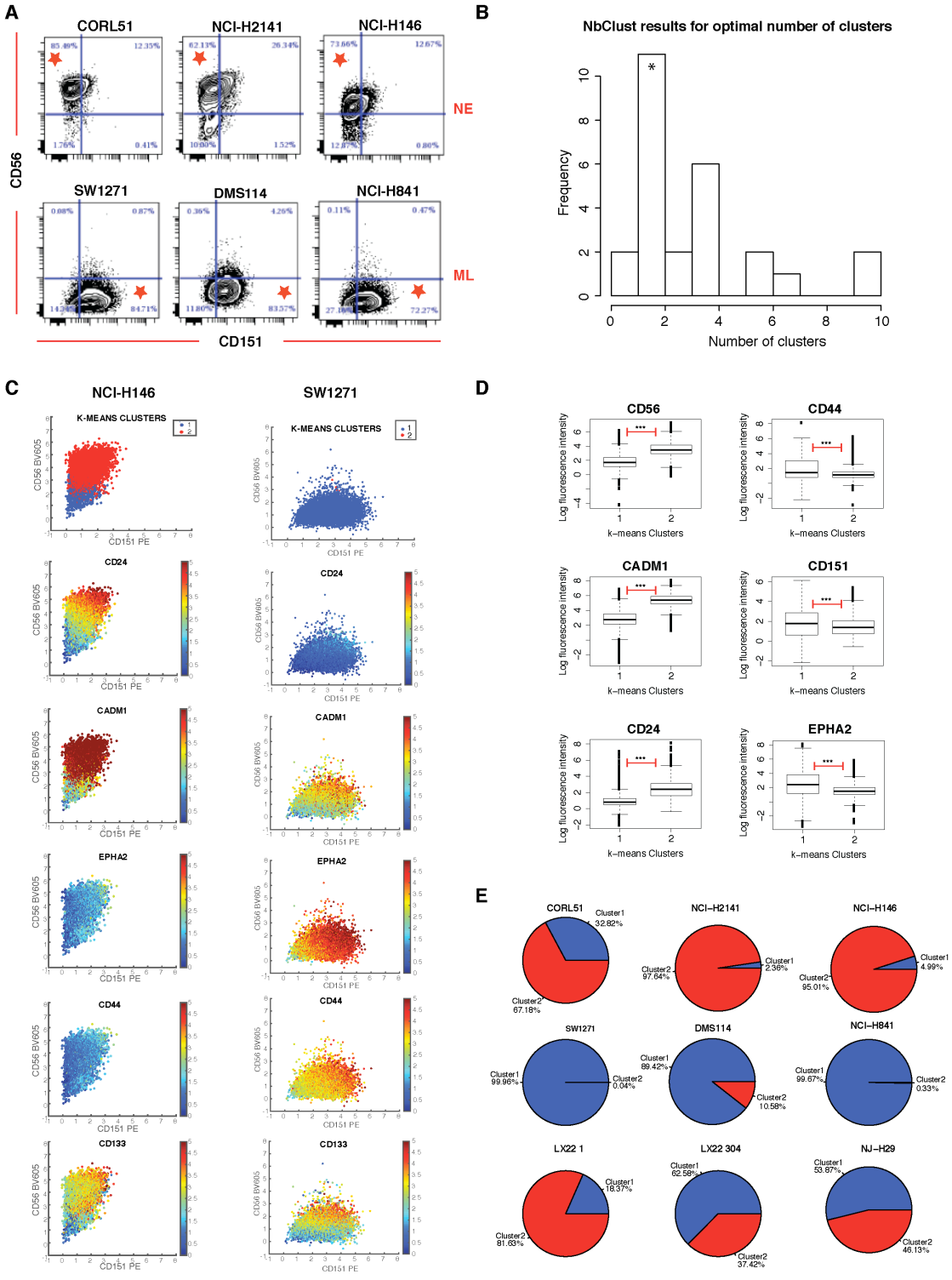


Figure 4.12: Multidimensional single-cell level analysis of attractor state space in SCLC cell lines and patients.

Figure 4.12: Multidimensional single-cell level analysis of attractor state space in SCLC cell lines and patients.

(A) Biaxial plots of neuroendocrine marker CD56 and mesenchymal marker CD151 measured via flow cytometry. Stars indicate the quadrant of the gate where majority of the cells fall. Each of the NE, NM and MC cell lines fall in a discrete phenotypic attractor state given by expression of the 2 markers. (B) To assess heterogeneity and co-expression of multiple markers at a single cell level, NbClust analysis to obtain the optimal number of clusters in the multidimensional flow cytometry data comprising of 3 NE markers (CD56, CD24, CADM1) and 3 ML markers (CD44, CD151, EPHA2) was conducted on 22 SCLC cell lines and 2 PDX. This analysis based on 30 metrics for optimal cluster determination predicted 2 k-means clusters in the dataset. Based on this k-means clustering was performed on this data to identify the 2 clusters of cell subpopulations. (C) and (D) denote the level of co-expression of NE and ML markers in the 2 k-means clusters, where the NE markers CD56, CD24 and CADM1 are co-overexpressed in Cluster 2 and the ML markers CD44, CD151 and EPHA2 are co-overexpressed in Cluster 1. Representative NE and ML cell lines NCI-H146 and SW1271 cell line plots overlaid on the CD56-CD151 biaxial plot as heat for different surface markers are shown in panel C. (E) shows the relative distribution of the NE (Cluster 2) and ML (Cluster 1) subpopulations in various SCLC cell lines including two PDX models LX-22 (1 and 304 - two independent mice) and NJ-H29.

Interestingly, the same PDX LX-22 grown two independent mice (#302 and #1) show differences in the presence of the NE and ML clusters, underscoring the impact of *in vivo* tumor microenvironment in regulation of heterogeneity.

These data strongly suggest that a single cell within a SCLC cell line and PDX can exist in distinct NE or ML attractors where cells in the NE attractor co-express NE surface markers and low expression of ML markers, and vice versa for cells in the ML attractor.

Discussion

Understanding the origins and roles of both inter- and intra-tumor heterogeneity remains a significant challenge facing cancer researchers. Molecular and genetic subtyping has introduced the promise of personalized therapies, however success has been limited in practice by a lack of well classified subtypes and the emergence of treatment resistant tumors. Intra-tumor heterogeneity has been shown in many studies to play a significant role in the emergence of resistance in many cancers (Almendro et al., 2014b; Chong and Jänne, 2013; Sharma et al., 2010), and developing strategies to abrogate these effects will be critical to the future development of more effective therapies.

Our work links statistical and bioinformatic analyses with the theoretical framework of epigenetic attractor landscapes in order to provide a unified picture of inter- and intra-tumor heterogeneity in human SCLC. It should be stressed that in this context, the term “epigenetic” does not necessarily equate with histone modifications and chromatin structure, but rather refers more generally to how the regulatory structure

imposed by genome can give rise to distinct, stable configuration. This combined approach leverages the vast amount of information in public databases to not only identify the presence of phenotypic subtypes, but provide mechanistic insights into how transcriptional regulatory networks maintain the stability of these distinct states.

Chemotherapy has remained the standard of care for SCLC patients and mechanisms of therapeutic resistance have been well documented. Adhesion to matrix elements such as laminin is one of the mechanisms of resistance in SCLC to chemotherapy (Hodkinson et al., 2006; Tsurutani, 2005). In 2011 Calbo et.al described distinct neuroendocrine and mesenchymal-like phenotypes in mouse p53-Rb null Cre-transgenic SCLC tumors (Calbo et al., 2011) strongly suggesting that intra-tumor heterogeneity was a phenomenon seen in SCLC even in genetically identical mice. We further explored this heterogeneity in human derived SCLC cell lines and clinical patient samples and identified a set of anti-correlated genes organized into the Blue and Turquoise modules. Increased expression of EMT markers such as CD44, NFKB, Vimentin, CD151, Zyxin,, and others potentially promote adhesion in the mesenchymal-like cells, suggesting a clinical significance for these findings.

SCLC has one of the highest mutation rates amongst all cancers (McFadden et al., 2014) and this genomic instability might contribute to the plasticity and diversity of cellular phenotypes. However, we find that a modeling approach including only epigenetic factors such as changes in transcription factor and marker expression is sufficient to capture the observed phenotypic states. We propose that this approach describes at a broad level the stable phenotypic states available to a cell, and that genomic heterogeneity affects the predilection an individual cell has for one state over

another. This interpretation is consistent with previous work describing non-genetic heterogeneity in murine SCLC (Calbo et al., 2011), and our current findings showing the existence of subpopulations of neuroendocrine and mesenchymal-like cells within cell lines which are overall classified with the opposite phenotype, and in which "neuroendocrine" cell lines had mostly neuroendocrine cells, but some mesenchymal, in contrast with "mesenchymal" cell lines which contained predominantly mesenchymal cells, but subpopulations of neuroendocrine.

Developing dynamical models of gene regulatory networks and signal transduction pathways promises to lead to new insights in systems level biological control and intervention. However such models often suffer from a dearth of experimental data with which to calibrate precise trajectories, partly due to the requirement of time course information for all relevant molecular species in order to determine kinetic parameters. Our work uses abundant "snapshot" data such as microarray or RNASeq along with protein interaction databases to build a coarse grain model, thereby sacrificing information on transient dynamics and focusing instead on steady state attractors. This method relies on the assumption that transcription factors function as binary units which are ON or OFF and act over similar timescales, as opposed to pathway signaling dynamics where the rate of each reaction is intimately dependent on the relative concentrations of the reactants. Since we were interested in identifying the transcriptional programs stabilizing these phenotypes, and not necessarily how cells process information through signaling pathways, we were able to safely omit signaling dynamics. Future work should unite signaling pathway models with

epigenetic regulatory models to build a more comprehensive understanding of how external cell signals can influence the internal phenotype.

One of the exciting predictions of the attractor landscape theory is that it permits cells to transition from one phenotypic attractor to another in the presence of noise or external perturbations. Because SCLC is believed to have a PNEC origin, one might ask how the mesenchymal tumor cells emerge from a neuroendocrine population. Calbo et.al were able to induce a phenotypic transition between neuroendocrine and mesenchymal states through KRAS mutation (Calbo et al., 2011), however this mutation is generally unobserved in human SCLC (Peifer et al., 2012; Rudin et al., 2012). Our model, in contrast, suggests that noise or differentiation signals could promote the transition from neuroendocrine to mesenchymal. The relatively recent work of Takahashi and Yamanaka (Takahashi and Yamanaka, 2006) has revealed the general possibility of reprogramming cell fates against the natural differentiation gradient. Many recent studies have advocated the use of attractor based models to uncover reprogramming strategies (Choi et al., 2012; Lang et al., 2014). Though we built this model in the context of SCLC, we believe that the framework is sufficiently general that similar models may be built to understand many types of heterogeneity in many different cancers. Future work should validate the ability of such models to inspire successful reprogramming strategies, ultimately paving a path for model based control of cell fate heterogeneity in cancer.

Acknowledgements

This work was supported by U54 CA113007-09 from the National Cancer Institute, Integrative Cancer Biology Program (NCI ICBP) awarded to Dr. Vito Quaranta and Grant 1I01CX000242 from the Department of Veterans Affairs CA90949 from the NCI SPORE program awarded to Dr. Pierre Massion. I would like to thank Megan Hoeksema for assistance with experiments and David Wooten for the boolean model simulations and analysis. I would like to acknowledge the Vanderbilt Flow Cytometry shared resource lab for the help and assistance in training for the use of flow cytometry instruments and the laboratory of Dr. Jonathan Irish in assistance and advice on flow cytometry data analysis. I would like to thank Dr. Charles Rudin (Memorial Sloan Kettering Center, NY) and Dr. Julien Sage (Stanford University, CA) for providing us patient-derived xenograft tumors.

CHAPTER V

DECIPHERING ROLE OF PHENOTYPIC HETEROGENEITY IN TREATMENT RESPONSE IN SCLC

Introduction

Genetic and non-genetic heterogeneity in cancer generates cells with a variety of differences in phenotypic traits such as signaling pathways, proliferation rates, invasive and metastatic ability, metabolism, drug sensitivity and resistance (Almendro et al., 2013). Mutations/amplifications in alternative signaling pathways in oncogene-addicted tumors such as c-MET amplification in EGFR-TKI resistant tumors or N-RAS mutations in TKI-resistant B-Raf mutant melanomas, lead to clonal selection and expansion of drug resistant clones. Non-genetic mechanisms of drug resistance include epithelial-to-mesenchymal transitions (EMT), mesenchymal-to-epithelial transition (MET) and cancer stem cells, mediated by large transcriptional changes in gene networks (Polyak and Weinberg, 2009). Phenotypic heterogeneity by itself measured via expression of certain phenotypic markers determines drug sensitivity (Almendro et al., 2014a). Furthermore, dynamic and reversible epigenetic changes can mediate drug resistance (Sharma et al., 2010).

We have previously uncovered two anti-correlated co-expression networks that define a heterogeneous phenotypic state space for SCLC. Within this framework, we identified a transcription factor (TF) network that dynamically regulates phenotypic heterogeneity in SCLC (Chapter IV). Briefly, the network is comprised of TF highly

represented in SCLC cell lines and patient specimens. Boolean modeling of the SCLC TF network gives rise to distinct attractor states that map to either a neuroendocrine (NE) or a mesenchymal-like (ML) differentiated phenotype that can be measured via expression of phenotypic biomarkers. TFs are known to act as master regulators of the differentiated state of a cell, especially in response to extrinsic factors in the context of tissue formation and homeostasis. The significance of our findings with the Boolean modeling of the SCLC TF network is that it defines combinations of active/inactive TFs that underlie the state of unconstrained proliferation characteristic of SCLC. Moreover, multiple SCLC proliferative states suggest the idea of phenotypic plasticity that may be responsible for escape from treatment. SCLC patients respond to combination chemotherapy (cisplatin-etoposide) but quickly develop resistance leading to fatal recurrent metastatic disease. We hypothesize that phenotypic state transitions in SCLC (mediated by changes in transcriptional regulatory networks) can be induced via drug treatment, potentially mediating drug resistance.

In this work, we aim to identify phenotypic states that might be more sensitive to drug treatment and understand phenotypic plasticity. Phenotypic plasticity in response to drug treatment can occur in two scenarios: (1) individual SCLC cells switch back and forth from NE to ML, or (2) there is a progenitor that goes in either direction with no way back.

Results

Differential sensitivity of phenotypic states to chemotherapy and epigenetic modifiers

We previously identified 2 phenotypic states or transcriptional subtypes in SCLC cell lines and patients - neuroendocrine (NE) and mesenchymal-like (ML). To test if the 2 phenotypic states are differentially sensitive to cisplatin and etoposide, we compared the response of 8 SCLC cell lines to increasing doses of the 2 chemotherapy agents for 5 days. At 5 days, viability was measured using hoescht-calcein-ethidium homodimer dyes (which label total nuclei, viable and dead cells respectively) and imaged on the Cellavista imager. Some of the cell lines proliferate at doubling times up to 40 hours, so earlier time points did not show an effect of the drug (data not shown). To account for differences in proliferation rates, we normalized the Day 5 viable cell counts to Day 0 counts, which also helps us distinguish between cytostasis (cell count above 0) versus cell death (cell count below 0). We found that NE cell lines exhibit more cell death at higher doses of cisplatin and etoposide while the ML cell lines show a cytostatic effect at higher doses of the drug (Figure 5.1A and B). The NE cell lines are more sensitive to cisplatin (lower IC50s) than ML lines, but both have similar high sensitivity to etoposide (Figure 5.1C and D). These results corroborate with the fact that most SCLC patients are initially highly sensitive to cisplatin-etoposide combination.

Epigenetic modifying drugs play an important role in non-genetic heterogeneity and modification of transcriptional network states, and also are important in eradication of drug-resistant clones (Sharma et al., 2010). So we tested if the SCLC cell lines would be sensitive to epigenetic modifiers. The NE and ML cell lines were very sensitive to

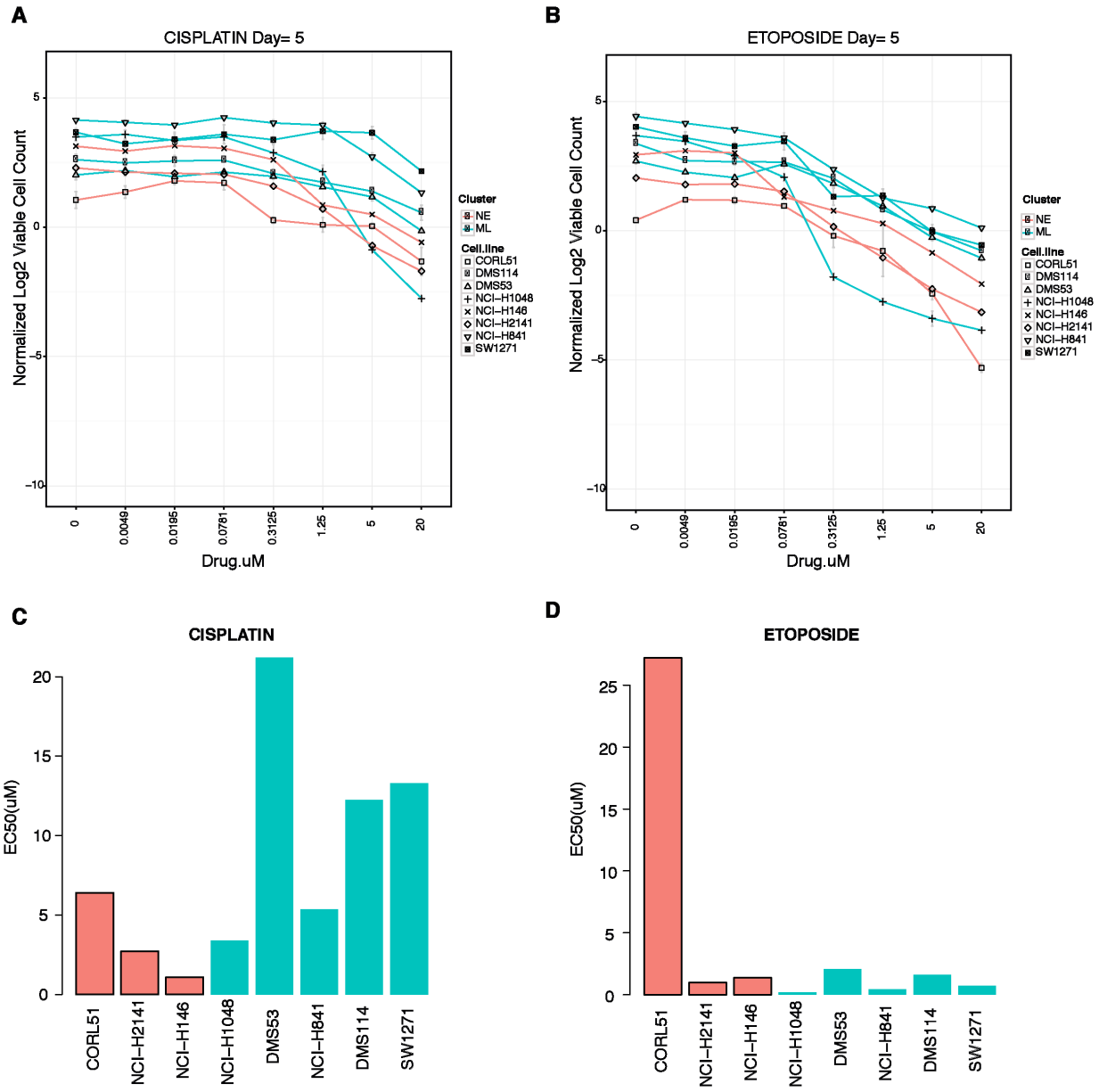


Figure 5.1. Differential sensitivity of neuroendocrine and mesenchymal cell lines to chemotherapy - cisplatin and etoposide.

(A) and (B) show growth inhibition curves of 8 SCLC cell lines to increasing concentrations of cisplatin and etoposide respectively, five days post treatment. Log2 normalized viable cell counts (X-axis) are calculated by normalizing the data to initial Day 0 cell counts to account for differences in proliferation rates of cell lines. Points below zero indicate that cells are dying in response to the drug. (C) and (D) are barplots of IC50 values calculated by fitting nonlinear regression models in R.

histone deacetylase HDAC inhibitors Valproic acid (VPA) and Trichostatin A (TSA) exhibiting high cell death (Figure 5.2A and B), while showing no response to demethylating agent Decitabine (5-azadeoxycytidine, DEC) (Figure 5.2C).

Multidimensional flow cytometry data reduction algorithms validate inter- and intra-tumor heterogeneity in SCLC cell lines.

We previously identified surface markers, kinases and adhesion molecules that would be representative of neuroendocrine/epithelial and mesenchymal-like differentiation. CD56/NCAM1, CD24 and CADM1/SynCAM1 are key markers of neuronal/neuroendocrine differentiation while CD44, CD151, EPHA2 and Zyxin are mesenchymal-like phenotype markers (Figure 4.11). A 2D view of NE and ML subpopulations is provided by a CD56 (NE marker) versus a CD151 (ML marker) (Figure 4.12 and 5.3). The quadrant gate was drawn based on unstained controls . Here, NE subpopulation is defined as CD56+CD151-, while ML population is defined by CD56-CD151+ (Figure 5.3A and B - top panels). The NE and ML cell lines did show the presence of a CD56+CD151+ double positive (DP) subpopulation, at varying proportions. Particularly some ML cell lines (defined by consensus clustering) showed a strong enrichment for CD56+CD151+ 'intermediate' subpopulations, but if this is a truly intermediate or a stem-cell like state remains to be seen. In terms of long-term stability of these states i.e. NE (CD56+CD151-), ML (CD56-CD151+) and intermediate (CD56+CD151-) subpopulations, we found that the relative proportions of these subpopulations remained the same in SCLC cell lines over time, indicating that these are in fact stable states in exponential growth culture conditions. We did not find any markers that were specific for the intermediate subpopulations.

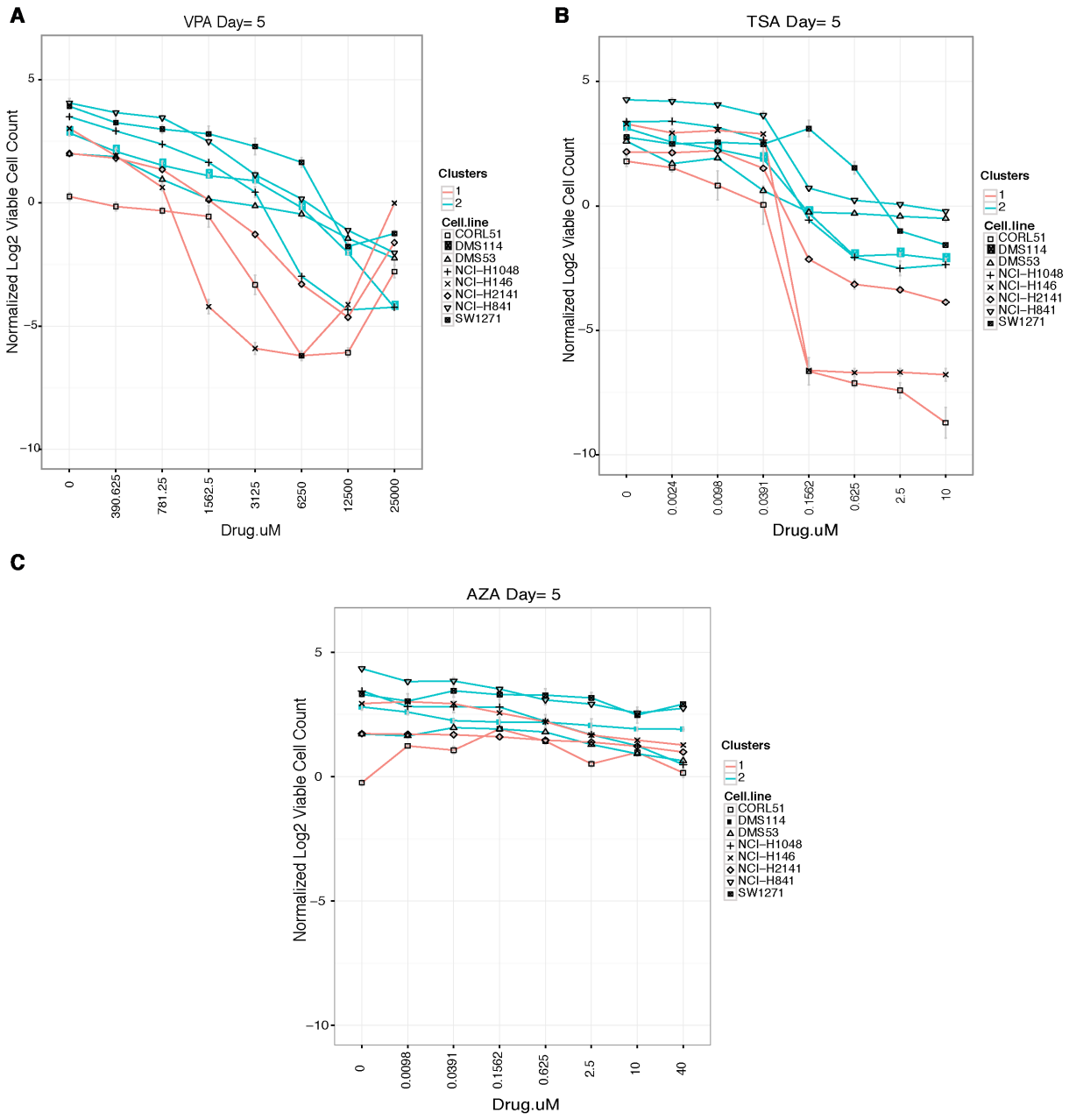


Figure 5.2. Response of NE and ML cell lines to epigenetic modifying agents.

(A) - (C) Growth inhibition curves of 8 SCLC cell lines to increasing concentrations of HDAC inhibitors (VPA and TSA respectively) and demethylating agent 5-azacytidine, five days post treatment. Log2 normalized viable cell counts (X-axis) are calculated by normalizing the data to initial Day 0 cell counts to account for differences in proliferation rates of cell lines. Points below zero indicate that cells are dying in response to the drug.

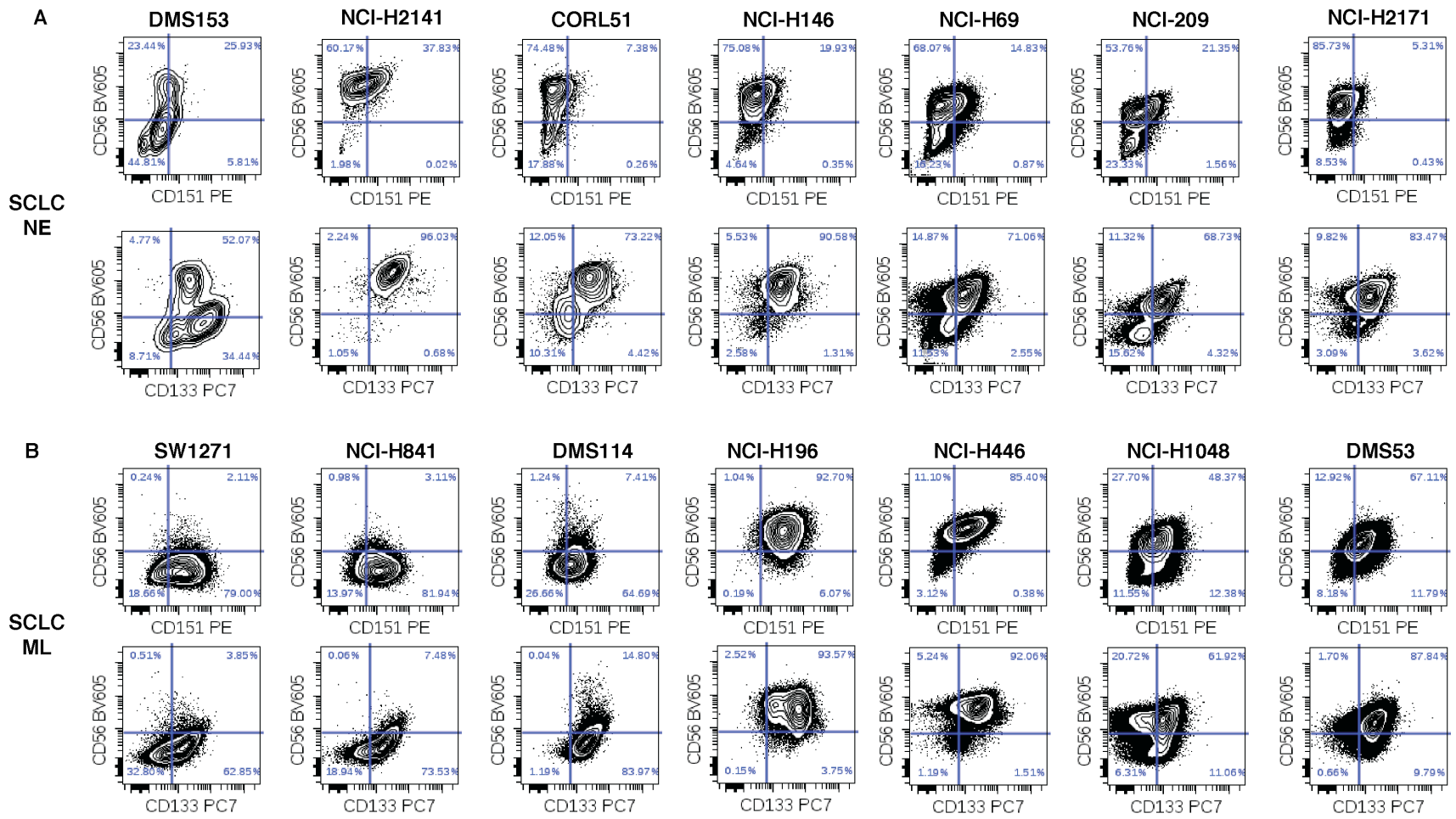


Figure 5.3. Establishment of a 2D view of phenotypic attractor states in SCLC using flow cytometry data.

Figure 5.3. Establishment of a 2D view of phenotypic attractor states in SCLC using flow cytometry data.

Biaxial contour plots of (A) NE marker CD56 versus ML marker CD151 (B) CD56 versus stem cell marker CD133 and (C) CD56 versus stem cell marker CD133, across 7 NE and ML cell lines. Percentages denote the the density of cells present in each quadrant of the gate. NE cell lines mainly consist of CD56+CD151- and a small percentage of CD56+CD151+ DP subpopulations. ML cell lines were more heterogeneous where some cell lines contained majority CD56-CD151+ subpopulations while others contained CD56+CD151+ DP subpopulations.

CD133, a cancer stem-cell marker, was highly expressed in both CD56+ and CD151+ populations in SCLC cell lines, suggesting that every subpopulation potentially possessed stem-like properties (Figure 4.12C, Figure 5.3A and B - bottom panels).

Using these markers, we determined the percentage of NE and ML subpopulations within each SCLC cell line (Figure 4.12) using k-means clustering analysis. Recently, several methods (k-means clustering, viSNE and SPADE) are used to study multidimensional flow cytometry datasets to understand the heterogeneity within a sample as well as to study relationships across samples at a single cell resolution. These are useful tools to study common and distinct subpopulations between samples and track phenotype of drug resistant/relapsed and rare populations over time (Amir et al., 2013; Simonds et al., 2011).

With the idea of using these methods to study and track subpopulations and plasticity upon treatment, we applied SPADE analysis to our multidimensional flow cytometry dataset consisting of 3 NE (CD56, CD24, CADM1) and 3 ML (CD151, CD44, EPHA2) markers of 22 SCLC cell lines and 2 PDX. SPADE analysis performs unsupervised clustering analysis of a multidimensional flow cytometry dataset and builds a minimum spanning tree (as nodes and branches) where nodes represent clusters of cells and the distance between the nodes is given by the branches (Figure 2.3). Clusters of nodes that are near one another represent cells that would be co-expressing various markers at a certain level. By assessment of the co-expression of NE markers versus ML markers in each cell line, we manually gated the tree as containing NE defined by high co-expression of NE markers, and low ML markers and vice versa for ML subpopulations. Representative examples of a NE (NCI-H146) and

ML (SW1271) cell line shown in Figure 5.4 and 5.5 respectively for co-expression of NE and ML markers. Thus these markers serve to define the features of NE and ML attractor states. To further verify the overlap of our manual gating of NE and ML subpopulations with previously established k-means clusters (Figure 4.12), we overlaid the NE and ML cluster information of individual cells from the k-means clustering analysis on the SPADE tree (Figure 5.6). 3 representative NE and 3 ML cell lines are shown in Figure 5.6 and the k-means cluster overlay confirmed that our manual gating was valid. Based on the SPADE analysis, we find distinct distributions of NE and ML subpopulations in various SCLC cell lines (Figure 5.7). The SPADE data suggest that there is further heterogeneity in the cells within a NE or ML cluster (given by k-means clustering). This analysis will help better track subpopulation 'movement' in response to treatment. These above described manual gates on the SPADE plots will be used for all future drug treatment analysis.

Drug treatment induces transitions between stable SCLC phenotypic states.

Attractor state landscapes define basins within which a phenotype will remain stable, however, extrinsic factors may lower transition barriers between basins or destabilize them entirely. To explore the plasticity of NE and ML states, we compared the response of a panel of 11 SCLC cell lines to 5-day treatment of cisplatin, etoposide, demethylating agent azacitidine, or HDAC inhibitor valproic acid (VPA) ordered from ME-Blue high (CORL51) to ME-Turquoise high (SW1271) expression.

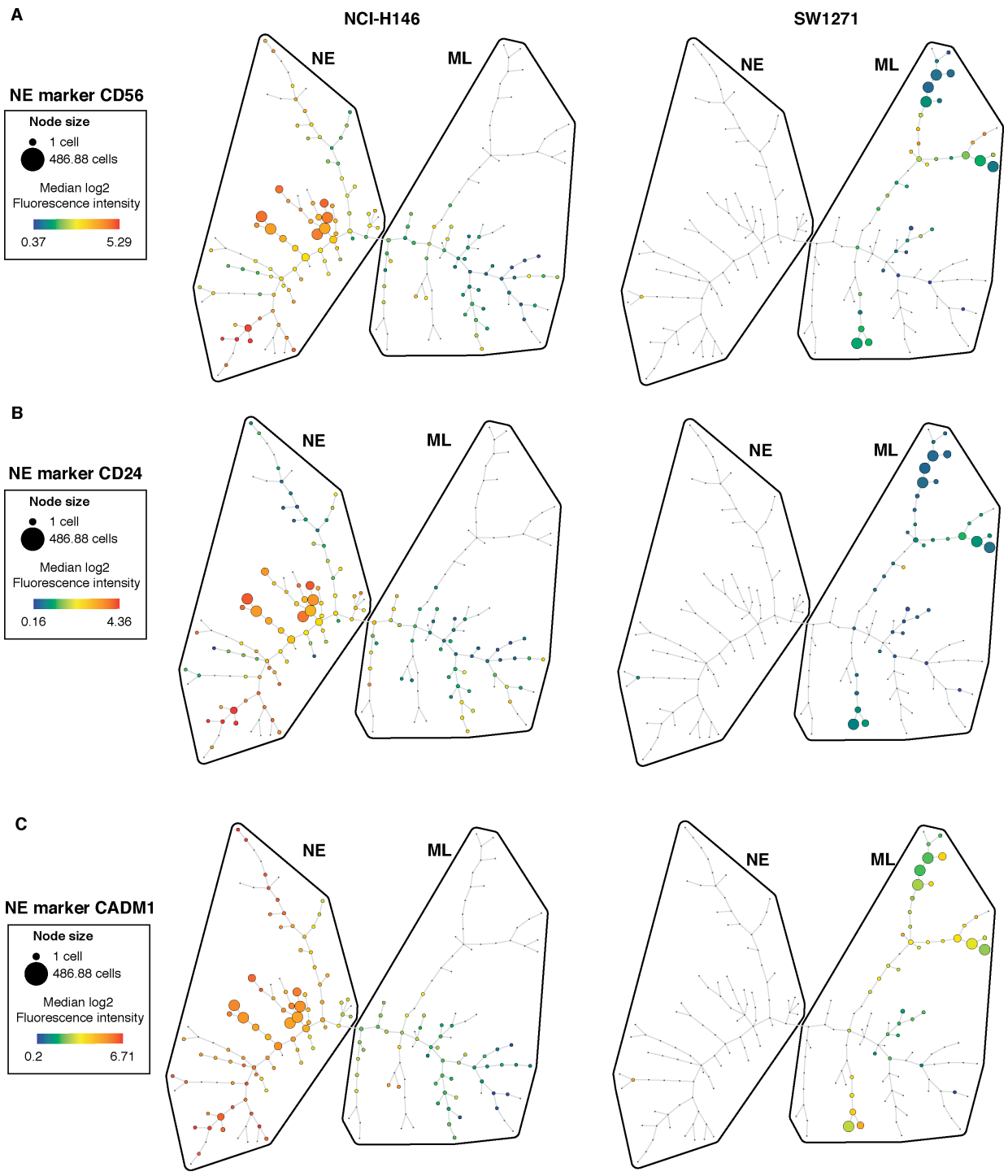


Figure 5.4. SPADE analysis of NE and ML subpopulations based on co-expression of NE markers.

Figure 5.4. SPADE analysis of NE and ML subpopulations based on co-expression of NE markers. This figure shows a minimum spanning tree given by SPADE analysis of a multidimensional flow cytometry dataset, composed of 22 SCLC cell lines spanning the 2 subtypes. 3 NE (CD56, CD24, CADM1) and 3 ML (CD44, CD151, EPHA2) markers were included in the analysis. Nodes denote clusters of cells grouped by unsupervised clustering and the branches denote the similarity between the nodes. Size of the node denotes the density of cells present in each cluster. Manual gating of NE and ML phenotypic states was done in Cytobank using the co-expression of NE or ML markers across the 22 cell lines. Log2 median intensity of expression (denoted by color - high given by red, low by blue) overlaid on 2 representative cell lines NCI-H146 (NE) and SW1271 (ML) are shown for 3 NE markers (A) CD56 , (B) CD24 and (C) CADM1.

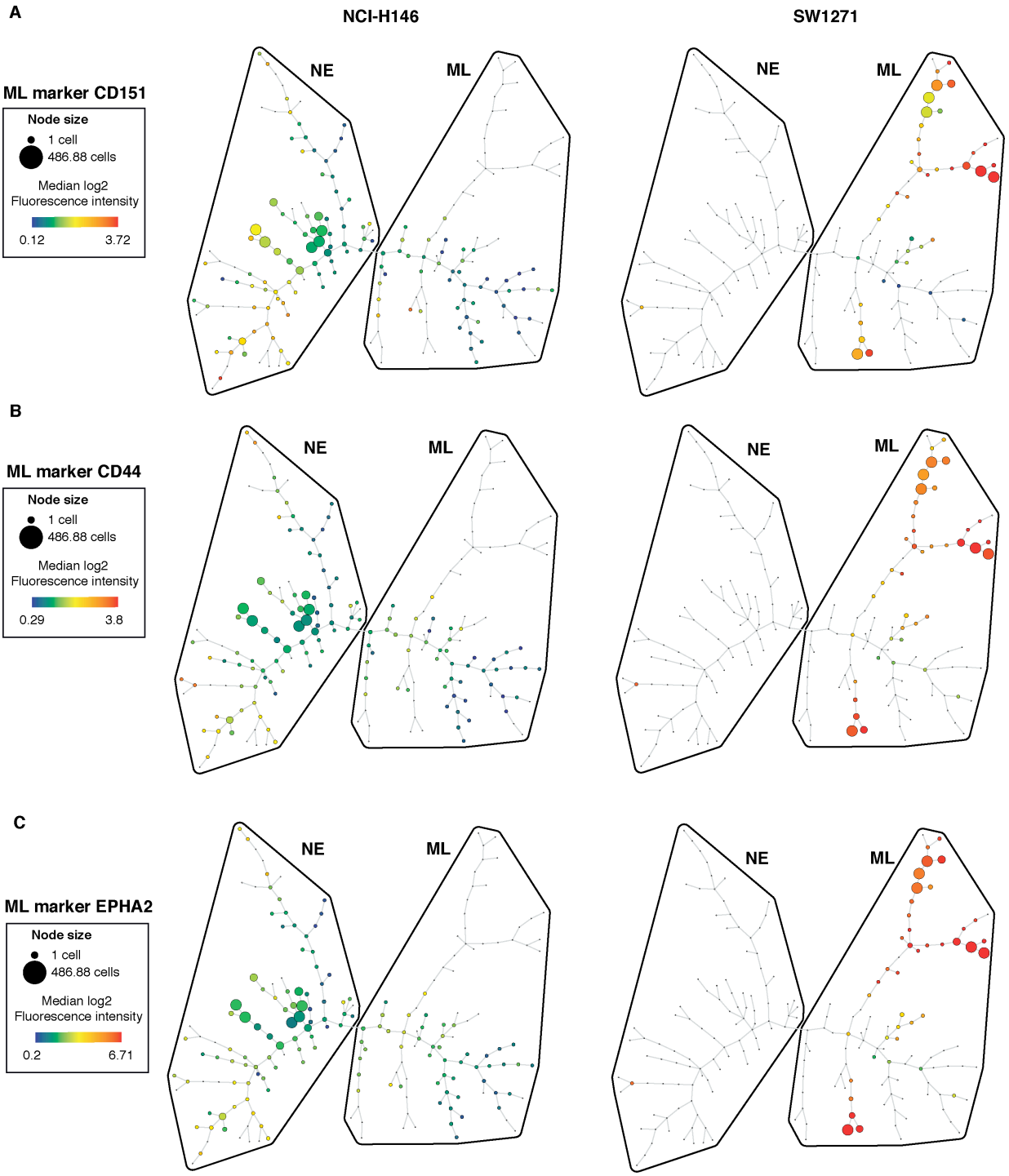


Figure 5.5. SPADE analysis of NE and ML subpopulations based on co-expression of ML markers.

Figure 5.5. SPADE analysis of NE and ML subpopulations based on co-expression of ML markers. This figure shows a minimum spanning tree given by SPADE analysis of a multidimensional flow cytometry dataset, composed of 22 SCLC cell lines spanning the 2 subtypes. 3 NE (CD56, CD24, CADM1) and 3 ML (CD44, CD151, EPHA2) markers were included in the analysis. Nodes denote clusters of cells grouped by unsupervised clustering and the branches denote the similarity between the nodes. Size of the node denotes the density of cells present in each cluster. Manual gating of NE and ML phenotypic states was done in Cytobank using the co-expression of NE or ML markers across the 22 cell lines. Log2 median intensity of expression (denoted by color - high given by red, low by blue) overlaid on 2 representative cell lines NCI-H146 (NE) and SW1271 (ML) are shown for 3 ML markers (A) CD151 , (B) CD44 and (C) EPHA2.

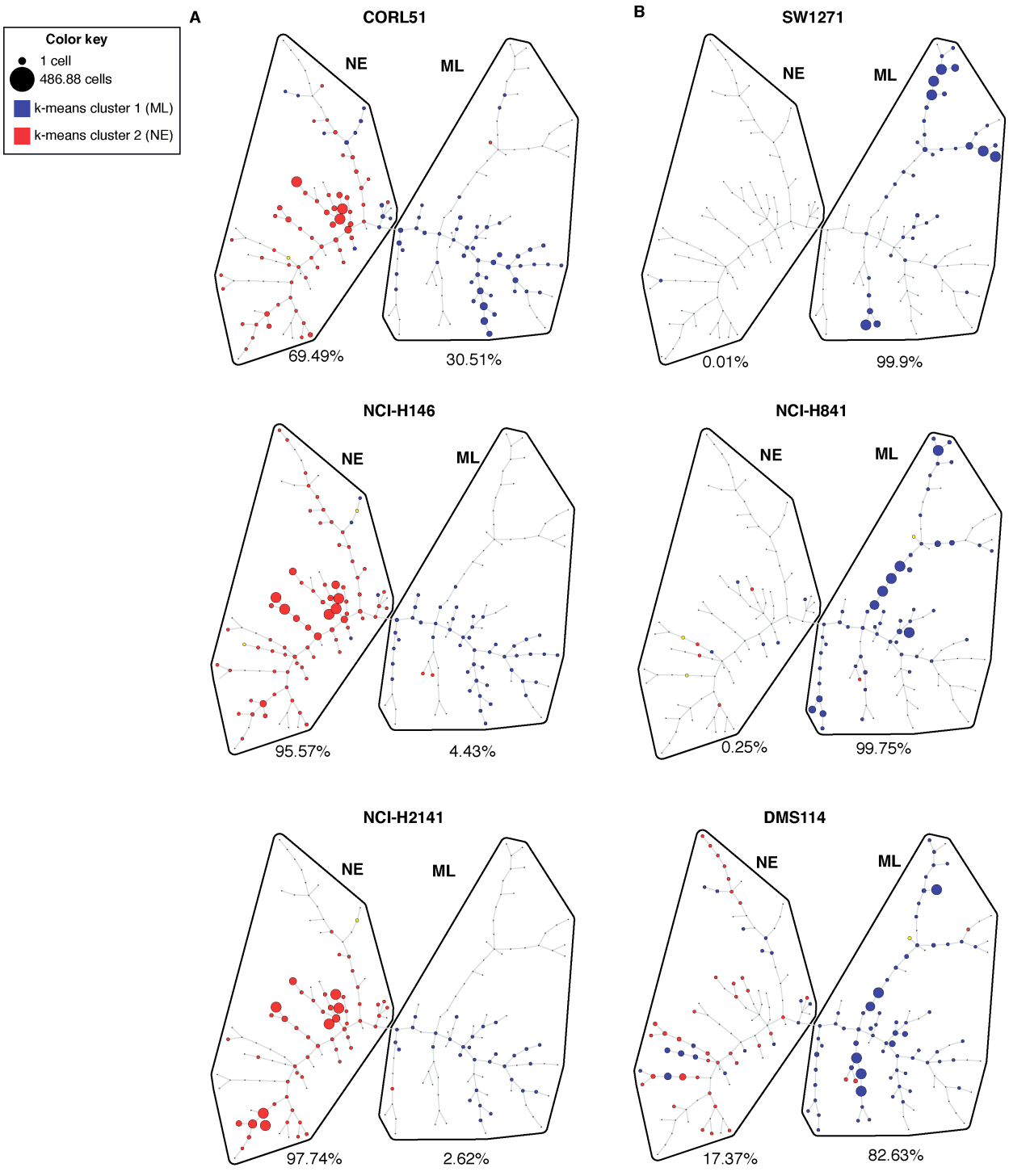


Figure 5.6. Independent validation of k-means clusters on SPADE tree.

Figure 5.6. Independent validation of k-means clusters on SPADE tree.

This figure shows a minimum spanning tree given by SPADE analysis of a multidimensional flow cytometry dataset, composed of 22 SCLC cell lines spanning the 2 subtypes. 3 NE (CD56, CD24, CADM1) and 3 ML (CD44, CD151, EPHA2) markers were included in the analysis. Nodes denote clusters of cells grouped by unsupervised clustering and the branches denote the similarity between the nodes. Manual gating of NE and ML phenotypic states was done in Cytobank using the co-expression of NE or ML markers across the 22 cell lines. The k-means clusters (k=2) established in Chapter IV (Figure 4.12) are overlaid here on the SPADE plots to verify the manual gating of NE and ML clusters. The cellular distribution of 3 representative NE and ML cell lines (denoted by the size of the nodes) are shown in (A) and (B) respectively with overlay of their respective k-means clusters (denoted by colors - red cluster 2/ NE and blue - cluster 1/ML).

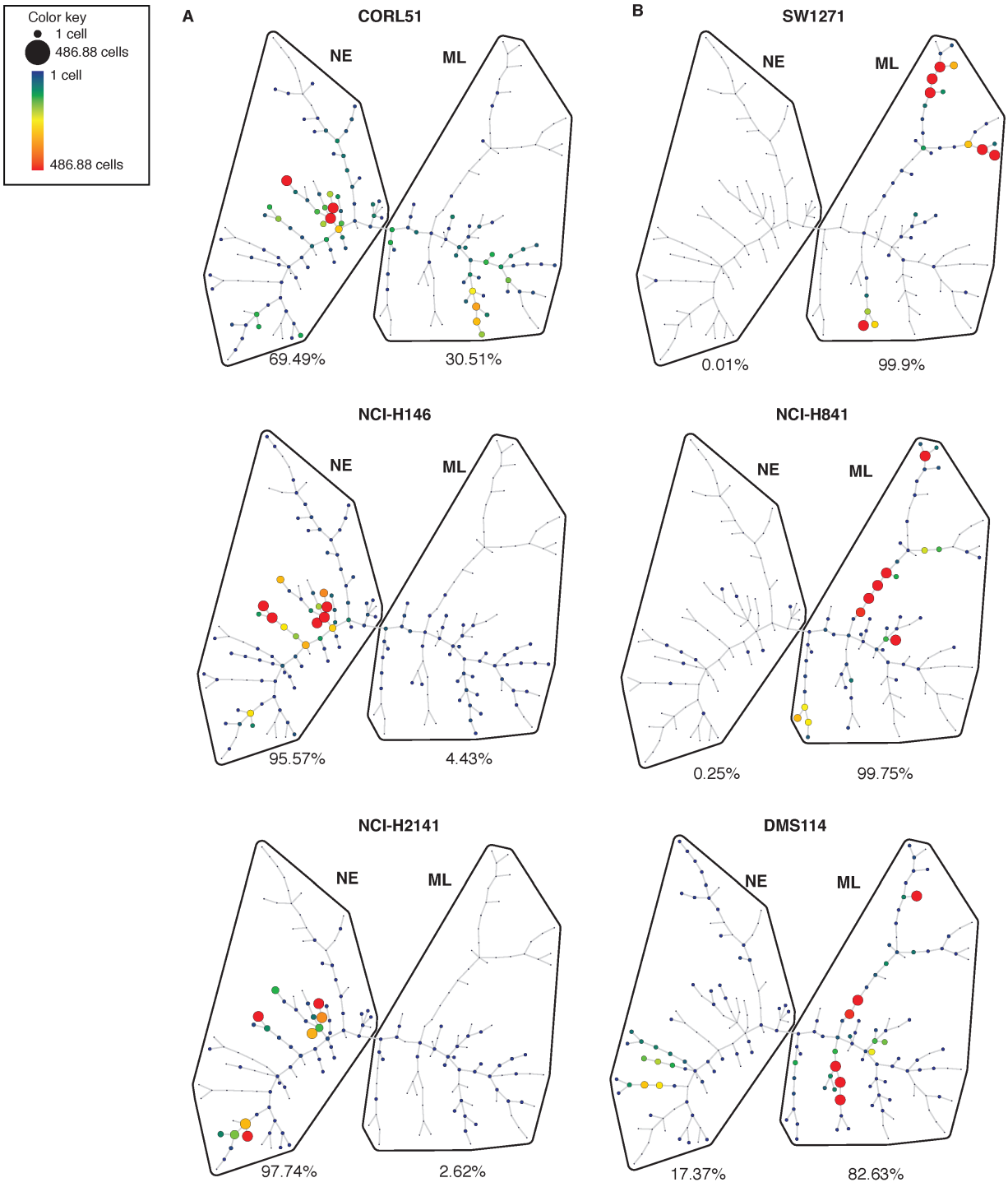


Figure 5.7. Density distribution of NE and ML cell lines on the SPADE tree.

Figure 5.7. Density distribution of NE and ML cell lines on the SPADE tree.

This figure shows a minimum spanning tree given by SPADE analysis of a multidimensional flow cytometry dataset, composed of 22 SCLC cell lines spanning the 2 subtypes. 3 NE (CD56, CD24, CADM1) and 3 ML (CD44, CD151, EPHA2) markers were included in the analysis. Nodes denote clusters of cells grouped by unsupervised clustering and the branches denote the similarity between the nodes. Manual gating of NE and ML phenotypic states was done in Cytobank using the co-expression of NE or ML markers across the 22 cell lines. The k-means clusters (k=2) established in Chapter IV (Figure 4.12) are overlaid here on the SPADE plots to verify the manual gating of NE and ML clusters. The cellular distribution of 3 representative NE and ML cell lines is denoted by the size of the nodes and the color.

Firstly, we observed that NE cell lines exhibited a much higher basal rate of cell death (measured by increase in A700 staining) than ML cell lines in normal cell culture conditions (Figure 5.8A). This was further increased upon drug treatment with either cisplatin, etoposide or VPA (Figure 5.8B). Azacitidine had a only modest increase in death in some cell lines. Interestingly, all SCLC cell lines were very sensitive to 5mM valproic acid with 75-90% cell death in 5 days. It is important to note that IC50s of SCLC cell lines upon 5-day VPA treatment (Figure 5.2) ranged from 2-8mM, so 5mM was within the range of dose sensitivity. VPA might potentially be an important target for SCLC to be pursued in the future via *in vivo* drug treatment.

Flow cytometry 2D plots can be used to visualize stable attractor states with a specific configuration of phenotypic markers (Figure 4.12) as well as to assess state transitions (Ho et al., 2012). State transitions in NE and ML cell lines can be described as gaining expression (>5%) of ML or NE markers respectively upon treatment, indicated by the red arrows (Figure 5.9). NE and ML cell lines showed differential susceptibility to death and state transitions in response to 5 days of drug treatment. NE and ML cell lines both showed an a significant increase in CD56+CD151+ DP subpopulations via gaining expression of CD151 and CD56 respectively. VPA induced a drastic change in ML phenotype with almost 75-85% increase in CD56+CD151+ DP subpopulations. Similarly, cisplatin and etoposide induced a drastic 25-50% increase in CD56+CD151+ DP in NE cell lines while VPA only showed a 15% increment in DP subpopulation but also increased CD56-CD151- and CD56-CD151+ subpopulations (particularly in NCI-H146) (Figure 5.9).

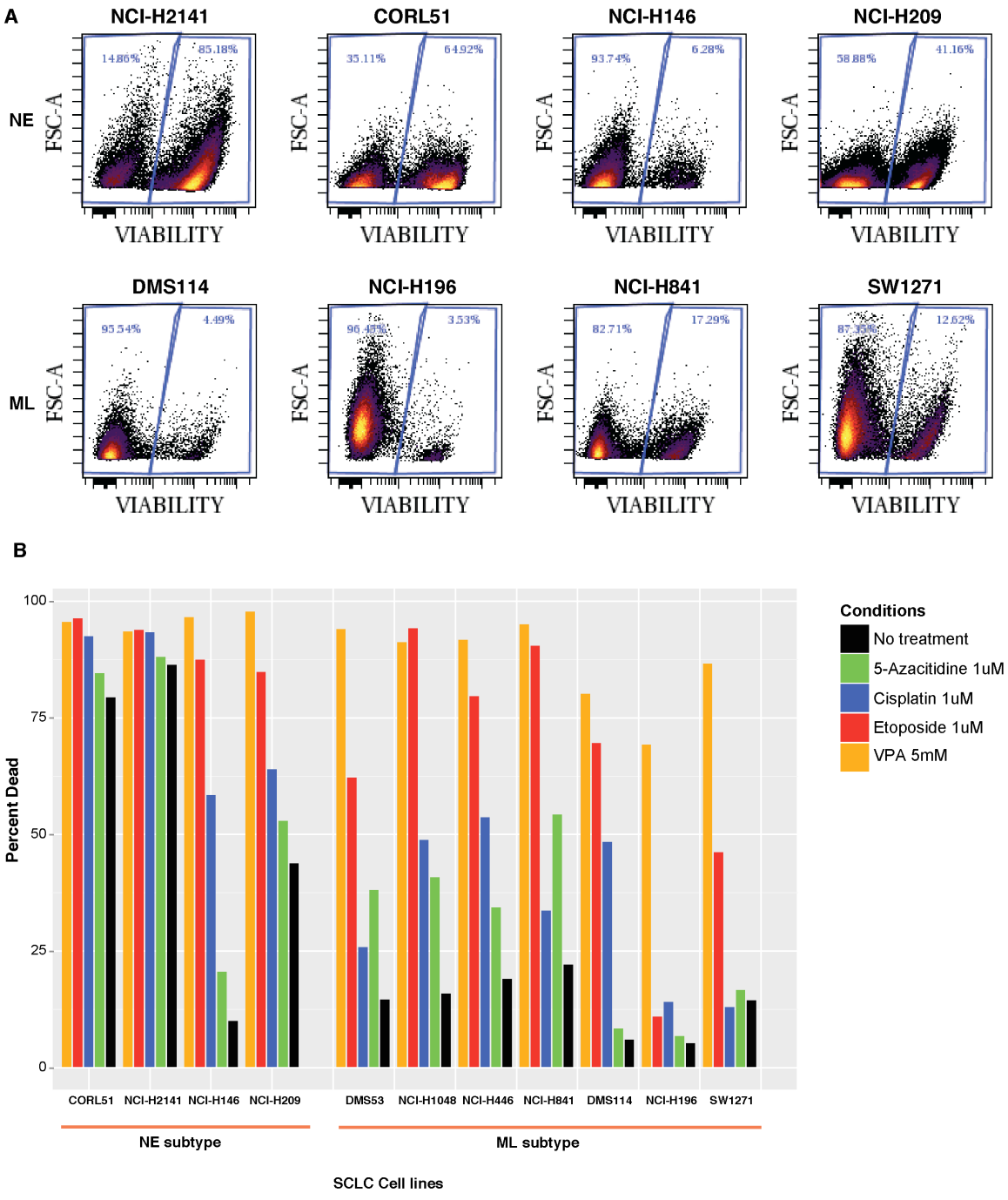


Figure 5.8. Basal and drug-induced cell death analysis in SCLC cell lines.

Figure 5.8. Basal and drug-induced cell death analysis in SCLC cell lines.

(A) Biaxial plot of Alexa700 (viability dye) versus Forward scatter area (FSC-A) for 4 NE and ML cell lines. Left and right gates denote percentage of viable and dead/dying cell populations respectively in basal i.e. no treatment conditions. NE cell lines display higher percentage of dead cells than ML cell lines in normal culture conditions suggesting spontaneous apoptosis.

(B) Barplot of percent dead (A700 positive) populations in each cell line with no treatment or upon 5-day treatment with either cisplatin (1uM), etoposide (1uM), valproic acid (VPA) 5mM and 5-azacytidine (1uM). These drug concentrations were chosen based on the range of IC50s (Figure 5.1 and 5.2). VPA and etoposide cause pronounced cell death in all SCLC cell lines.

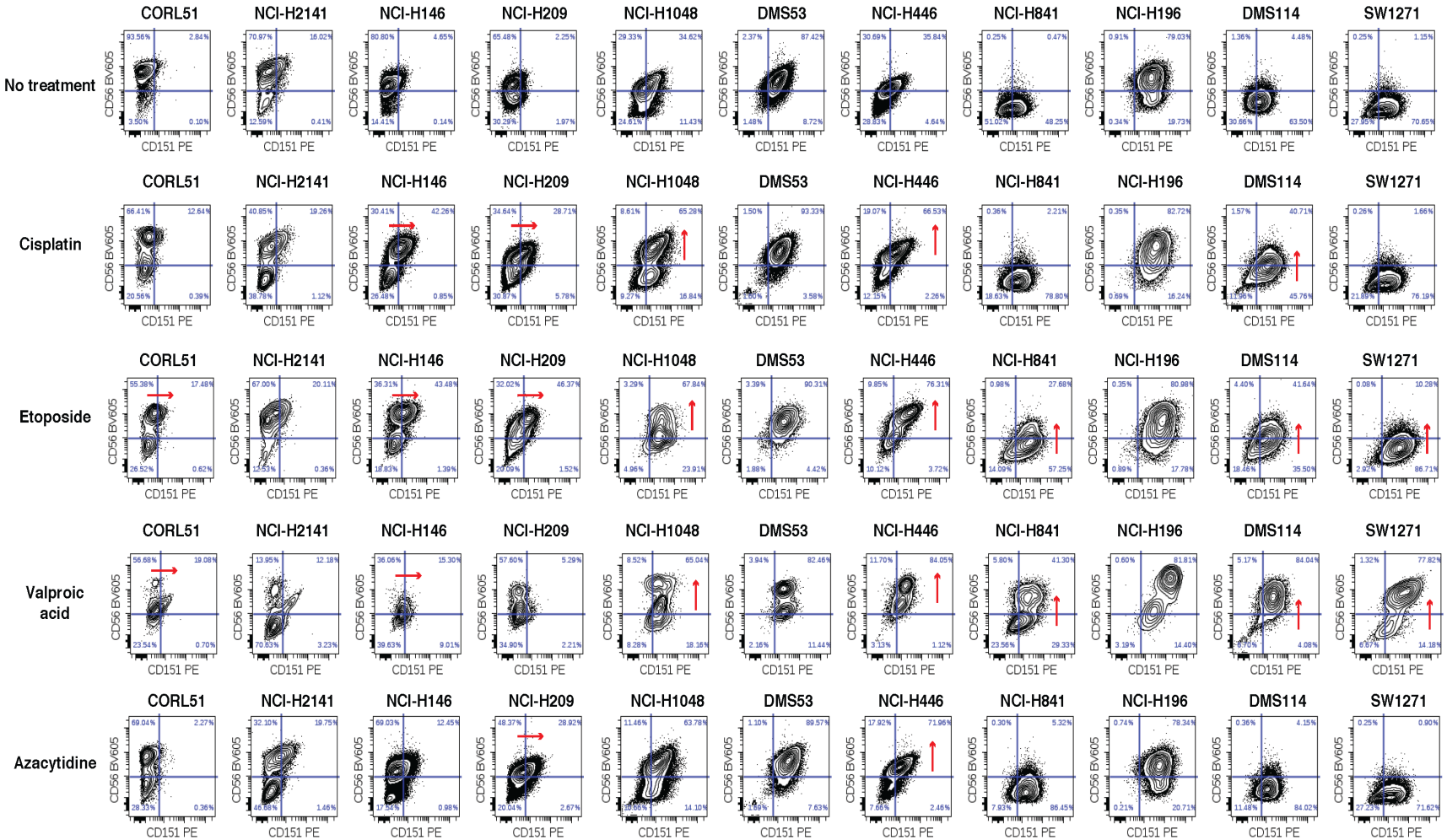


Figure 5.9. 2D view of phenotypic plasticity upon treatment in SCLC cell lines.

Figure 5.9: 2D view of phenotypic state transitions upon treatment in SCLC

This figure shows biaxial plots of NE marker CD56 and ML marker CD151 for representative NE, NM and ML cell lines under no treatment or 5-day treatment with either cisplatin, etoposide, 5-azacitidine or valproic acid. State transitions are denoted by the arrows showing > 5-10% shift from no treatment condition.

To assess global changes in phenotypic states upon treatment, we performed SPADE analysis of the 11 SCLC cell lines under various drug treatment conditions with the same 6 phenotypic markers (3 NE – CD56, CD24 and CADM1, and 3 ML surface markers – CD44, CD151 and EPHA2). Figure 5.10 shows SPADE plots for various drug treatments for 1 representative NE and ML cell line. These data indicate that changes are induced in other 4 markers as well for the cells to be different enough to be now part of a different gate. Percentages below each plot denote the percent of cells in the NE and ML gates within the SPADE plots. In the no treatment condition, the majority of NE, ML, and NM cells fall into their respective cluster (Figure 5.10A and B), however after treatment, cell lines exhibited a drug/phenotype dependent plasticity. Cisplatin had an effect on NE cell line NCI-H146, transitioning cells into the ML clusters, but no effect on ML (Figure 5.10A, top middle panels). Etoposide had a major effect on the NE cell line, but none on ML (Figure 5.10A and B, top right panels). VPA had a major impact on all cell lines, causing NE cell lines to enter the ML state, while making the ML transition to a more NE state, suggesting that broad epigenetic modifications may be able to upset the regulatory balance maintaining cells in attractor basins (Figure 5.10A and B, lower left panels). Conversely Azacitidine, a histone demethylating agent, has little or no effect on any of the cell lines (Figure 5.10A and B, lower right panels).

Epigenetic plasticity reverses state transitions upon drug removal

SCLC plasticity could potentially be induced via reversible phenotypic state transitions or permanent changes in phenotypic states. To test if the state transitions

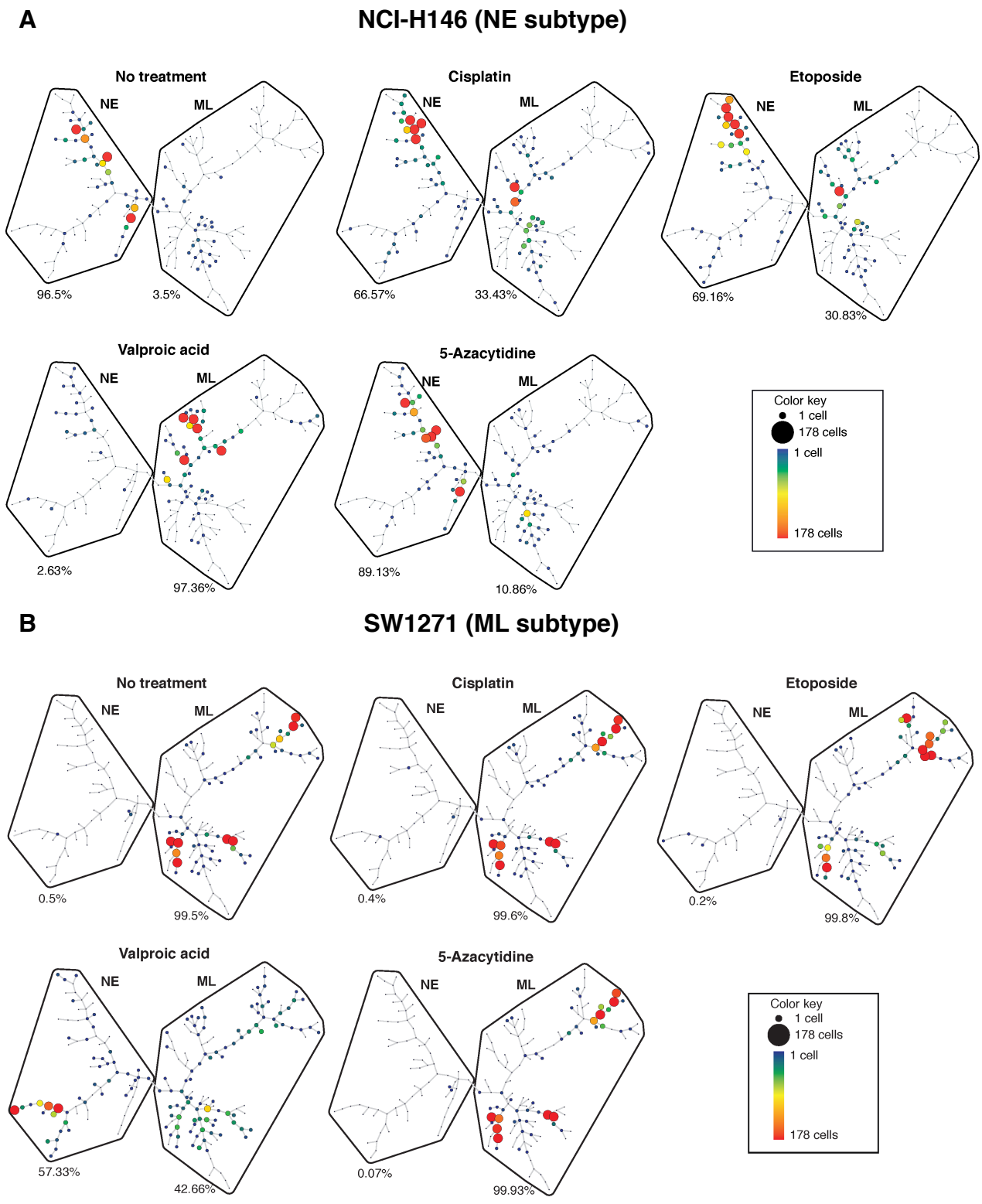


Figure 5.10. SPADE analysis based assessment of phenotypic state transitions upon drug treatment in SCLC cell lines.

Figure 5.10. SPADE analysis based assessment of phenotypic state transitions upon drug treatment in SCLC cell lines.

This figure shows a minimum spanning tree given by SPADE analysis of a multidimensional flow cytometry dataset, composed of 11 SCLC cell lines (Figure 5.9) spanning the 2 subtypes under various drug conditions (no treatment, cisplatin 1uM, etoposide 1uM VPA 5mM, and 5-azacytidine 1uM) treated for 5 days. 3 NE (CD56, CD24, CADM1) and 3 ML (CD44, CD151, EPHA2) markers were included in the analysis. Nodes denote clusters of cells grouped by unsupervised clustering and the branches denote the similarity between the nodes. Manual gating of NE and ML phenotypic states was done in Cytobank using the co-expression of NE or ML markers across the 12 cell lines (as shown in Figure 5.4 and 5.5). The cellular distribution of 1 representative NE and ML cell lines is denoted by the size of the nodes and the color. Percentages denote the number of cells within each gate given the total number of cells in each sample.

were reversible (implying dynamic epigenetic changes) or permanent (selection of a drug resistant subpopulation), we treated 6 SCLC cell lines with 0.5uM etoposide and 1.25mM VPA for either (1) 14 days with addition of fresh drug on day 7 or (2) 7 days with drug and allowed to rebound for 7 days without drug. The concentration was reduced from previous experiment so as to obtain sufficient number of cells for flow experiments since even 5-day treatment of the cells was inducing massive cell death (Figure 5.8B). We observed similar state transitions upon drug treatment as seen in Figure 5.9 and 5.10, although at a lower efficiency possibly due to lower drug dosage (Figures 5.11). In the 3 NE cell lines tested, drug-induced state transition upon withdrawal was not reversible, at least not in the timeframe of measurement (7 days without drug) (Figures 5.11A). In fact, the percentage of cells in the ML state further increased in rebound than treated samples.

In contrast, ML cell lines showed partial to almost complete reversal of the ML--> NE state transitions upon drug withdrawal (Figures 5.11B). These data suggest that NE-->ML transition might be relatively permanent while ML-->NE transition is temporary, implying differential stabilities of the 2 states in these SCLC subtypes.

Discussion

SCLC patients are treated with standard of care chemotherapy – cisplatin and etoposide combination, where initial response rates are upto 80% but invariably tumor recurrence leads to fatal metastasis (Demedts et al., 2009; Rosti et al., 2006).

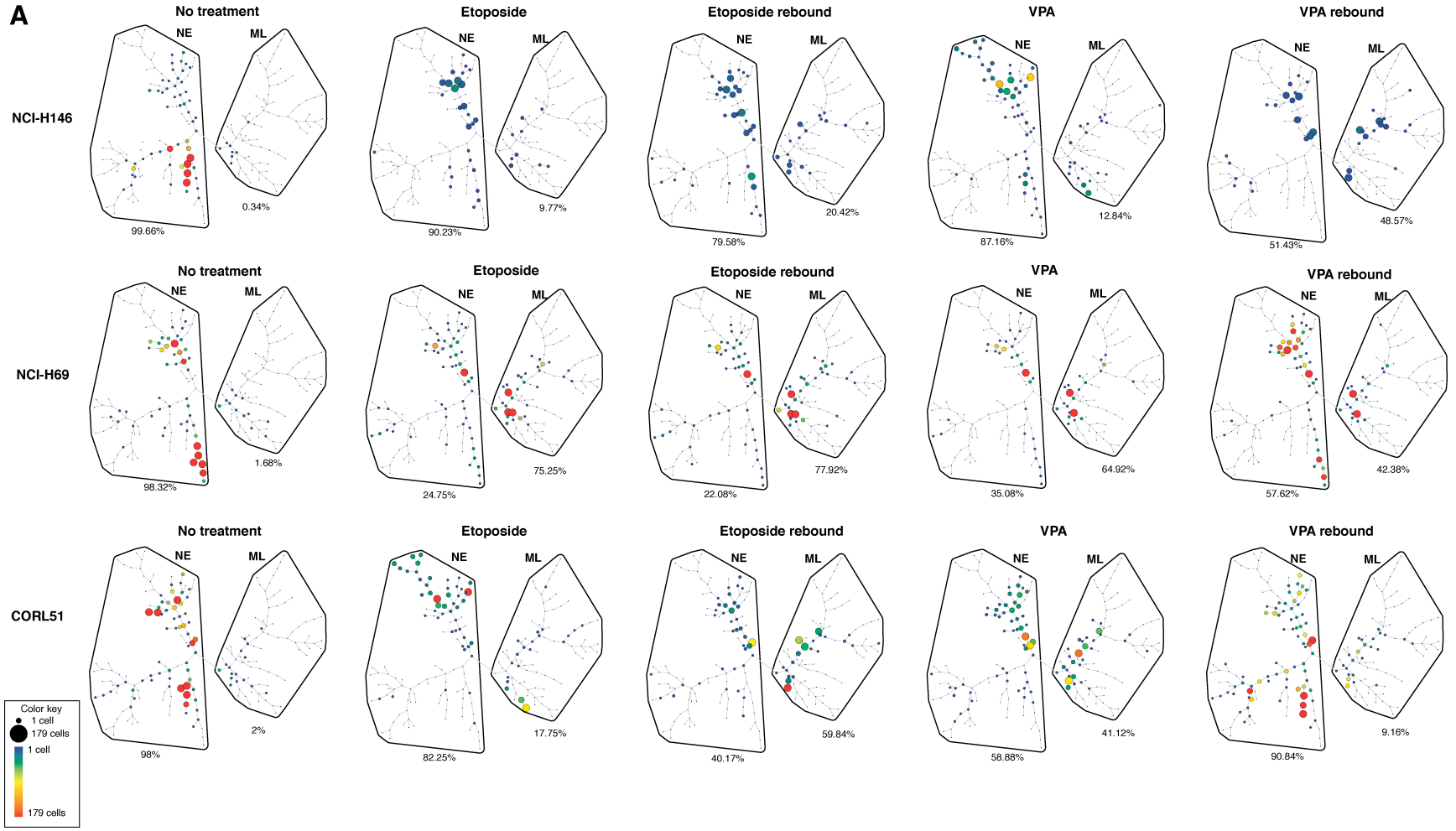


Figure 5.11 : SPADE analysis of long-term drug treatment and rebound in NE and ML cell lines.

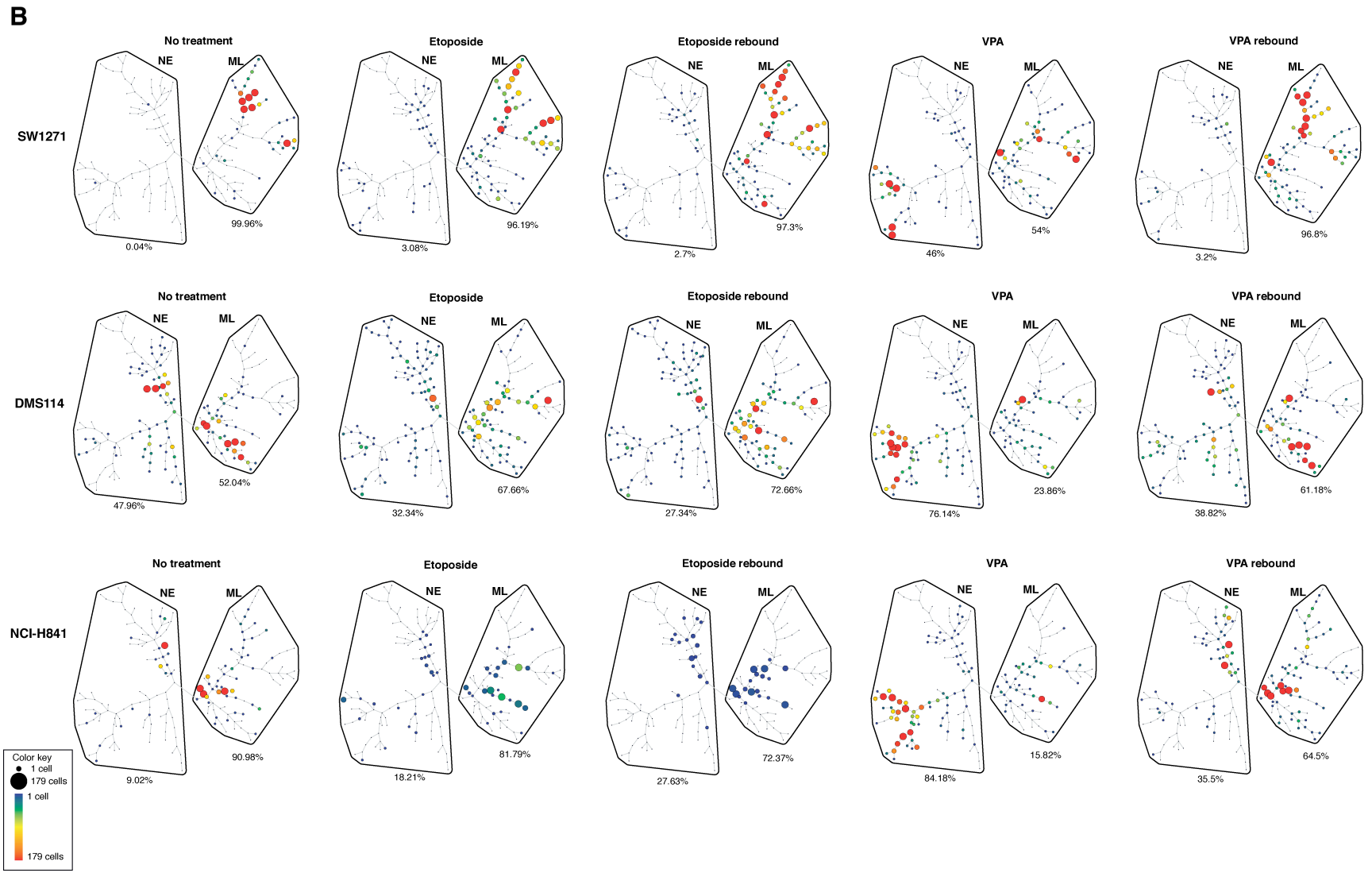


Figure 5.11 (continued): SPADE analysis of long-term drug treatment and rebound in NE and ML cell lines.

Figure 5.11 : SPADE analysis of long-term drug treatment and rebound in NE and ML cell lines.

This figure shows a minimum spanning tree given by SPADE analysis of a multidimensional flow cytometry dataset, composed of 6 SCLC cell lines spanning the 2 subtypes under various drug conditions (no treatment, 0.5uM etoposide or 1.25mM VPA) treated for (1) 14 days with fresh drug addition at day 7 or (2) 7 days treatment followed by rebound (without drugs) for additional 7 days. 3 NE (CD56, CD24, CADM1) and 3 ML (CD44, CD151, EPHA2) markers were included in the analysis. Nodes denote clusters of cells grouped by unsupervised clustering and the branches denote the similarity between the nodes. Manual gating of NE and ML phenotypic states was done in Cytobank using the co-expression of NE or ML markers across the 8 cell lines (as shown in Figure 5.4 and 5.5). The cellular distribution of 1 representative NE and ML cell lines is denoted by the size of the nodes and the color. Percentages denote the number of cells within each gate given the total number of cells in each sample. (A) and (B) show SPADE plots for 3 NE and ML cell lines respectively.

Cellular phenotypic state dynamics is potentially an important factor determining drug response and resistance. Hence, studying the population dynamics at a single cell level is essential to follow the tumor response over time.

Our work strongly suggests that SCLC cell lines, specifically the NE subtype, exhibits phenotypic plasticity in response to standard of care chemotherapy (cisplatin and etoposide) and has relatively permanent transition from NE--> ML even after withdrawal of the drug. On the other hand, ML cell lines are relatively resistant to cisplatin, and don't exhibit state transitions to NE with chemotherapy. Valproic acid, a HDAC inhibitor, induced significant cell death in most, if not all, SCLC cell lines and drastic phenotypic state transitions from NE-->ML and ML--> NE. If the drug-induced transition was due to selection of resistant clones, there would be two possibilities : (1) the resistant clones possess high proliferation rates to be able to increase in cell density (2) the resistant clones remain unchanged upon treatment withdrawal. Based on our current knowledge of proliferation rates of SCLC cell lines (doubling time 28-50 hours), we couldn't possibly obtain such a large increase in drug-resistant clones at the end of 5 days. In addition, we observe that drug-induced state transitions are reversible, particularly in ML cell lines. NE cell lines did exhibit a relatively permanent state transition, but it might be possible that these cell lines transition back into their original states given more time. This possibility remains to be tested in the future.

It is important to first identify phenotypic state defining markers that would hopefully also change upon drug treatment. To our surprise, the phenotypic markers we identified in Chapter IV were capable of broadly assessing drug-induced state transitions. But currently we have only assessed the role of these biomarkers in drugs

that induce broad changes in DNA damage or epigenetics. It remains to be seen if these markers would also work for specific targeted therapies.

No surgical resection is performed on SCLC patients due to early metastatic disease. However, several reports indicate that serum from SCLC patients contains circulating tumor cells (CTCs) that can be captured efficiently and studied for drug treatment studies (Hodgkinson et al., 2014). Our single-cell based flow cytometry approach could potentially be applied during SCLC treatment in the clinic to follow the phenotypic state dynamics of the tumor over time, which would be useful in tracking tumor evolution.

Acknowledgements

I would like to acknowledge Keisha Hardeman for assistance with Cellavista viability experiments, Dr. Jonathan Irish and David Wooten for ideas on flow cytometry data analysis.

CHAPTER VI

CONCLUSIONS AND FUTURE DIRECTIONS

Identification of a robust network signature for SCLC

SCLC have been considered a monolithic disease with respect to its diagnosis and treatment over the past several decades. Robust signatures for SCLC have thus far been lacking as well as targeted therapies that are effective in the clinic. Not being a mutated oncogenic driver addicted cancer like NSCLC, a novel approach and outlook towards SCLC biology is warranted. Overall, our work defines a combined bioinformatic-experimental-theoretical approach to delineating heterogeneity in lung cancer, specifically in SCLC.

We have identified a robust co-expression network based signature (SSHN) for SCLC tumors on three independent platforms (microarrays, RNAseq and shotgun proteomics). This signature was also conserved in SCLC cell lines. Future evaluation of this signature as a diagnostic tool requires a larger set of patients for independent validation. SSHN is composed of 287 genes, which is clinically not practical for use in diagnostic testing, so further reduction of this signature to a 25-50 gene set would prove to be useful in the future.

SSHN was composed of several well known players in cell cycle checkpoints (CDK4, CDC2, CDK2, CDC7, CDK5R1, BUB1) and DNA repair pathways (BRCA1). BRCA1 mutations significantly increases risk of breast and ovarian cancer (Antoniou et al., 2003) and reduced expression of BRCA1 is associated with EMT in breast cancer

(Wu et al., 2012). We found that BRCA1 was expressed at different levels in SCLC patients (Figure 6.1A). Also BRCA1 was co-expressed with other SSHN genes FYN and SYK (Figure 6.1B), and high expression of these 3 genes together suggests a better survival in this small cohort of SCLC patients (Figure 6.1C). Currently PARP1 inhibitors are being tried in patients with BRCA1 mutations providing some clinical benefit (Fong et al., 2009). PARP1 being a newly discovered target in SCLC (Byers et al., 2012), further analysis of BRCA1 mutations are warranted to identify patients sensitive to PARP1 inhibition, given our evidence of its varying levels of expression in SCLC.

SYK as a targeted therapy for a subset of SCLC

Within this SSHN network, we found twenty targetable kinases that were overexpressed in most, if not all of these platforms (Table 3.1). In Chapter 2, we validated 2 of the tyrosine kinases - SYK and FYN. SYK and FYN were overexpressed significantly in SCLC patients and cell lines by several independent bioinformatics and experimental methods, and distinguished between two potential groups of patients – SYK/FYN positive and negative. The SYK/FYN positive SCLC cell lines exhibited significant loss of viability and increased cell death in response to SYK siRNA, providing evidence for SYK as a novel oncogenic driver for SCLC (Udyavar et al., 2013). All SCLC patients get treated with combination chemotherapy (cisplatin, etoposide) without distinction. Our work suggests that stratifying patients with respect to SYK/FYN expression may open avenues to personalized medicine in SCLC, given that SYK small-molecule inhibitors are already in clinical trials for other disease conditions.

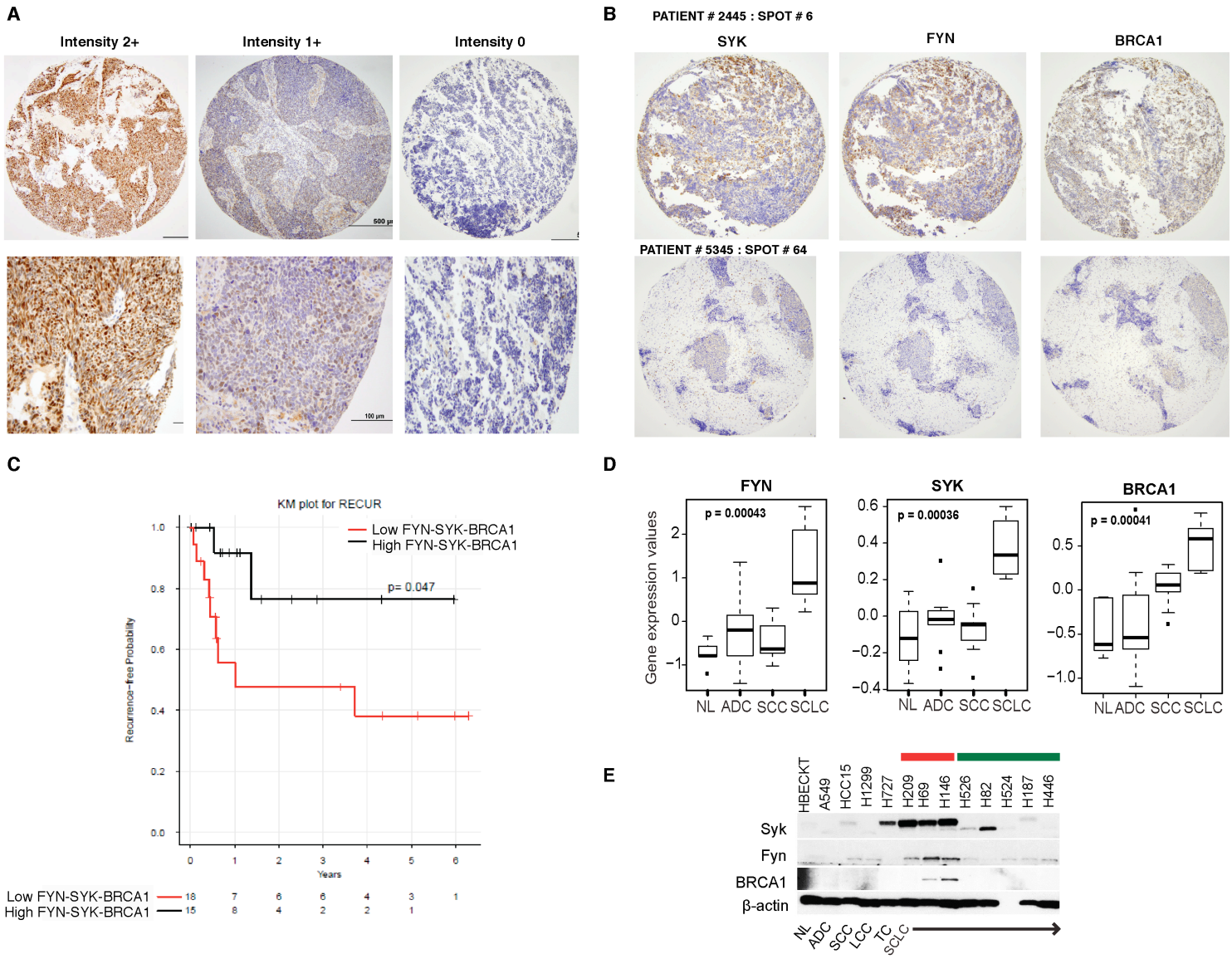


Figure 6.1. Variability of BRCA1 expression in SCLC patients and cell lines.

Figure 6.1. Variability of BRCA1 expression in SCLC patients and cell lines.

(A) Immunohistochemistry of BRCA1 protein in SCLC patient TMAs where patients exhibit high, medium and low expression. (B) Representative images of 2 patient subsets of patients with high and low FYN-SYK-BRCA1 co-expression identified. (C) Kaplan-Maier survival analysis of high and low FYN-SYK-BRCA1 patient subsets, indicating higher expression of these markers provides better survival with lower recurrence. (D) FYN, SYK and BRCA1 gene expression in SCLC patients from GSE6044 dataset Rohrbeck:2008em}. The outliers are denoted by dots. P-value shows statistical significance by Kruskal-Wallis nonparametric test (Kruskal and Wallis, 1952). FYN,SYK and BRCA1 are co-overexpressed in SCLC patients versus NSCLC (ADC, SCC) and normal lung. (E) Western blot of FYN, SYK and BRCA1 in SCLC cell lines where high and low expression is denoted by the red and green bars respectively. These high and low expressing cell lines are further defined as NE and ML phenotypic states aka transcriptional subytpes in Chapter IV.

SYK expression was maintained *in vivo* in xenografts tumors where H69 and H146 still exhibited high SYK expression, while NCI-H82 and NCI-H524 had low expression (Figure 6.2A). However, we were unable to detect any basal activation of SYK (i.e. phospho-SYK) in SCLC cell lines, unless treated via hydrogen peroxide (oxidative stress induction) and SYK inhibitors did not show any selective growth inhibition of SYK-expressing cell lines (data not shown). We also found variable expression of SYK in syngeneic p53/Rb knockout SCLC mouse models (Figure 6.2A). LX-22, a patient derived xenograft (PDX), also exhibited high SYK expression in both western blots and IHC, and displayed SYK activation *in vivo* (Figure 6.2B). Future work will have to determine whether in fact SYK may represent a potential actionable target in SCLC, by itself or in combination with chemo or radiation therapy.

Transcriptional regulation of a heterogeneous phenotypic state space of SCLC

Lung cancer heterogeneity today, is largely defined by genetic mutations, especially in NSCLC (Pao and Hutchinson, 2012). Attempts are currently being made to identify genetic heterogeneity in SCLC, however given the high mutation rate, not many novel oncogenic drivers have been discovered (Peifer et al., 2012; Rudin et al., 2012). With our current studies, we have defined a phenotypic state space that describes heterogeneity in human SCLC via two anti-correlated gene co-expression networks. This state space can be described as a landscape of stable phenotypic attractors - neuroendocrine (NE) and mesenchymal-like (ML) - driven by alternating dynamics of a transcriptional regulatory network that drive these anti-correlated co-expression modules using a boolean network model (Figure 6.2).

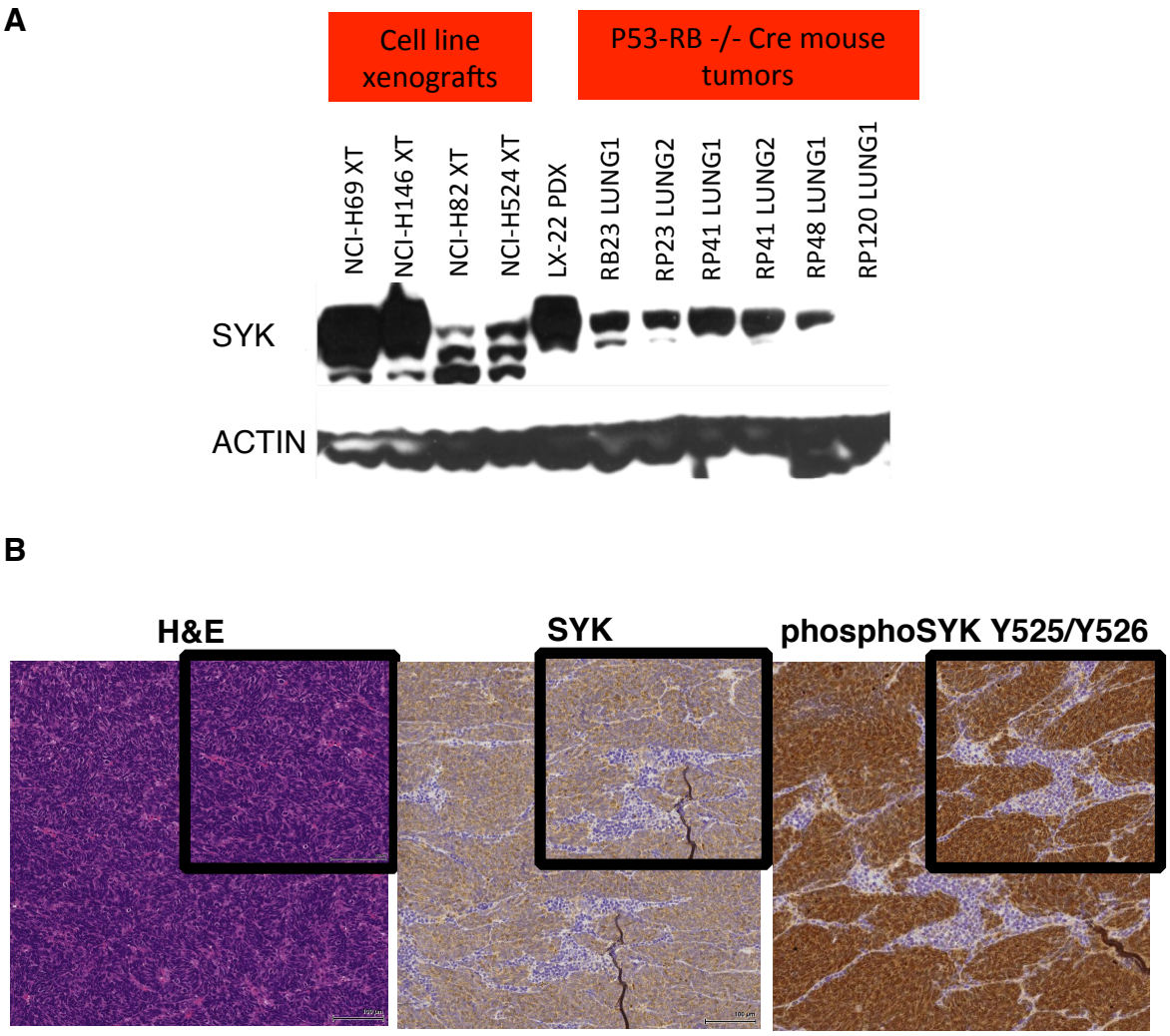


Figure 6.2. SYK expression and activation status in mouse models of SCLC.

(A) Western blot showing SYK expression of xenograft tumors of SCLC cell lines, genetically engineered p53/Rb knockout SCLC mouse models and PDX LX-22. (B) Immunohistochemistry staining for hematoxylin-eosin (H&E), SYK and phosphoSYK expression in PDX LX-22.

In addition, we validated the model predictions with several orthogonal approaches such as (1) consensus clustering, which also identified phenotypic states as transcriptional subtypes in both SCLC cell lines and patients; (2) transcription factor network ON/OFF “status” in the two subtypes; and (3) surface markers and signaling differences in SCLC (Figure 6.3). Furthermore, we verified the stability of the NE and ML attractor states via co-expression of NE and ML markers respectively at a single cell level using multidimensional flow cytometry analysis. Inter-tumor heterogeneity in SCLC can be described by the intra-tumor heterogeneity by percentage of NE and ML subpopulations.

To our knowledge, this is the first time that phenotypic heterogeneity has been defined in human SCLC patients with the identification of transcriptional subtypes (Figure 4.9). Previously the theoretical framework of attractor states has been proposed to be applied to explain cancer plasticity and heterogeneity (Creixell et al., 2012; Huang, 2011), but our work is the one of the first few attempts at applying this theoretical framework to cancer using boolean network modeling and validating it with a series of experiments in cell lines and patient samples.

Inter-tumor heterogeneity in SCLC can potentially arise from (1) distinct cells of origin i.e. there are distinct hits of p53/Rb loss in 2 phenotypically distinct cells; (2) cancer stem cell hypothesis - the 2 phenotypes occur from one common cancer stem cell population that gives rise to heterogeneity; (3) a balanced mix of cell types (or phenotypic states) exist in SCLC that generate intra-tumor heterogeneity, the majority of the ‘dominant’ population drives inter-tumor heterogeneity and maintains an equilibrium of phenotypic states. Our data suggest that the third possibility might be true, although

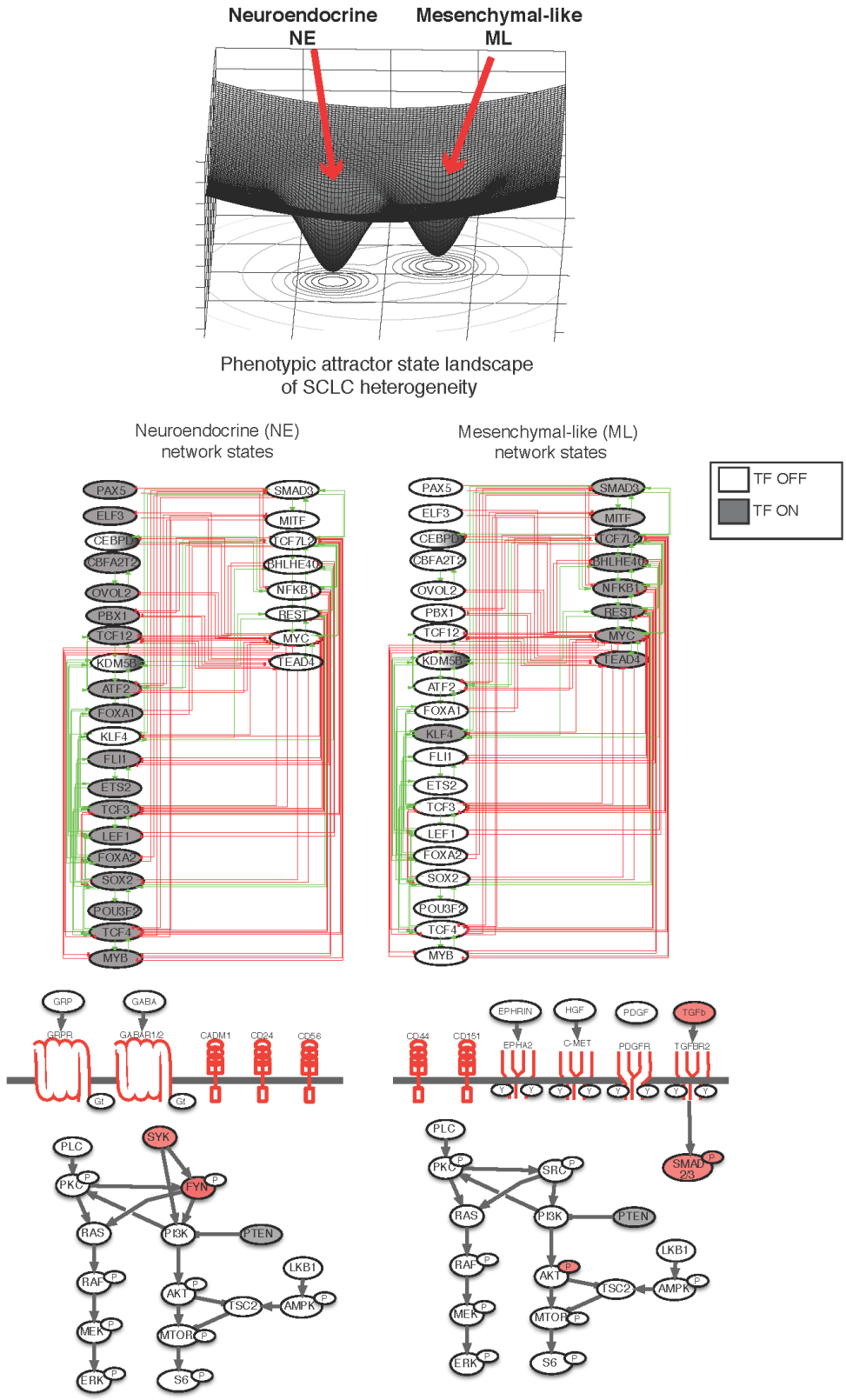


Figure 6.3. A heterogeneous phenotypic state space for SCLC defined as two stable attractors - neuroendocrine (NE) and mesenchymal-like (ML).

Figure 6.3. A heterogeneous phenotypic state space for SCLC defined as two stable attractors - neuroendocrine (NE) and mesenchymal-like (ML).

This figure summarizes the findings from Chapter IV where we identified two distinct phenotypic states NE and ML driven by specific ON/OFF configurations of a transcriptional regulatory network. These states also show differences in phenotypic surface markers, basal signaling and kinases.

the distinct cells of origin possibility cant be entirely ruled out. GEMM of SCLC also suggest that the individual phenotypes (NE and non-NE) are stable and together co-operate in some yet to be discovered manner to drive metastasis (Calbo et al., 2011). This also goes along with the idea of greater initial tumor diversity might correlate with worse clinical outcome with increased propensity of tumor metastasis.

Proposed attractor landscape model for Lung cancer

Based on our mixed experimental-bioinformatic-theoretical foundation, we propose an attractor state landscape model for SCLC and lung cancer in general whereby in normal lung development, cells exist in stable attractor states such as AT2, Clara, Basal cells, PNEC, etc each dictated by a distinct TF network state. During lung injury, several cell types such as PNEC, AT2, Clara cells exhibit the capacity to "transdifferentiate" into neighboring cell types in order to repopulate the lung (Desai et al., 2014; Li and Linnoila, 2012; Song et al., 2012) (Figure 6.4, top panel - 'normal phenotypic landscape').

In cancer, after loss of p53 (and RB in case of SCLC) and amplification/mutation of oncogenic drivers, this barrier is crossed over to cancer due to changes in genomic stability and epigenetics. Similar types of plasticity as seen during lung injury are observed in cancer cells where a particular cell type can give rise to SCLC or NSCLC (or different types of NSCLC) depending on the mutation hit (Figure 6.4- 'epigenetic barrier'). For example, *Sutherland et.al* generated several mouse models with Cre-recombination mediated knockout of p53 and Rb in PNEC, Clara and AT2 cells. Even though the majority of tumors arising from PNECs were neuroendocrine tumors

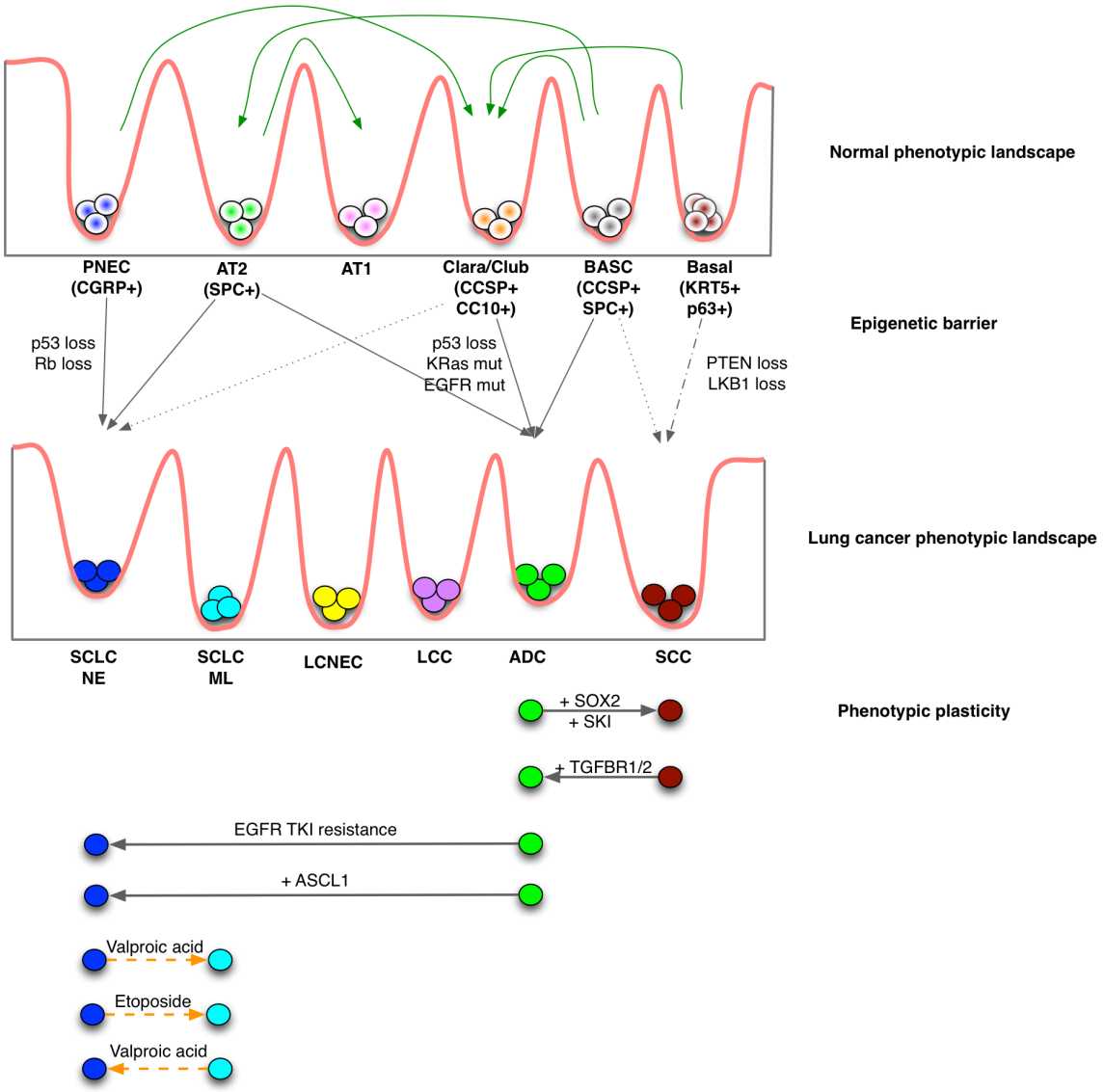


Figure 6.4. Attractor landscape model for lung cancer heterogeneity.

Normal cell types and lung cancer phenotypic landscapes are shown as attractor basins where cells at the bottom of the basin denote a stable attractor. State transitions are denoted by the arrows in normal phenotypic landscape and phenotypic plasticity. Genetically engineered mouse model studies predict the cell of origin for various lung cancer subtypes, where multiple cell types can give rise to a particular cancer subtype. These transitions are possible in a dynamical system via global changes in transcriptional regulatory networks without acquisition of new mutations. Literature-based evidence for ADC<-> SCC, and NSCLC --> SCLC state transitions are denoted by solid lines with mechanisms. Our drug treatment analysis suggests NE-->ML state transitions can be made via etoposide or valproic acid, while ML--> NE state transitions can be possible via Valproic acid (denoted by dotted orange lines).

resembling human SCLC, other cell types were also capable of generating neuroendocrine tumors admittedly at a lower efficiency (Sutherland et al., 2011). Such examples were also seen in NSCLC where LKB1;PTEN knockout specifically gave rise to SCC, while LKB1;PTEN;p53 combined knockout generated adeno-squamous tumors, and KRAS mutation;LKB1;p53 loss generated both murine lung ADC and SCC tumors (Xu et al., 2014). Over the past few years, several cell types have been shown to yield ADC tumors such as AT2, Clara cells and Bronchioalveolar stem cells (BASCs) (Kim et al., 2005; Sutherland et al., 2014). Such time-course progression studies are impossible to validate in patients, since early stage disease detection is not common in lung cancer. Nevertheless, clinically patients do exhibit mixed SCLC phenotype (with NSCLC or LCNEC) (Nicholson et al., 2002), ADC-SCC (Filosso et al., 2011), as well as NSCLC with neuroendocrine features (Walker et al., 2005) suggesting that this attractor state view of lung cancer might be useful to consider while treating these patients. Based on our preliminary analysis with SCLC NE and ML surface markers, we do find similarities between SCLC and NSCLC with a few NSCLC cell lines (specifically K-Ras mutant) expressing NE marker CD56 (Figure 6.5). Further comparative analysis with ADC and SCC specific markers (Table 1.1), kinases (EGFR, KRAS) and TFs might provide a global view of the phenotypic heterogeneity of lung cancer. Building a TF-network dynamic model composed of ADC, SCC and SCLC-specific TFs would provide potential insights into the regulatory control of these distinct phenotypic states.

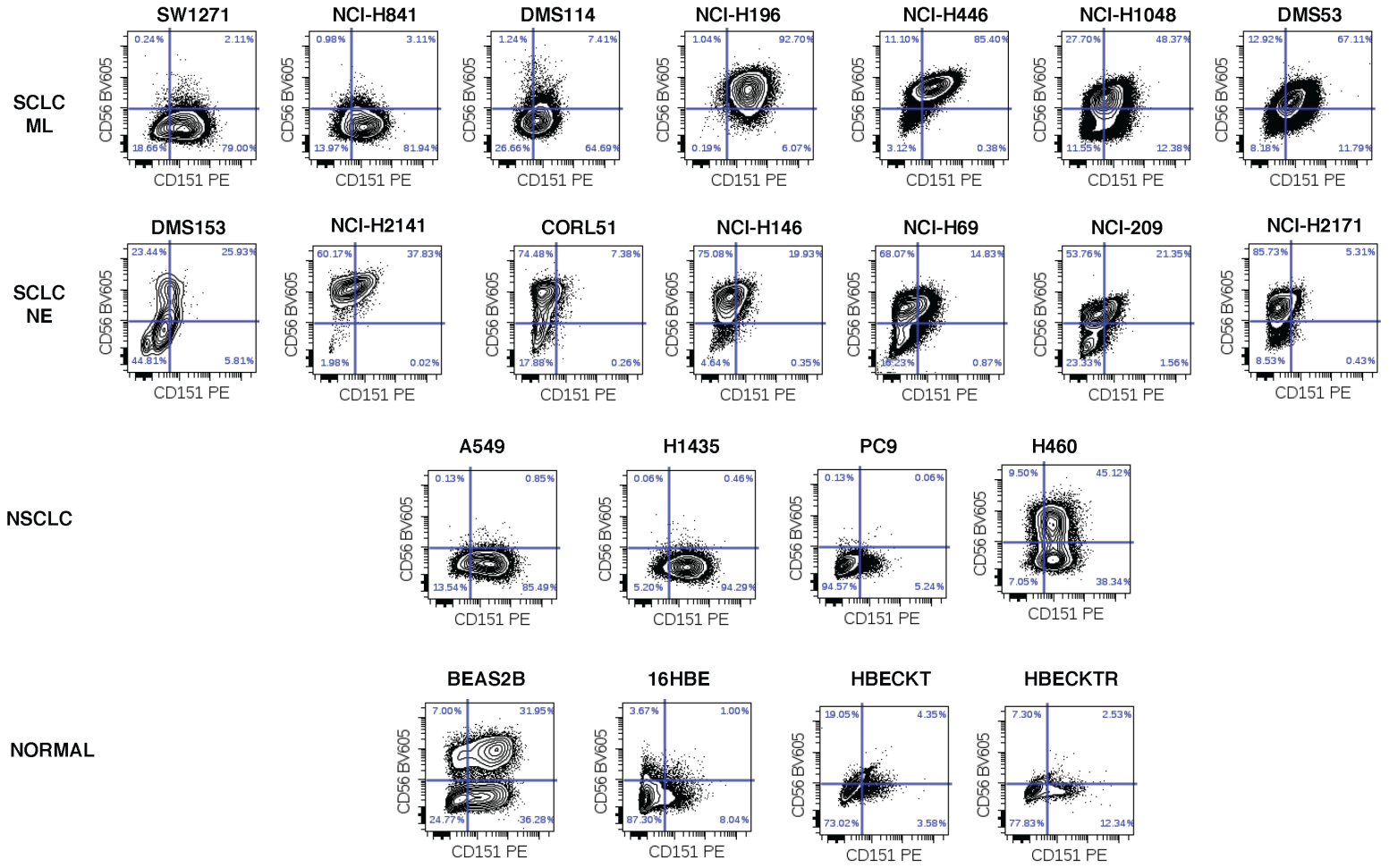


Figure 6.5. Preliminary analysis of NE and ML phenotypic state markers on NSCLC and normal cell lines.

Figure 6.5. Preliminary analysis of NE and ML phenotypic state markers on NSCLC and normal cell lines.

Flow cytometry analysis of NE marker CD56 and ML marker CD151 on NE and ML SCLC, ADC (EGFR mutant PC9, K-Ras mutant A549, NCI-H1435), LCC (KRas mutant NCI-H460) and normal cell lines. NE cell lines show high percentages of CD56+CD151- cells and some CD56+CD151+ populations. On the other hand, ML cell lines are more heterogeneous where some cell lines show almost exclusive expression of CD56-CD151+ subpopulations while others show a high percentage of CD56+CD151+ subopulations. These double positive cell lines also exhibit somewhat mixed adherent and suspension cells in culture. Interestingly, K-Ras mutant A549 and H1435, but not EGFR mutant PC9, ADC cell lines express CD151, while NCI-H460 a K-Ras mutant LCC, exhibits both CD56+CD151+ and CD56-CD151+ subpopulations. Most transformed normal cell lines except BEAS2B (SV40 transformed) exhibit low expression of both markers. This suggests that NE and ML subpopulations are present in NSCLC cell lines, especially ones with KRas mutations. Further analysis with NSCLC specific markers are warranted to test the extent of NSCLC differentiation in SCLC cell lines.

Reversible reprogramming due to changes in epigenetics (i.e. transcriptional networks) has been observed in cancers such as melanoma (Quintana et al., 2010), breast (Gupta et al., 2011) and gastrointestinal cancers (Schwitalla et al., 2013). Alveolar epithelial subpopulations are capable of generating both lung ADC and SCC, that can switch between ADC and SCC phenotypes depending on the TGFbeta pathway activity (Figure 6.4- 'phenotypic plasticity')(Ischenko et al., 2014). *Calbo et.al* showed that NE subpopulations within murine SCLC models can be switched to non-NE phenotypes via induced expression of mutant K-Ras (Calbo et al., 2011). However, K-Ras mutations are not found in SCLC patients, leading to the speculation if the NE-->non-NE transition was in fact a SCLC-->NSCLC transition, which remains to be seen. Treatment-induced plasticity leading to drug resistance is also observed in cancer. 14% of EGFR-mutant NSCLC patients that develop acquired resistance to EGFR TKI do so by switching to a SCLC phenotype which are then sensitive to SCLC standard of care combination chemotherapy (Sequist et al., 2011). EGFR-mutant lung cancer also show reversible sensitivity to EGFR TKI maintained by altered epigenetic state (Sharma et al., 2010).

Transcriptional networks largely govern the phenotypic state space and cellular heterogeneity in normal lung during development and injury, and potentially also cancer (Warburton et al., 1998). ASCL1 expressing progenitor cells differentiate into PNEC and other airway cell types such as Clara and AT2 during development (Li and Linnoila, 2012). Expression of ASCL1 also governs NE differentiation in NSCLC tumors and is prognostic indicator of poorer survival (Osada, 2005). Notch1 via Hes1 downregulates

ASCL1 expression in development and prevents PNEC differentiation in normal lung and SCLC (Ito et al., 2000; Shan et al., 2007). Interestingly we also found that ASCL1 and Notch1, both at gene and protein level, are anti-correlated in SCLC cell lines (protein data cor: -0.585, p-value: 0.006702), where ASCL1 is highly expressed in NE cell lines, but not ML, and vice versa for Notch1 expression (Figure 6.6). This reinforces the fact that distinct differentiation programs regulated via differential network states operate in these two transcriptional subtypes of SCLC. Similarly, SOX2 expression in ADC promotes conversion to SCC (Ischenko et al., 2014).

In summary, understanding the dynamics of the TF networks that result in distinct states can be useful for treatment, in terms of identifying vulnerable TF 'nodes' or its targets. The idea of attractor states naturally permits plasticity since the dynamics of the TF network drive the stability of a particular attractor state, which can be noisy given the varying tumor microenvironment. Under drug treatment, this TF network state might change dramatically, leading to a phenotypic 'switch' from one attractor state to another. This is in contrast to a hierarchical view of heterogeneity where a few cancer stem cells are typically the drug resistant cells, that repopulate the majority tumorigenic population. Identification of attractor states most vulnerable to a drug will be critical for therapeutic targeting as well as strategies to push the drug-resistant attractor states into drug-sensitive ones, suggesting the strong need for rational and effective therapeutic combinations.

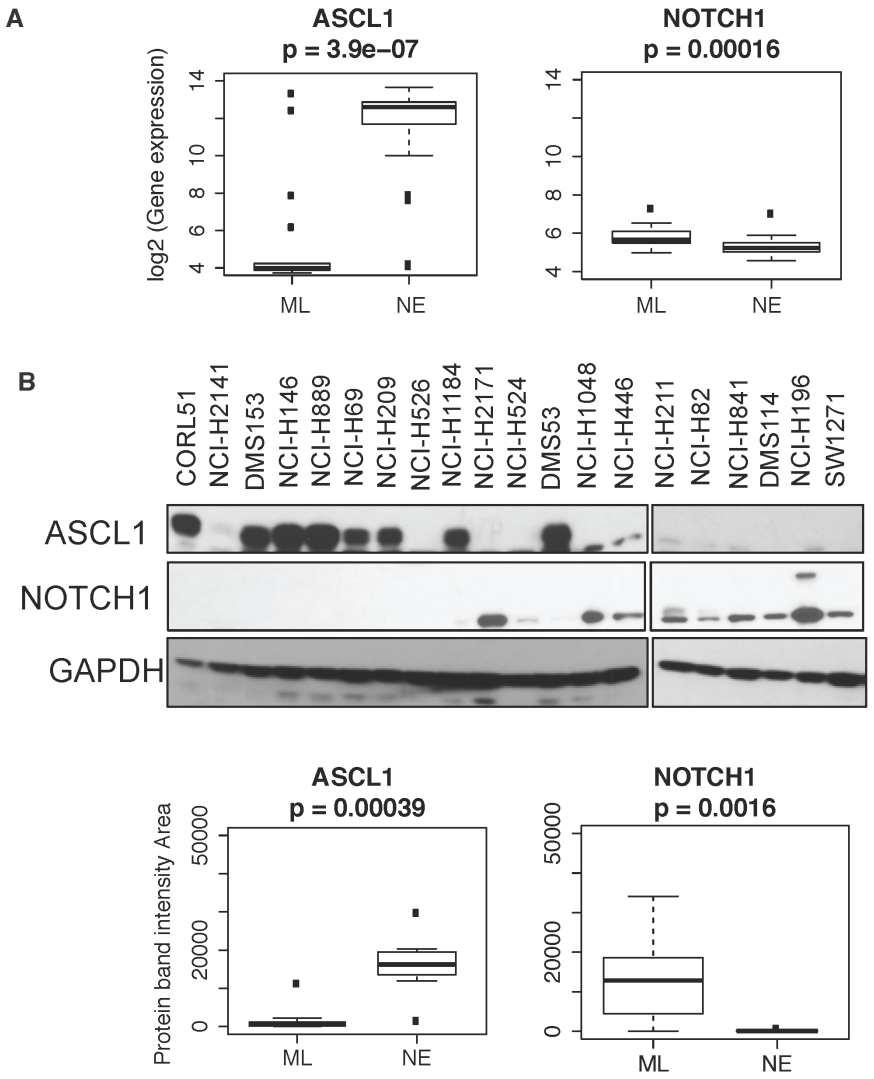


Figure 6.6. Anti-correlated pattern of expression of ASCL1 and NOTCH1 in NE and ML transcriptional subtypes.

(A) Boxplots comparing ASCL1 and NOTCH1 gene expression in 53 SCLC cell lines in the CCLE dataset divided into the NE and ML phenotypic states. (B) Western blot of ASCL1 and NOTCH1 across 20 SCLC cell lines ordered from high ME-Blue (CORL51) to high ME-Turquoise (SW1271). Protein band intensity quantification boxplots are shown below in the 2 phenotypic states. P-values are given by Kruskal-Wallis test.

Identification of drug-sensitive phenotypic attractor states

Resistance to chemotherapy is a major reason for SCLC treatment failure. Study of phenotypic states that arise during course of treatment in SCLC might be helpful to determine mechanisms of drug resistance. The stable phenotypes might remain in equilibrium until a change occurs in tumor microenvironment or drug treatment. This might explain initial high response rates of SCLC patients to chemotherapy but rapid development of recurrent disease. Thus following the tumor evolution over the course of treatment might be key in SCLC patients.

As we now know SCLC is phenotypically heterogeneous, the goal would be to first identify which subpopulations are most sensitive to a particular therapy, use drugs that would reduce the heterogeneity and transition the phenotypic state to a more vulnerable one, and then co-target a specific target that could potentially abolish the drug-sensitive populations. Pathway-based approaches such as described in Chapters III and IV are warranted for identification of novel targets. We showed SYK is an important target in the NE subset of SCLC (Udyavar et al., 2013). EPHA2, a RTK, important in KRAS mutant NSCLC (Amato et al., 2014), could also be a potential therapeutic target of ML subtype of SCLC. We also observed high TGFbeta, phospho-SMAD2/3 expression in ML subtype, which could be potentially targeted via TGFBR2 inhibitors. We have also identified a transcriptional regulatory network driving the distinct NE and ML phenotypic states in SCLC. This network could be simulated *in silico* for assessment of critical TF nodes that could be important for NE and ML differentiation and survival (singly or in combination) and potentially promote a state transition to a more drug-sensitive state. TFs are currently not targetable in the clinic, but targets of

TFs or synthetic lethal combinations (given a particular TF overexpression such as Aurora kinase inhibitors for MYC amplified SCLC (Sos et al., 2012)) could be of therapeutic relevance in SCLC. We also identified valproic acid (VPA), a HDAC inhibitor, that was effective in both NE and ML cell lines and induced massive cell death. It also induced state transitions (NE-->ML and vice versa), which were relatively permanent in NE and reversible in ML cell lines. VPA supports the idea that the drugs can be pushed into a drug-sensitive or resistant state upon treatment. Since ML cell lines are in general more resistant to treatment, it might be a useful tool to push ML states into NE which might be more amenable to treatment. However, its effect on tumor growth and state transitions in *in vivo* setting remains to be seen.

Monitoring dynamics of therapeutic response *ex vivo*

Tumor heterogeneity is a major concern in cancer biology and most likely also an issue in SCLC. Tumor evolution is currently being studied by either single-cell sequencing of tumors in the hope to find clonal lineage trees, or studying tumors at various stages of progression (de Bruin et al., 2014; Polyak, 2011; Zhang et al., 2014b). It is difficult to study tumor evolution or identify cell of origin in SCLC by single-cell sequencing attempts since most patients usually present with metastatic widespread disease. This has recently been attempted in mouse models of SCLC (McFadden et al., 2014).

Studying tumor dynamics over time would be ideal using biopsies, but that may be impossible since surgical resections are rare in SCLC. SCLC patients do exhibit a lot of circulating tumor cells (CTCs) in their blood circulation which are prognostic

biomarkers of survival and response to chemotherapy (Hiltermann et al., 2012). CTCs have also been used to generate patient derived xenografts and study course of treatment *ex vivo* in mice (Hodgkinson et al., 2014). One potential idea would be to study if this phenotypic heterogeneity (Chapter IV) exists *in vivo* and how it is impacted during treatment. Ultimate goal would be to chase the tumor over time and follow its course using fine-needle biopsy and circulating tumor cells (CTC) *in vivo*. Change in diagnosis and therapeutic monitoring needs to take place in SCLC, for example, fine needle biopsy over course of treatment and flow cytometry-based tracking analysis of cellular phenotypic states.

Signaling and metabolic heterogeneity in SCLC

Blue and Turquoise modules were composed of 1179 and 3450 genes. Comparative Pathway enrichment analysis using BINGO and Enrichment map in Cytoscape and Ingenuity Pathway analysis® (www.ingenuity.com) showed common pathways and some key differences in the biological function of these modules such as differentiation, signaling, adhesion, metabolism and stress response (Figure 6.7A). The differentiation and signaling pathway differences are described in Chapter IV (Figure 4.2).

Interestingly, both modules showed an enrichment of metabolic pathways (Figure 6.7B). Blue module was enriched for cholesterol/sterol biosynthesis while the turquoise module was enriched for fatty acid catabolic processes namely, beta-oxidation. In addition, the turquoise module was particularly enriched for glucose and glutamine

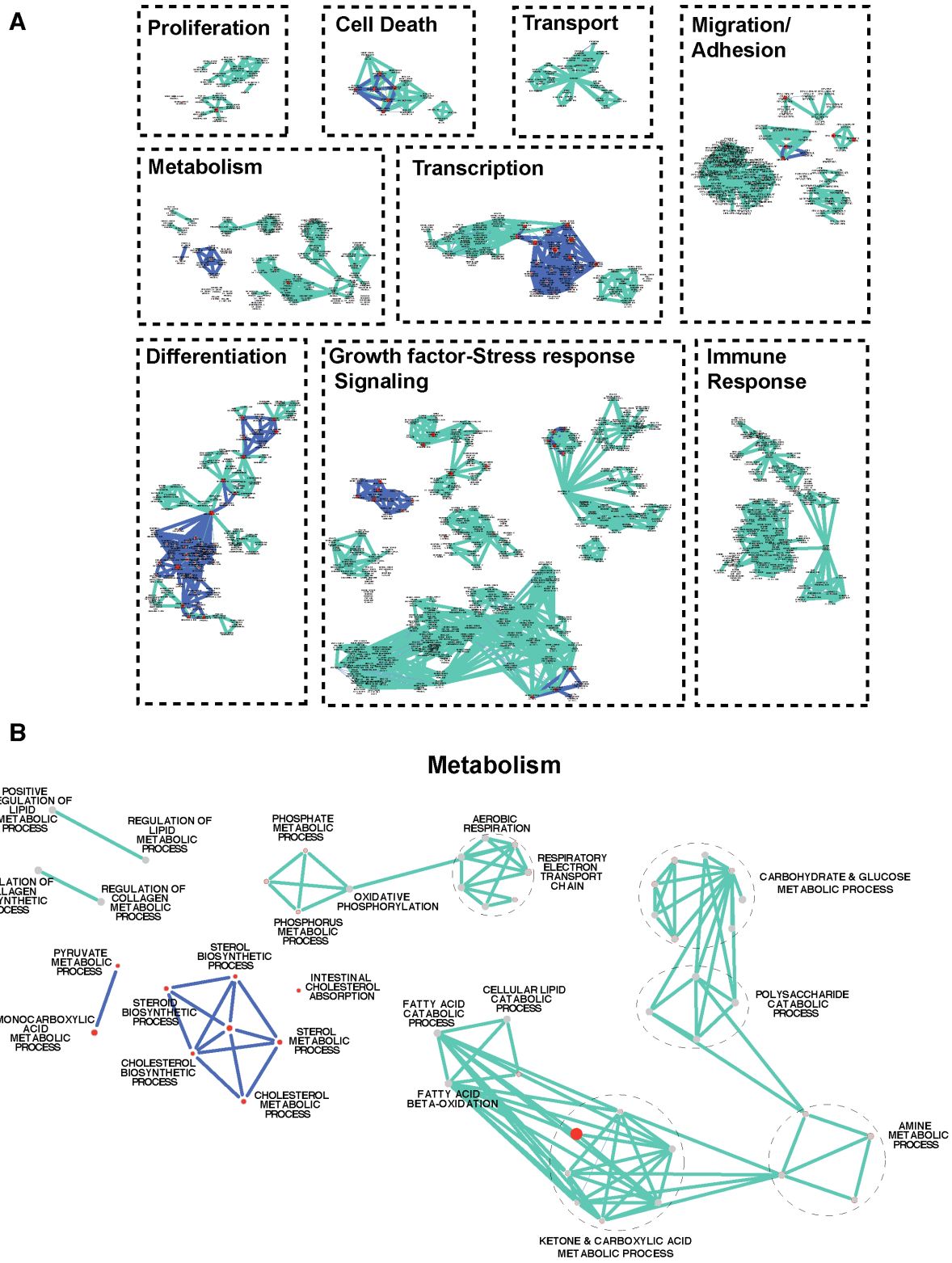


Figure 6.7. Pathway enrichment analysis of Blue and Turquoise modules with focus on metabolism in SCLC.

Figure 6.7. Pathway enrichment analysis of Blue and Turquoise modules with focus on metabolism in SCLC.

This figure describes comparative enrichment analysis of Gene Ontology (GO) pathways enriched in Blue versus Turquoise networks using BINGO and EnrichmentMap in Cytoscape®. (A) Blue and turquoise modules show statistically significant differences in metabolism, signaling, stress response, adhesion, differentiation, transcription, proliferation and apoptosis. Nodes denote the enriched GO categories. Edges denote the connections between the pathways, Blue and Turquoise colors of the edges indicate pathway groups enriched specifically in Blue or Turquoise modules respectively. (B) Zoom-in figure of the metabolic pathway differences. Blue module is enriched in cholesterol and steroid biosynthetic process while Turquoise module is enriched in amine/glucose/fatty acid catabolic processes.

catabolic pathways including oxidative phosphorylation, electron transport chain, aerobic glycolysis. This suggests a fundamental difference in the metabolic programming of cell lines enriched in these respective networks where one is more proficient in utilizing resources for building blocks while the other is proficient in catabolic processes using glucose, glutamine and fatty acids for generation of ATP.

Metabolic reprogramming goes hand-in-hand with increased proliferation and tumor progression, and inhibition of cancer metabolism can serve as useful therapeutic targets (Hanahan and Weinberg, 2011; Zhao et al., 2013). Preliminary analysis of oxygen consumption with various metabolic substrates and inhibitors (see Chapter 2 for details) showed that ML cell lines (NCI-H82, NCI-H524) are more glutamine-avid and consume more oxygen to utilize it while NE cell lines (NCI-H146, NCI-H69) are more reliant on oxidative phosphorylation (Figure 6.8A and B). In addition, SLC1A5, a glutamine-avid transporter, shown to be an important regulator of glutamine uptake in NSCLC (Hassanein et al., 2013), is also overexpressed specifically in the ML cell lines (NCI-H82, NCI-H524), reinforcing the fact that the ML cell lines might be more glutamine-dependent for their growth than NE cell lines (Figure 6.8C). More detailed analysis with specific inhibitors of OXPHOS and glutamine metabolism are warranted, thus providing a novel metabolic reprogramming-based treatment strategy, currently an unexplored area of research in SCLC.

In terms of signaling differences, the blue module was enriched for neuronal signaling pathways such as neurotransmitter release, reelin, Her3, IGF-1, cell-to-cell signaling while the turquoise module was enriched for TGFβ, Jak-STAT, Ephrin, MAPK, cytokine (IL-6, IL-10), and NFKB signaling pathways, suggesting a more EMT-like

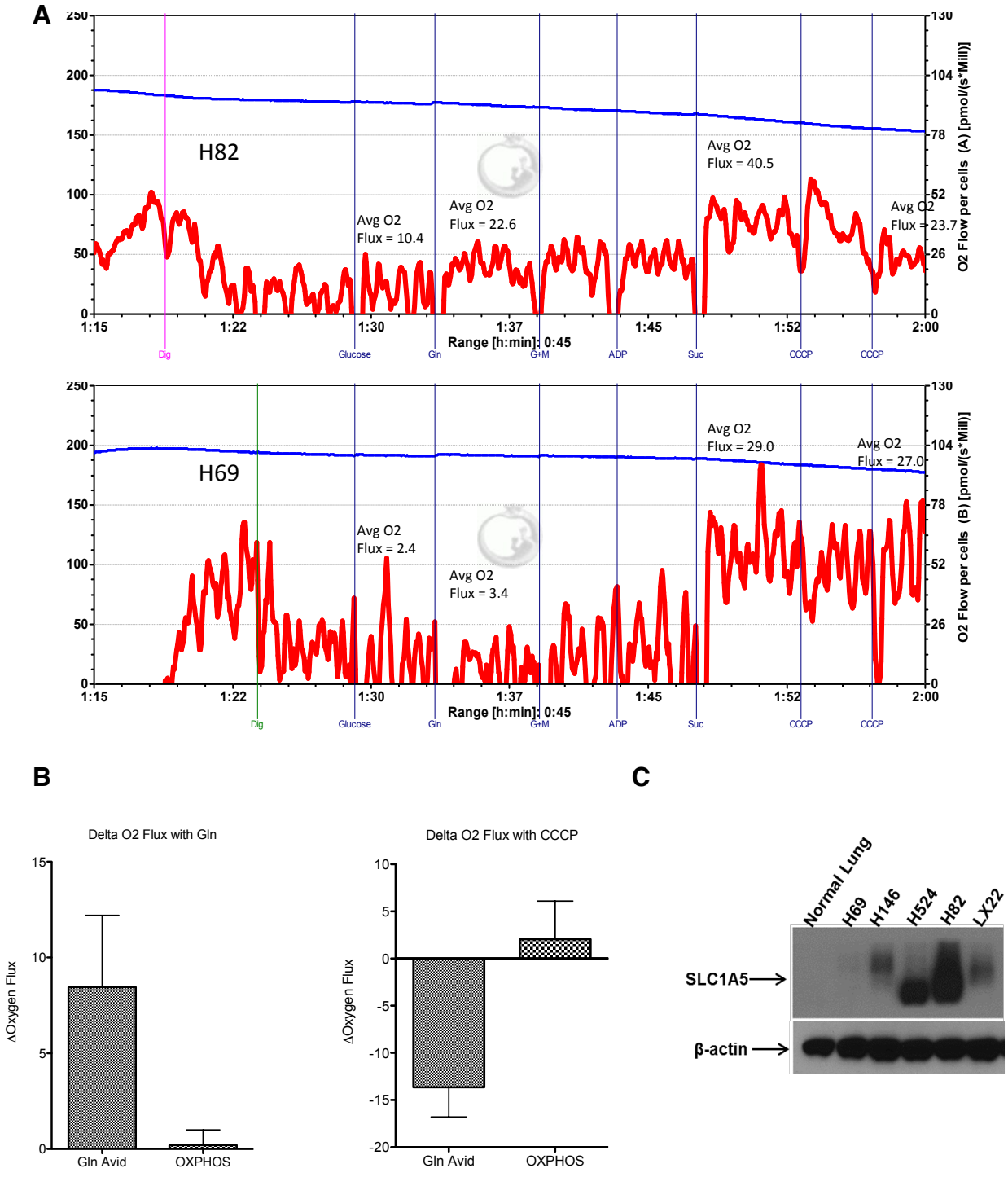


Figure 6.8: Metabolic variability in NE and ML SCLC cell lines.

Figure 6.8: Metabolic variability in NE and ML SCLC cell lines.

(A) is an Oxygraph trace of NE SCLC line NCI-H69 and ML cell line NCI-H82. X-axis is time. Left Y-axis is oxygen concentration. Right Y-axis is specific oxygen flux, which is the negative derivative of the O2 concentration with respect to time, normalized to cell number. Along the x-axis are denoted a number of events. Here they are in order:

Digitonin - This was addition of digitonin to permeabilize the plasma membrane to allow for efficient diffusion of substrates into and out of the cells. This removes the effects of transporters, diffusion barriers from plasma membrane, etc. It also allows endogenous substrates to leak out of the cell, so that you can see the true effect of addition of the substrates and inhibitors of interest.

Glucose - Cells were in glucose-free media, so first glucose was added as a substrate.

Glutamine - Glutamine was then added. The idea here was that we would see an increase in O2 flux with a glutamine-avid (turquoise high) cell line, but not with an OXPHOS reliant (blue high) line. Recall that cells are permeabilized, so there is no differential effect of glutamine transporters here.

G + M - Glutamate and malate, to fuel Complex I (assuming intact TCA function).

ADP - This is to induce state 3 respiration.

Suc - Succinate addition. This usually gives a fairly impressive increase in O2 flux, but it's fairly blunted in these cells. One of the other lines gave a more robust succinate response. With addition of this substrate, the electron transport system is more or less fully engaged (no fatty acid substrate, but wouldn't expect these cells to metabolize a lot of fat).

CCCP - An uncoupler, works by dissipating the proton gradient. In these cells, we either saw a puny response or outright inhibition.

(B) There's a measurable and reproducible increase in O2 flux with addition of glutamine for glutamine-avid lines (n = 2 lines tested thus far), but not for OXPHOS reliant lines (n = 2 lines tested). With addition of CCCP, glutamine-avid lines show a fairly pronounced decrease in O2 flux, but the OXPHOS reliant lines don't.

(C) Western blot of SLC1A5, a glutamine avid transporter, expression in SCLC cell lines and PDX. Normal lung lysate is used as negative control. SLC1A5 is overexpressed in ML cell lines NCI-H82 and NCI-H524, with low/no expression in NE cell lines NCI-H146 and NCI-H69.

These data together suggest that ML cell lines are Glutamine-avid while NE cell lines are more OXPHOS-dependent.

phenotype. Interestingly, reelin signaling is a negative regulator of TGFβ1 induced EMT and migration in other cancers (Yuan et al., 2012). EMT transition is accompanied by increased glucose catabolic pathways such as glycolysis that promote faster proliferation, and reduce Reactive oxygen species (ROS) in cells via increased NADPH and GSH (ROS scavengers) (Dong et al., 2013). Redox signaling or response to oxidative stress (particulary hydrogen peroxide) was enriched in both the modules, a key PNEC function (Buttigieg et al., 2012; Cutz and Jackson, 1999; Domnik and Cutz, 2011) as well as in important mechanism in EMT (Tobar et al., 2010), suggesting that redox sensing and signaling might be a key mechanism in SCLC tumors.

To obtain a global idea of the signaling response in SCLC NE and ML subtypes, we performed a preliminary analysis of fluorescent barcoding signaling experiments (see Chapter 2- Materials and Methods for details) by stimulating NE and ML cell lines with 22 different mitogens. ML cell lines in general were more responsive to RTK stimuli such as EGF, IGF-1 and HGF (Figure 6.9B-D), while NE cell line was more responsive to Interleukins (Figure 6.9A). All cell lines were responsive to H2O2. ML cell lines exhibited a high phospho-S6 signal upon stimulation via various RTK stimuli, while NE cell line had reduced phospho-S6. NE cell line NCI-H146 showed high phospho-PLCgamma (Figure 6.9A), which is also downstream of SYK signaling cascade. Particularly, in ML cell lines NCI-H82, NCI-H1048 and SW1271, high phospho-AKT signal was detected upon EGF, IGF, SCF, Neuregulin (NRG) stimulation, consistent with the fact that RTK signaling is more important in ML cell lines than NE (Chapter IV). We also observed high basal phospho-AKT in ML cell lines, but not in NE (Figure 4.11). G-protein coupled receptor signaling mediated via neuropeptides is important in NE

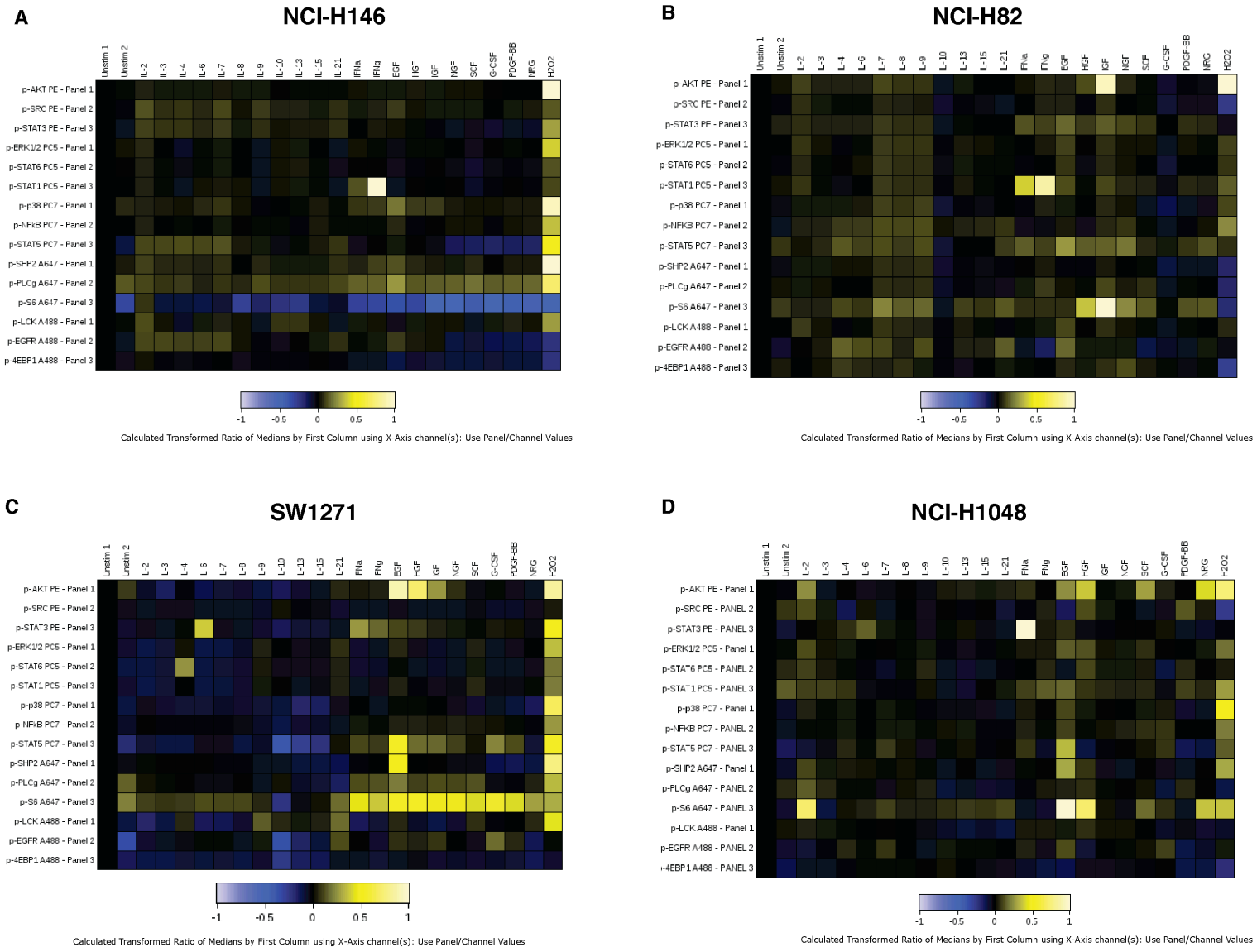


Figure 6.9 : Global changes in signaling response dynamics to various mitogenic stimuli in NE and ML SCLC cell lines.

Figure 6.9 : Global changes in signaling response dynamics to various mitogenic stimuli in NE and ML SCLC cell lines.

(A) NE SCLC cell line (NCI-H146) and (B-D) 3 ML cell lines (NCI-H82, NCI-H1048 and SW1271) were stimulated with 22 mitogens for 15minutes followed by measurement of 15 signaling readouts using fluorescent barcoding technique (see Chapter II - fluorescent barcoding section for details). This was a broad screen intended to capture signaling differences in NE and ML cell lines to cytokines and RTK mitogens. H2O2, a oxidative stress inducer as well as inhibitor of phosphatases, is used as positive control. The signaling readouts are summarized as signaling vs readout matrix heatmap for each cell line, where increased signal induction than basal (unstimulated) is given by yellow and decrease in signal upon stimulation is denoted by blue.

subtype. These stimuli were not present in this current mitogen panel. In future, other NE specific stimuli such as reelin, GRP/bombesin, somatostatin, glutamate/GABA, and ML-specific stimuli such as TGFbeta and Ephrins could be added to this panel to better distinguish between NE and ML signaling and identify signaling pathway-based therapeutic targets for the SCLC subsets.

Acknowledgements

I would like to acknowledge Dr. Joshua Fessell and Dr. Mohamed Hassanein for their advice and assistance on the metabolism experiments. Dr. Mohamed Hassanein has led the project on the discovery of SLC1A5 in NSCLC and has graciously shared the SCLC data on SLC1A5 with me. I would also like to acknowledge Nalin Leelathian from Dr. Jonathan Irish's lab for all the help, teaching and advice on the barcoding signaling experiments. Many thanks to Dr. Charles Rudin (MSKCC, NY) and Dr. Julien Sage (Stanford University) for providing the LX-22 and NJ-H29 PDXs respectively, and Dr. Kwok-Kin Wong (Dana Farber Cancer Institute) for providing samples (protein lysates and tissues) of p53/Rb knockout SCLC mouse models.

REFERENCES

Albeck, J.G., Mills, G.B., and Brugge, J.S. (2013). Frequency-modulated pulses of ERK activity transmit quantitative proliferation signals. Molecular Cell *49*, 249–261.

Almendro, V., Cheng, Y.-K., Randles, A., Itzkovitz, S., Marusyk, A., Ametller, E., Gonzalez-Farre, X., Muñoz, M., Russnes, H.G., Helland, A., et al. (2014a). Inference of Tumor Evolution during Chemotherapy by Computational Modeling and In Situ Analysis of Genetic and Phenotypic Cellular Diversity. Cell Reports *6,* 514-527.

Almendro, V., Kim, H.J., Cheng, Y.-K., Gonen, M., Itzkovitz, S., Argani, P., van Oudenaarden, A., Sukumar, S., Michor, F., and Polyak, K. (2014b). Genetic and phenotypic diversity in breast tumor metastases. Cancer Research *74*, 1338–1348.

Almendro, V., Marusyk, A., and Polyak, K. (2013). Cellular Heterogeneity and Molecular Evolution in Cancer. Annu. Rev. Pathol. Mech. Dis. *8*, 277–302.

Amato, K.R., Wang, S., Hastings, A.K., Youngblood, V.M., Santapuram, P.R., Chen, H., Cates, J.M., Colvin, D.C., Ye, F., Brantley-Sieders, D.M., et al. (2014). Genetic and pharmacologic inhibition of EPHA2 promotes apoptosis in NSCLC. J. Clin. Invest. *124*, 2037–2049.

Amir, E.-A.D., Davis, K.L., Tadmor, M.D., Simonds, E.F., Levine, J.H., Bendall, S.C., Shenfeld, D.K., Krishnaswamy, S., Nolan, G.P., and Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat Biotechnol *31*, 545–552.

Antoniou, A., Pharoah, P.D.P., Narod, S., Risch, H.A., Eyfjord, J.E., Hopper, J.L., Loman, N., Olsson, H., Johannsson, O., Borg, A., et al. (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am. J. Hum. Genet. *72*, 1117–1130.

Arriola, E., Cañadas, I., Arumí, M., Rojo, F., Rovira, A., and Albanell, J. (2008). Genetic changes in small cell lung carcinoma. Clin Transl Oncol *10*, 189–197.

Asamura, H. (2006). Neuroendocrine Neoplasms of the Lung: A Prognostic Spectrum. Journal of Clinical Oncology *24*, 70–76.

Augustyn, A., Borromeo, M., Wang, T., Fujimoto, J., Shao, C., Dospoy, P.D., Lee, V., Tan, C., Sullivan, J.P., Larsen, J.E., et al. (2014). ASCL1 is a lineage oncogene providing therapeutic targets for high-grade neuroendocrine lung cancers. Proceedings of the National Academy of Sciences *111*, 14788–14793.

Ban, K., Gao, Y., Amin, H.M., Howard, A., Miller, C., Lin, Q., Leng, X., Munsell, M., Bar-Eli, M., Arlinghaus, R.B., et al. (2008). BCR-ABL1 mediates up-regulation of Fyn in chronic myelogenous leukemia. Blood *111*, 2904–2908.

Barbareschi, M., Cantaloni, C., Del Vescovo, V., Cavazza, A., Monica, V., Carella, R., Rossi, G., Morelli, L., Cucino, A., Silvestri, M., et al. (2011). Heterogeneity of Large Cell Carcinoma of the Lung: An Immunophenotypic and miRNA-Based Analysis. American Journal of Clinical Pathology *136*, 773–782.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2013). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–307.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., et al. (2010). NCBI GEO: archive for functional genomics data sets--10 years on. Nucleic Acids Research *39*, D1005–D1010.

Bass, A.J., Watanabe, H., Mermel, C.H., Yu, S., Perner, S., Verhaak, R.G., Kim, S.Y., Wardwell, L., Tamayo, P., Gat-Viks, I., et al. (2009). SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet *41*, 1238–1242.

Basu, D., Nguyen, T.-T.K., Montone, K.T., Zhang, G., Wang, L.-P., Diehl, J.A., Rustgi, A.K., Lee, J.T., Weinstein, G.S., and Herlyn, M. (2010). Evidence for mesenchymal-like sub-populations within squamous cell carcinomas possessing chemoresistance and phenotypic plasticity. Oncogene *29*, 4170–4182.

Bean, J., Brennan, C., Shih, J.-Y., Riely, G., Viale, A., Wang, L., Chitale, D., Motoi, N., Szoke, J., Broderick, S., et al. (2007). MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib. Proceedings of the National Academy of Sciences *104*, 20932–20937.

Ben Bolker, Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., and Venables, B. (2012). gplots: Various R programming tools for plotting data.

Benjamini, Y., and Y, H. (1995). Controlling the false discovery rate- a practical and powerful approach to multiple testing. J. R. Statist. Soc. Series B (Methodological) *57*, 289–300.

Binda, E., Visioli, A., Giani, F., Lamorte, G., Copetti, M., Pitter, K.L., Huse, J.T., Cajola, L., Zanetti, N., DiMeco, F., et al. (2012). The EphA2 receptor drives self-renewal and tumorigenicity in stem-like tumor-propagating cells from human glioblastomas. Cancer Cell *22*, 765–780.

Bodenmiller, B., Zunder, E.R., Finck, R., Chen, T.J., Savig, E.S., Bruggner, R.V., Simonds, E.F., Bendall, S.C., Sachs, K., Krutzik, P.O., et al. (2012). Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. Nat Biotechnol *30*, 857–866.

Brambilla, E., Travis, W.D., Colby, T.V., Corrin, B., and Shimosato, Y. (2001). The new World Health Organization classification of lung tumours. European Respiratory Journal *18*, 1059–1068.

Brock, A., Chang, H., and Huang, S. (2009). Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours. Nat Rev Genet *10*, 336–342.

Buchner, M., Fuchs, S., Prinz, G., Pfeifer, D., Bartholome, K., Burger, M., Chevalier, N., Vallat, L., Timmer, J., Gribben, J.G., et al. (2009). Spleen Tyrosine Kinase Is Overexpressed and Represents a Potential Therapeutic Target in Chronic Lymphocytic Leukemia. Cancer Research *69*, 5424–5432.

Burrell, R.A., McGranahan, N., Bartek, J., and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. Nature *501*, 338–345.

Buttigieg, J., Pan, J., Yeger, H., and Cutz, E. (2012). NOX2 (gp91phox) is a predominant O2 sensor in a human airway chemoreceptor cell line: biochemical, molecular, and electrophysiological evidence. AJP: Lung Cellular and Molecular Physiology *303*, L598–L607.

Byers, L. A. & Rudin, C. M. Small cell lung cancer: Where do we go from here? *Cancer* 121, 664–672 (2015).

Byers, L.A., Diao, L., Wang, J., Saintigny, P., Girard, L., Peyton, M., Shen, L., Fan, Y., Giri, U., Tumula, P.K., et al. (2013). An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. Clin. Cancer Res. *19*, 279–290.

Byers, L.A., Wang, J., Nilsson, M.B., Fujimoto, J., Saintigny, P., Yordy, J., Giri, U., Peyton, M., Fan, Y.H., Diao, L., et al. (2012). Proteomic profiling identifies dysregulated pathways in small cell lung cancer and novel therapeutic targets including PARP1. Cancer Discovery *2*, 798–811.

Cai, H., Smith, D.A., Memarzadeh, S., Lowell, C.A., Cooper, J.A., and Witte, O.N. (2011). Differential transformation capacity of Src family kinases during the initiation of prostate cancer. Proceedings of the National Academy of Sciences *108*, 6579–6584.

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.

Cardnell, R.J., Feng, Y., Diao, L., Fan, Y.H., Masrorpour, F., Wang, J., Shen, Y., Mills, G.B., Minna, J.D., Heymach, J.V., et al. (2013). Proteomic Markers of DNA Repair and PI3K Pathway Activation Predict Response to the PARP Inhibitor BMN 673 in Small Cell Lung Cancer. Clin. Cancer Res. *19*, 6322–6328.

Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature *463*, 318–325.

Chanda, S., Ang, C.E., Davila, J., Pak, C., Mall, M., Lee, Q.Y., Ahlenius, H., Jung, S.W., Südhof, T.C., and Wernig, M. (2014). Generation of Induced Neuronal Cells by the Single Reprogramming Factor ASCL1. Stem Cell Reports 1–15.

Charrad, M., Ghazzali, N., Boiteau, and Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software *61*, 1–36.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Chen, L., Monti, S., Juszczynski, P., Daley, J., Chen, W., Witzig, T.E., Habermann, T.M., Kutok, J.L., and Shipp, M.A. (2008). SYK-dependent tonic B-cell receptor signaling is a rational treatment target in diffuse large B-cell lymphoma. Blood *111*, 2230–2237.

Chen, Z., Fillmore, C.M., Hammerman, P.S., Kim, C.F., and Wong, K.-K. (2014). Non-small-cell lung cancers:a heterogeneous set of diseases. Nature Reviews Cancer *14*, 535–546.

Cheng, A.M., Rowley, B., Pao, W., Hayday, A., Bolen, J.B., and Pawson, T. (1995). Syk tyrosine kinase required for mouse viability and B-cell development. Nature *378*, 303–306.

Cheng, S., Coffey, G., Zhang, X.H., Shaknovich, R., Song, Z., Lu, P., Pandey, A., Melnick, A.M., Sinha, U., and Wang, Y.L. (2011). SYK inhibition and response prediction in diffuse large B-cell lymphoma. Blood *118*, 6342–6352.

Chin, L., Tam, A., Pomerantz, J., Wong, M., Holash, J., Bardeesy, N., Shen, Q., O'Hagan, R., Pantginis, J., Zhou, H., et al. (1999). Essential role for oncogenic Ras in tumour maintenance. Nature *400*, 468–472.

Choi, M., Shi, J., Jung, S.H., Chen, X., and Cho, K.-H. (2012). Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage. Science Signaling *5*, 251, ra83, 1-13.

Chong, C.R., and Jänne, P.A. (2013). The quest to overcome resistance to EGFR-targeted therapies in cancer. Nature Medicine *19*, 1389–1400.

Christensen, C.L., Kwiatkowski, N., Abraham, B.J., Carretero, J., Al-Shahrour, F., Zhang, T., Chipumuro, E., Herter-Sprie, G.S., Akbay, E.A., Altabef, A., et al. (2014). Targeting Transcriptional Addictions in Small Cell Lung Cancer with a Covalent CDK7 Inhibitor. Ccell *26*, 909–922.

Chu, L., Eric, S., and Kondo, T. (2001). GeneSpring: Tools for Analyzing Microarray Expression Data. Genome Informatics *12*, 227–229.

Clinical Lung Cancer Genome Project (CLCGP), Network Genomic Medicine NGM (2013). A genomics-based classification of human lung tumors. Science Translational Medicine *5*, 209ra153–209ra153.

Coopman, P.J., Do, M.T., Barth, M., Bowden, E.T., Hayes, A.J., Basyuk, E., Blancato, J.K., Vezza, P.R., McLeskey, S.W., Mangeat, P.H., et al. (2000). The Syk tyrosine kinase suppresses malignant growth of human breast cancer cells. Nature *406*, 742–747.

Coopman, P.J., and Mueller, S.C. (2006). The Syk tyrosine kinase: A new negative regulator in tumor growth and progression. Cancer Letters *241*, 159–173.

Creixell, P., Schoof, E.M., Erler, J.T., and Linding, R. (2012). Navigating cancer network attractors for tumor-specific therapy. Nat Biotechnol *30*, 842–848.

Cutz, E., and Jackson, A. (1999). Neuroepithelial bodies as airway oxygen sensors. Respiration Physiology *115*, 201–214.

Dakir, E.L.H., Feigenbaum, L., and Linnoila, R.I. (2008). Constitutive expression of human keratin 14 gene in mouse lung induces premalignant lesions and squamous differentiation. Carcinogenesis *29*, 2377–2384.

Daniel, V.C., Marchionni, L., Hierman, J.S., Rhodes, J.T., Devereux, W.L., Rudin, C.M., Yung, R., Parmigiani, G., Dorsch, M., Peacock, C.D., et al. (2009). A Primary Xenograft Model of Small-Cell Lung Cancer Reveals Irreversible Changes in Gene Expression Imposed by Culture In vitro. Cancer Research *69*, 3364–3373.

Davidich, M.I., and Bornholdt, S. (2008). Boolean network model predicts cell cycle sequence of fission yeast. PLoS ONE *3 (2)*, e1672.

de Bruin, E.C., McGranahan, N., Mitter, R., Salm, M., Wedge, D.C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A.J., et al. (2014). Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science *346*, 251–256.

Demedts, I.K., Vermaelen, K.Y., and van Meerbeeck, J.P. (2009). Treatment of extensive-stage small cell lung carcinoma: current status and future prospects. European Respiratory Journal *35*, 202–215.

Desai, T.J., Brownfield, D.G., and Krasnow, M.A. (2014). Alveolar progenitor and stem cells in lung development, renewal and cancer. Nature *507*, 190–194.

Domnik, N.J., and Cutz, E. (2011). Pulmonary neuroepithelial bodies as airway sensors: putative role in the generation of dyspnea. Current Opinion in Pharmacology *11*, 211–217.

Dong, C., Yuan, T., Wu, Y., Wang, Y., Fan, T.W.M., Miriyala, S., Lin, Y., Yao, J., Shi, J., Kang, T., et al. (2013). Loss of FBP1 by Snail-Mediated Repression Provides Metabolic Advantages in Basal-like Breast Cancer. Cancer Cell *23*, 316–331.

Dutta, B., Pusztai, L., Qi, Y., eacute, F.A., Lazar, V., Bianchini, G., Ueno, N., Agarwal, R., Wang, B., Shiang, C.Y., et al. (2012). A network-based, integrative study to identify core biological pathways that drive breast cancer clinical subtypes. British Journal of Cancer 1–10.

Engelman, J.A., Zejnullahu, K., Mitsudomi, T., Song, Y., Hyland, C., Park, J.O., Lindeman, N., Gale, C.M., Zhao, X., Christensen, J., et al. (2007). MET Amplification Leads to Gefitinib Resistance in Lung Cancer by Activating ERBB3 Signaling. Science *316*, 1039–1043.

Ettinger, D.S. (2006). Changing Face of Small-Cell Lung Cancer: Real and Artifact. Journal of Clinical Oncology *24*, 4526–4527.

Felsher, D.W., and Bishop, J.M. (1999). Reversible tumorigenesis by MYC in hematopoietic lineages. Molecular Cell *4*, 199–207.

Filosso, P.L., Ruffini, E., Asioli, S., Giobbe, R., Macri, L., Bruna, M.C., Sandri, A., and Oliaro, A. (2011). Adenosquamous lung carcinomas: A histologic subtype with poor prognosis. Lung Cancer *74*, 25–29.

Fischer, B., Marinov, M., and Arcaro, A. (2007). Targeting receptor tyrosine kinase signalling in small cell lung cancer (SCLC): What have we learned so far? Cancer Treatment Reviews *33*, 391–406.

Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M.J., et al. (2009). Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCAMutation Carriers. N Engl J Med *361*, 123–134.

Friedberg, J.W., Sharman, J., Sweetenham, J., Johnston, P.B., Vose, J.M., LaCasce, A., Schaefer-Cutillo, J., De Vos, S., Sinha, R., Leonard, J.P., et al. (2010). Inhibition of Syk with fostamatinib disodium has significant clinical activity in non-Hodgkin lymphoma and chronic lymphocytic leukemia. Blood *115*, 2578–2585.

Gallego, G. A., Villaamil, V. M. & Puerta, A. C (2012). Neuroendocrine Tumours of the Lung. Lung Diseases – Selected State of the Art Reviews, 203-230.

Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*, 570–575.

Gaughan, E.M., and Costa, D.B. (2011). Genotype-driven therapies for non-small cell lung cancer: focus on EGFR, KRAS and ALK gene abnormalities. Ther Adv Med Oncol *3*, 113–125.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307–315.

Guix, M. et al. Acquired resistance to EGFR tyrosine kinase inhibitors in cancer cells is mediated by loss of IGF-binding proteins. J. Clin. Invest. 118, 2609–2619 (2008).

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic State Transitions Give Rise to Phenotypic Equilibrium in Populations of Cancer Cells. Cell *146*, 633–644.

Gustafsson, B.I., Kidd, M., Chan, A., Malfertheiner, M.V., and Modlin, I.M. (2008). Bronchopulmonary neuroendocrine tumors. Cancer *113*, 5–21.

Hahn, C.K., Berchuck, J.E., Ross, K.N., Kakoza, R.M., Clauser, K., Schinzel, A.C., Ross, L., Galinsky, I., Davis, T.N., Silver, S.J., et al. (2009). Proteomic and Genetic Approaches Identify Syk as an AML Target. Cancer Cell *16*, 281–294.

Hammerman, P.S., Lawrence, M.S., Voet, D., Jing, R., Cibulskis, K., Sivachenko, A., Stojanov, P., McKenna, A., Lander, E.S., Gabriel, S., et al. (2012). Comprehensive genomic characterization of squamous cell lung cancers. Nature *489*, 519–525.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674.

Hann, C.L., and Rudin, C.M. (2007). Fast, hungry and unstable: finding the Achilles' heel of small-cell lung cancer. Trends in Molecular Medicine *13*, 150–157.

Hassanein, M., Callison, J.C., Callaway-Lane, C., Aldrich, M.C., Grogan, E.L., and Massion, P.P. (2012). The State of Molecular Biomarkers for the Early Detection of Lung Cancer. Cancer Prevention Research *5*, 992–1006.

Hassanein, M. et al. SLC1A5 mediates glutamine transport required for lung cancer cell growth and survival. Clin. Cancer Res. 19, 560–570 (2013).

He, X., and Zhang, J. (2006). Why Do Hubs Tend to Be Essential in Protein Networks? PLoS Genet *2*, e88.

Hiltermann, T.J.N., Pore, M.M., van den Berg, A., Timens, W., Boezen, H.M., Liesker, J.J.W., Schouwink, J.H., Wijnands, W.J.A., Kerner, G.S.M.A., Kruyt, F.A.E., et al. (2012). Circulating tumor cells in small-cell lung cancer: a predictive and prognostic factor. Annals of Oncology *23*, 2937–2942.

Hiroshima, K., Iyoda, A., Shida, T., Shibuya, K., Iizasa, T., Kishi, H., Tanizawa, T., Fujisawa, T., and Nakatani, Y. (2006). Distinction of pulmonary large cell neuroendocrine carcinoma from small cell lung carcinoma: a morphological, immunohistochemical, and molecular analysis. Mod Pathol *19*, 1358–1368.

Hirsch, F.R. (2003). Epidermal Growth Factor Receptor in Non-Small-Cell Lung Carcinomas: Correlation Between Gene Copy Number and Protein Expression and Impact on Prognosis. Journal of Clinical Oncology *21*, 3798–3807.

Hirsch, F.R., MJ, M., S, A., O, C., JD, E., Gazdar, A.F., B, M., M, N., Shimosato, Y., RH, S., et al. (1988). Histopathologic Classification of Small Cell Lung Cancer. Cancer *62*, 973–977.

Hirsch, F., Varella-Garcia, M., Cappuzzo, F., McCoy, J., Bemis, L., Xavier, A., Dziadziuszko, R., Gumerlock, P., Chansky, K., West, H., et al. (2006). Combination of EGFR gene copy number and protein expression predicts outcome for advanced non-small-cell lung cancer patients treated with gefitinib. Annals of Oncology *18*, 752–760.

Ho, H.J., Lin, T.I., Chang, H.H., Haase, S.B., Huang, S., and Pyne, S. (2012). Parametric modeling of cellular state transitions as measured with flow cytometry. BMC Bioinformatics *13*, S5.

Hodgkinson, C.L., Morrow, C.J., Li, Y., Metcalf, R.L., Rothwell, D.G., Trapani, F., Polanski, R., Burt, D.J., Simpson, K.L., Morris, K., et al. (2014). Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. Nature Medicine *20*, 897–903.

Hodkinson, P.S., Elliott, T., Wong, W.S., Rintoul, R.C., Mackinnon, A.C., Haslett, C., and Sethi, T. (2006). ECM overrides DNA damage-induced cell cycle arrest and apoptosis in small-cell lung cancer cells through β1 integrin-dependent activation of PI3-kinase. Cell Death and Differentiation *13*, 1776–1788.

Hohla, F., and Schally, A.V. (2010). Targeting gastrin releasing peptide receptors: New options for the therapy and diagnosis of cancer. Cell cycle *9*, 1738–1741.

Horváth, S., Zhang, B., Carlson, M., Lu, K.V., Zhu, S., Felciano, R.M., Laurance, M.F., Zhao, W., Qi, S., Chen, Z., et al. (2006). Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proc. Natl. Acad. Sci. U.S.A. *103*, 17402–17407.

Huang, J. et al. EphA2 promotes epithelial-mesenchymal transition through the Wnt/β-catenin pathway in gastric cancer cells. Oncogene 33, 2737–2747 (2014).

Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. Development *136*, 3853–3862.

Huang, S. (2011). On the intrinsic inevitability of cancer: from foetal to fatal attraction. Seminars in Cancer Biology *21*, 183–199.

Huang, S. (2012). The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology? Bioessays *34*, 149–157.

Huang, S. (2013). Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. Cancer Metastasis Rev *32*, 423–448.

Hubaux, R., KL, T., Coe, B.P., MacAulay, C., Lam, S., and Lam, W.L. (2013). EZH2 Promotes E2F-Driven SCLC Tumorigenesis through Modulation of Apoptosis and Cell-Cycle Regulation. Journal of Thoracic Oncology *8*, 1–5.

Humphrey, L.L., Deffebach, M., Pappas, M., Baumann, C., Artis, K., Mitchell, J.P., Zakher, B., Fu, R., and Slatore, C.G. (2013). Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. Ann. Intern. Med. *159*, 411–420.

Inukai, M., Toyooka, S., Ito, S., Asano, H., Ichihara, S., Soh, J., Suehisa, H., Ouchida, M., Aoe, K., Aoe, M., et al. (2006). Presence of epidermal growth factor receptor

gene T790M mutation as a minor clone in non-small cell lung cancer. Cancer Research *66*, 7854–7858.

Irish, J.M., Myklebust, J.H., Alizadeh, A.A., Houot, R., Sharman, J.P., Czerwinski, D.K., Nolan, G.P., and Levy, R. (2010). B-cell signaling networks reveal a negative prognostic human lymphoma cell subset that emerges during tumor progression. Proceedings of the National Academy of Sciences *107*, 12747–12754.

Ischenko, I., Liu, J., Petrenko, O., and Hayman, M.J. (2014). Transforming growth factor-beta signaling network regulates plasticity and lineage commitment of lung cancer cells. *21*, 1218–1228.

Ito, T., Udaka, N., Yazawa, T., Okudela, K., Hayashi, H., Sudo, T., Guillemot, F., Kageyama, R., and Kitamura, H. (2000). Basic helix-loop-helix transcription factors regulate the neuroendocrine differentiation of fetal mouse pulmonary epithelium. Development *127*, 3913–3921.

Jahchan, N. S. et al. A Drug Repositioning Approach Identifies Tricyclic Antidepressants as Inhibitors of Small Cell Lung Cancer and Other Neuroendocrine Tumors (2013). Cancer Discovery 3, 1364–1377.

Jeong, H., Mason, S.P., Barabasi, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature *411*, 41–42.

Ji, H., Ramsey, M.R., Hayes, D.N., Fan, C., McNamara, K., Kozlowski, P., Torrice, C., Wu, M.C., Shimamura, T., Perera, S.A., et al. (2007). LKB1 modulates lung cancer differentiation and metastasis. Nature *448*, 807–810.

Jiang, S.-X., Kameya, T., Asamura, H., Umezawa, A., Sato, Y., Shinada, J., Kawakubo, Y., Igarashi, T., Nagai, K., and Okayasu, I. (2003). hASH1 expression is closely correlated with endocrine phenotype and differentiation extent in pulmonary neuroendocrine tumors. Mod Pathol *17*, 222–229.

Jiang, T., Collins, B.J., Jin, N., Watkins, D.N., Brock, M.V., Matsui, W., Nelkin, B.D., and Ball, D.W. (2009). Achaete-Scute Complex Homologue 1 Regulates Tumor-Initiating Capacity in Human Small Cell Lung Cancer. Cancer Research *69*, 845–854.

Johnson, B.E., Mazor, T., Hong, C., Barnes, M., Aihara, K., McLean, C.Y., Fouse, S.D., Yamamoto, S., Ueda, H., Tatsuno, K., et al. (2014). Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. Science *343*, 189–193.

Jones, M.H., Virtanen, C., Honjoh, D., Miyoshi, T., Satoh, Y., Okumura, S., Nakagawa, K., Nomura, H., and Ishikawa, Y. (2004). Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. The Lancet *363*, 775–781.

Ke, A.-W. et al. CD151 amplifies signaling by integrin α6β1 to PI3K and induces the epithelial-mesenchymal transition in HCC cells. Gastroenterology 140, 1629–41.e15 (2011).

Khatri, P., Sirota, M., and Butte, A.J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol *8*, e1002375.

Kim, C.F.B., Jackson, E.L., Woolfenden, A.E., Lawrence, S., Babar, I., Vogel, S., Crowley, D., Bronson, R.T., and Jacks, T. (2005). Identification of Bronchioalveolar Stem Cells in Normal Lung and Lung Cancer. Cell *121*, 823–835.

Kotton, D.N., and Morrisey, E.E. (2014). Lung regeneration: mechanisms, applications and emerging stem cell populations. Nature Medicine *20*, 822–832.

Krumsiek, J., Marr, C., Schroeder, T., and Theis, F.J. (2011). Hierarchical Differentiation of Myeloid Progenitors Is Encoded in the Transcription Factor Network. PLoS ONE *6*, e22649.

Kruskal, W.H., and Wallis, W.A. (1952). Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association *47*, 583–621.

Kuehl, R.O. (2000). Design of Experiments (Duxbury Press).

Kuno, Y. (2001). Constitutive kinase activation of the TEL-Syk fusion gene in myelodysplastic syndrome with t(9;12)(q22;p12). Blood *97*, 1050–1055.

Kutner, M.H. (2005). Applied Linear Statistical Models (McGraw-Hill Education).

Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., and Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics *26*, 2438–2444.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research *22*, 1813–1831.

Lang, A.H., Li, H., Collins, J.J., and Mehta, P. (2014). Epigenetic Landscapes Explain Partially Reprogrammed Cells and Identify Key Reprogramming Genes. PLoS Comput Biol *10*, e1003734.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lapointe, J., Li, C., Higgins, J.P., van de Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U., et al. (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc. Natl. Acad. Sci. U.S.A. *101*, 811–816.

Lee, M.J., Ye, A.S., Gardino, A.K., Heijink, A.M., Sorger, P.K., MacBeath, G., and Yaffe, M.B. (2012). Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. Cell *149*, 780–794.

Lefebvre, C., Rieckhof, G., and Califano, A. (2012). Reverse-engineering human regulatory networks. WIREs Syst Biol Med *4*, 311–325.

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., and Pietenpol, J.A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. *121*, 2750–2767.

Levitt, M.L., Gazdar, A.F., Oie, H.K., Schuller, H., and Thacher, S.M. (1990). Cross-linked envelope-related markers for squamous differentiation in human lung cancer cell lines. Cancer Research *50*, 120–128.

Li, M., Gray, W., Zhang, H., Chung, C.H., Billheimer, D., Yarbrough, W.G., Liebler, D.C., Shyr, Y., and Slebos, R.J.C. (2010). Comparative Shotgun Proteomics Using Spectral Count Data and Quasi-Likelihood Modeling. J. Proteome Res. *9*, 4295–4305.

Li, Y., and Linnoila, R.I. (2012). Multidirectional Differentiation of Achaete–Scute Homologue–1–Defined Progenitors in Lung Development and Injury Repair. American Journal of Respiratory Cell and Molecular Biology *47*, 768–775.

Lockwood, W.W., Chari, R., Coe, B.P., Girard, L., MacAulay, C., Lam, S., Gazdar, A.F., Minna, J.D., and Lam, W.L. (2008). DNA amplification is a ubiquitous mechanism of oncogene activation in lung and other cancers. Oncogene *27*, 4615–4624.

Van Lommel, A. Pulmonary neuroendocrine cells (PNEC) and neuroepithelial bodies (NEB): chemoreceptors and regulators of lung development. Paediatr Respir Rev 2, 171–176 (2001).

Loo, L.-H., Wu, L. F. & Altschuler, S. J. Image-based multivariate profiling of drug responses from single cells. Nature Methods 4, 445–453 (2007).

Lovly, C.M., and Pao, W. (2012). Escaping ALK inhibition: mechanisms of and strategies to overcome resistance. Science Translational Medicine *4*, 120ps2–120ps2.

Lu, K.V., Zhu, S., Cvrljevic, A., Huang, T.T., Sarkaria, S., Ahkavan, D., Dang, J., Dinca, E.B., Plaisier, S.B., Oderberg, I., et al. (2009). Fyn and Src Are Effectors of Oncogenic Epidermal Growth Factor Receptor Signaling in Glioblastoma Patients. Cancer Research *69*, 6889–6898.

Ma, Z.-Q., Dasari, S., Chambers, M.C., Litton, M.D., Sobecki, S.M., Zimmerman, L.J., Halvey, P.J., Schilling, B., Drake, P.M., Gibson, B.W., et al. (2009). IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. J. Proteome Res. *8*, 3872–3881.

Maher, B. (2012). ENCODE: The human encyclopaedia. Nature *489*, 46–48.

Malkoski, S.P., Cleaver, T.G., Thompson, J.J., Sutton, W.P., Haeger, S.M., Rodriguez, K.J., Lu, S.-L., Merrick, D., and Wang, X.-J. (2013). Role of PTEN in basal cell derived lung carcinogenesis. Mol. Carcinog. *53*, 841–846.

Margolin, A.A., Wang, K., Lim, W.K., Kustagi, M., Nemenman, I., and Califano, A. (2006). Reverse engineering cellular networks. Nat Protoc *1*, 662–671.

Marro, S., Pang, Z.P., Yang, N., Tsai, M.-C., Qu, K., Chang, H.Y., Südhof, T.C., and Wernig, M. (2011). Short Article. Stem Cell *9*, 374–382.

Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? Nature Reviews Cancer *12*, 323–334.

Ma Z-Q, Tabb D.L, Burden J, Chambers M.C, Cox M.B, Cantrell M.J, Ham A-J. L, Litton M.D, Oreto M.R, Schultz W.C, Sobecki S.M, Tsui T.Y, Wernke G.R, Liebler D.C (2011). Supporting tool suite for production proteomics. 27, 3214–3215.

Massion, P.P. (2014). Biomarkers to the rescue in a lung nodule epidemic. J. Clin. Oncol. *32*, 725–726.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.-Y., Chou, A., Ienasescu, H., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Research *42*, D142–D147.

Matys, V., Fricke, E., Geffers, R., Gössling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., et al. (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Research *31*, 374–378.

McFadden, D.G., Papagiannakopoulos, T., Taylor-Weiner, A., Stewart, C., Carter, S.L., Cibulskis, K., Bhutkar, A., McKenna, A., Dooley, A., Vernon, A., et al. (2014). Genetic and Clonal Dissectionof Murine Small Cell Lung Carcinoma Progression by Genome Sequencing. Cell *156*, 1298–1311.

Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G.D. (2010). Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. PLoS ONE *5*, e13984.

Meyerson, M., Franklin, W.A., and Kelley, M.J. (2004). Molecular classification and molecular genetics of human lung cancers. Seminars in Oncology *31*, 4–19.

Micalizzi, D.S., Farabaugh, S.M., and Ford, H.L. (2010). Epithelial-Mesenchymal Transition in Cancer: Parallels Between Normal Development and Tumor Progression. J Mammary Gland Biol Neoplasia *15*, 117–134.

Miller, J.A., Cai, C., Langfelder, P., Geschwind, D.H., Kurian, S.M., Salomon, D.R., and Horvath, S. (2011). Strategies for aggregating gene expression data: The collapseRows R function. BMC Bioinformatics *12*, 322.

Miwa, S., Yano, S., Kimura, H., Yamamoto, M., Toneri, M., Matsumoto, Y., Uehara, F., Hiroshima, Y., Murakami, T., Hayashi, K., et al. (2014). Cell-cycle fate-monitoring distinguishes individual chemosensitive and chemoresistant cancer cells in drug-treated heterogeneous populations demonstrated by real-time fucci imaging. Cell cycle, 14:4, 621-629.

Mori, M., Nakagami, H., Koibuchi, N., Miura, K., Takami, Y., Koriyama, H., Hayashi, H., Sabe, H., Mochizuki, N., Morishita, R., et al. (2009). Zyxin Mediates Actin Fiber Reorganization in Epithelial–Mesenchymal Transition and Contributes to Endocardial Morphogenesis. Molecular Biology of the Cell *20*, 3115–3124.

Mócsai, A., Ruland, J., and Tybulewicz, V.L.J. (2010). The SYK tyrosine kinase: a crucial player in diverse biological functions. Nature Publishing Group *10*, 387–402.

Network, T.C.G.A.R. (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543–550.

Nicholson, S.A., Beasley, M.B., Brambilla, E., Hasleton, P.S., Colby, T.V., Sheppard, M.N., Falk, R., and Travis, W.D. (2002). Small Cell Lung Carcinoma (SCLC) - A Clinicopathologic Study of 100 Cases With Surgical Specimens. The American Journal of Surgical Pathology *26*, 1184–1197.

Ocak, S., Yamashita, H., Udyavar, A.R., Miller, A.N., Gonzalez, A.L., Zou, Y., Jiang, A., Yi, Y., Shyr, Y., Estrada, L., et al. (2010). DNA copy number aberrations in small-cell lung cancer reveal activation of the focal adhesion pathway. Oncogene *29*, 6331–6342.

Ocak, S., Chen, H., Callison, C., Gonzalez, A.L., and Massion, P.P. (2011). Expression of focal adhesion kinase in small-cell lung carcinoma. Cancer *118*, 1293–1301.

Osada, H. (2005). ASH1 Gene Is a Specific Therapeutic Target for Lung Cancers with Neuroendocrine Features. Cancer Research *65*, 10680–10685.

Pao, W., and Hutchinson, K.E. (2012). Chipping away at the lung cancer genome. Nature Medicine *18*, 349–351.

Pao, W., Iafrate, A.J., and Su, Z. (2010). Genetically informed lung cancer medicine. J. Pathol. *223*, 231–241.

Pao, W., Miller, V.A., and Kris, M.G. (2004). “Targeting” the epidermal growth factor receptor tyrosine kinase with gefitinib (Iressa®) in non-small cell lung cancer (NSCLC). Seminars in Cancer Biology *14*, 33–40.

Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism. Cell *155*, 1008–1021.

Park, K.-S., Liang, M.-C., Raiser, D.M., Zamponi, R., Roach, R.R., Curtis, S.J., Walton, Z., Schaffer, B.E., Roake, C.M., Zmoos, A.-F., et al. (2011a). Characterization of the cell of origin for small cell lung cancer. Cell cycle *10*, 2806–2815.

Park, K.-S., Martelotto, L.G., Peifer, M., Sos, M.L., Karnezis, A.N., Mahjoub, M.R., Bernard, K., Conklin, J.F., Szczepny, A., Yuan, J., et al. (2011b). A crucial

requirement for Hedgehog signaling in small cell lung cancer. Nature Medicine *17*, 150–1508.

Park, S.Y., Gonen, M., Kim, H.J., Michor, F., and Polyak, K. (2010). Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. J. Clin. Invest. *120*, 636–644.

Parsons, S.J., and Parsons, J.T. (2004). Src family kinases, key regulators of signal transduction. Oncogene *23*, 7906–7909.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

Peifer, M., Fernández-Cuesta, L., Sos, M.L., George, J., Seidel, D., Kasper, L.H., Plenker, D., Leenders, F., Sun, R., Zander, T., et al. (2012). Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat Genet *44*, 1104–1110.

Perez-Moreno, P., Brambilla, E., Thomas, R., and Soria, J.C. (2012). Squamous Cell Carcinoma of the Lung: Molecular Subtypes and Therapeutic Opportunities. Clinical Cancer Research *18*, 2443–2451.

Pietanza, M.C., and Ladanyi, M. (2012). Bringing the genomic landscape of small-cell lung cancer into focus. Nat Genet *44*, 1074–1075.

Pietanza, M.C., Kadota, K., Huberman, K., Sima, C.S., Fiore, J.J., Sumner, D.K., Travis, W.D., Heguy, A., Ginsberg, M.S., Holodny, A.I., et al. (2012). Phase II trial of temozolomide in patients with relapsed sensitive or refractory small cell lung cancer, with assessment of methylguanine-DNA methyltransferase as a potential biomarker. Clin. Cancer Res. *18*, 1138–1145.

Polyak, K. (2011). Heterogeneity in breast cancer. J. Clin. Invest. *121*, 3786–3788.

Polyak, K., and Weinberg, R.A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nature Reviews Cancer *9*, 265–273.

Ponta, H., Sherman, L., and Herrlich, P.A. (2003). CD44: From adhesion molecules to signalling regulators. Nature Reviews Molecular Cell Biology *4*, 33–45.

Popper, H.H. (2011). Large cell carcinoma of the lung – a vanishing entity? Memo *4*, 4–9.

Posadas, E.M., Al-Ahmadie, H., Robinson, V.L., Jagadeeswaran, R., Otto, K., Kasza, K.E., Tretiakov, M., Siddiqui, J., Pienta, K.J., Stadler, W.M., et al. (2009). FYNis overexpressed in human prostate cancer. BJU International *103*, 171–177.

Prinos, P., Garneau, D., Lucier, J.-F., Gendron, D., Couture, S., Boivin, M., Brosseau, J.-P., Lapointe, E., Thibault, P., Durand, M., et al. (2011). Alternative splicing of SYK regulates mitosis and cell survival. Nature Publishing Group *18*, 673–679.

Pruszak, J., Ludwig, W., Blak, A., Alavian, K., and Isacson, O. (2009). CD15, CD24, and CD29 Define a Surface Biomarker Code for Neural Lineage Differentiation of Stem Cells. Stem Cells *27*, 2928–2940.

Quintana, E., Shackleton, M., Foster, H.R., Fullen, D.R., Sabel, M.S., Johnson, T.M., and Morrison, S.J. (2010). Phenotypic Heterogeneity among Tumorigenic Melanoma Cells from Patients that Is Reversible and Not Hierarchically Organized. Cancer Cell *18*, 510–523.

Rekhtman, N. (2010). Neuroendocrine Tumors of the Lung: An Update. Arch Pathol Lab Med *134*, 1628–1638.

Renshaw, A.A., Voytek, T.M., Haja, J., Wilbur, D.C., and Cytology Committee, College of American Pathologists (2005). Distinguishing small cell carcinoma from non-small cell carcinoma of the lung: correlating cytologic features and performance in the College of American Pathologists Non-Gynecologic Cytology Program. Arch Pathol Lab Med *129*, 619–623.

Reth, M. (2002). Hydrogen peroxide as second messenger in lymphocyte activation. Nature Immunology *3*, 1129–1134.

Riccaboni, M., Bianchi, I., and Petrillo, P. (2010). Spleen tyrosine kinases: biology, therapeutic targets and drugs. Drug Discovery Today *15*, 517–530.

Riely, G.J., Marks, J., and Pao, W. (2009). KRAS Mutations in Non-Small Cell Lung Cancer. Proceedings of the American Thoracic Society *6*, 201–205.

Rigby, S., Huang, Y., Streubel, B., Chott, A., Du, M.Q., Turner, S.D., and Bacon, C.M. (2009). The Lymphoma-associated Fusion Tyrosine Kinase ITK-SYK Requires

Pleckstrin Homology Domain-mediated Membrane Localization for Activation and Cellular Transformation. Journal of Biological Chemistry *284*, 26871–26881.

Rock, J.R., Gao, X., Xue, Y., Randell, S.H., Kong, Y.-Y., and Hogan, B.L.M. (2011). Notch-Dependent Differentiation of Adult Airway Basal Stem Cells. Stem Cell *8*, 639–648.

Rohrbeck, A., Neukirchen, J., Rosskopf, M., Pardillos, G.G., Geddert, H., Schwalen, A., Gabbert, H.E., Haeseler, von, A., Pitschke, G., Schott, M., et al. (2008). Gene expression profiling for molecular distinction and characterization of laser captured primary lung cancers. J Transl Med *6*, 69.

Rossi, A., Maione, P., Palazzolo, G., Sacco, P.C., Ferrara, M.L., Falanga, M., and Gridelli, C. (2008). New Targeted Therapies and Small-Cell Lung Cancer. Clinical Lung Cancer *9*, 271–279.

Rosti, G., Bevilacqua, G., Bidoli, P., Portalone, L., Santo, A., and Genestreti, G. (2006). Small cell lung cancer. Annals of Oncology *17*, ii5–ii10.

Rudin, C.M., Durinck, S., Stawiski, E.W., Poirier, J.T., Modrusan, Z., Shames, D.S., Bergbower, E.A., Guan, Y., Shin, J., Guillory, J., et al. (2012). Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nat Genet *44*, 1111–1116.

Russell, P.A., Wainer, Z., Wright, G.M., Daniels, M., Conron, M., and Williams, R.A. (2011). Does lung adenocarcinoma subtype predict patient survival?: A clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification. J Thorac Oncol *6*, 1496–1504.

Saadatpour, A., Albert, I., and Albert, R. (2010). Journal of Theoretical Biology. Journal of Theoretical Biology *266*, 641–656.

Sada, K., Minami, Y., and Yamamura, H. (1997). Relocation of Syk Protein-Tyrosine Kinase to the Actin Filament Network and Subsequent Association with Fak. Eur. J. Biochem *248*, 827–833.

Sada, K., Takano, T., Yanagi, S., and Yamamura, H. (2001). Structure and function of Syk protein-tyrosine kinase. Journal of Biochemistry *130*, 177–186.

Sadanandam, A., Lyssiotis, C.A., Homicsko, K., Collisson, E.A., Gibb, W.J., Wullschleger, S., Ostos, L.C.G., Lannon, W.A., Grotzinger, C., Del Rio, M., et al. (2013). A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nature Medicine 1–8.

Saito, Y.D., Jensen, A.R., Salgia, R., and Posadas, E.M. (2010). Fyn - A Novel Molecular Target in Cancer. Cancer *116*, 1629–1637.

Sato, T., Kaneda, A., Tsuji, S., Isagawa, T., Yamamoto, S., Fujita, T., Yamanaka, R., Tanaka, Y., Nukiwa, T., Marquez, V.E., et al. (2013). PRC2 overexpression and PRC2-target gene repression relating to poorer prognosis in small cell lung cancer. Sci. Rep. 3, 1911.

Scagliotti, G.V., Novello, S., Rapetti, S., and Papotti, M. (2013). Current state-of-the-art therapy for advanced squamous cell lung cancer. Am Soc Clin Oncol Educ Book 354–358.

Schultheis, A.M., Bos, M., Schmitz, K., Wilsberg, L., Binot, E., Wolf, J.U.R., ttner, R.B.U., and Schildhaus, H.-U. (2013). Fibroblast growth factor receptor 1 (FGFR1)amplification is a potential therapeutic target in small-cell lung cancer. Modern Pathology, 1–8.

Schwitalla, S., Fingerle, A.A., Cammareri, P., Nebelsiek, T., Göktuna, S.I., Ziegler, P.K., Canli, O., Heijmans, J., Huels, D.J., Moreaux, G., et al. (2013). Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. Cell *152*, 25–38.

Sequist, L. V. et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors (2011). Science Translational Medicine 3, 75ra26–75ra26.

Shan, L., Aster, J.C., Sklar, J., and Sunday, M.E. (2007). Notch-1 regulates pulmonary neuroendocrine cell differentiation in cell lines and in transgenic mice. Am. J. Physiol. Lung Cell Mol. Physiol. *292*, L500–L509.

Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., et al. (2010). A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. Cell *141*, 69–80.

Shi, Z., Derow, C.K., and Zhang, B. (2010). Co-expression module analysis reveals biological processes, genomic gain, and regulatory mechanisms associated with breast cancer progression. BMC Systems Biology *4*, 74.

Siegel, R., Naishadham, D., and Jemal, A. (2012). Cancer statistics, 2012. CA: a Cancer Journal for Clinicians *62*, 10–29.

Simonds, E.F., Bendall, S.C., Gibbs, K.D., Bruggner, R.V., Linderman, M.D., Sachs, K., Nolan, G.P., Plevritis, S.K., and Qiu, P. (2011). nbt.1991. Nat Biotechnol *29*, 886–891.

Singh, D.K., Ku, C.-J., Wichaidit, C., Steininger, R.J., Wu, L.F., and Altschuler, S.J. (2010). Patterns of basal signaling heterogeneity can distinguish cellular populations with different drug sensitivities. Molecular Systems Biology *6*, 369.

Slebos, R.J.C., Brock, J.W.C., Winters, N.F., Stuart, S.R., Martinez, M.A., Li, M., Chambers, M.C., Zimmerman, L.J., Ham, A.J., Tabb, D.L., et al. (2008). Evaluation of strong cation exchange versus isoelectric focusing of peptides for multidimensional liquid chromatography-tandem mass spectrometry. J. Proteome Res. *7*, 5286–5294.

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. Proceedings of the National Academy of Sciences *109*, 17531–17536.

Sos, M.L., Dietlein, F., Peifer, M., Schöttle, J., Balke-Want, H., Müller, C., Koker, M., Richters, A., Heynck, S., Malchers, F., et al. (2012). A framework for identification of actionable cancer genome dependencies in small cell lung cancer. Proceedings of the National Academy of Sciences *109*, 17034–17039.

Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature *459*, 428–432.

Sperry, R.B., Bishop, N.H., Bramwell, J.J., Brodeur, M.N., Carter, M.J., Fowler, B.T., Lewis, Z.B., Maxfield, S.D., Staley, D.M., Vellinga, R.M., et al. (2010). Zyxin controls migration in epithelial-mesenchymal transition by mediating actin-membrane linkages at cell-cell junctions. J. Cell. Physiol. *222*, 612–624.

Sprung, R.W., Brock, J.W.C., Tanksley, J.P., Li, M., Washington, M.K., Slebos, R.J.C., and Liebler, D.C. (2009). Equivalence of protein inventories obtained from formalin-fixed paraffin-embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis. Mol. Cell Proteomics *8*, 1988–1998.

Staaf, J., Isaksson, S., Karlsson, A., Jönsson, M., Johansson, L., Jönsson, P., Botling, J., Micke, P., Baldetorp, B., and Planck, M. (2012). Landscape of somatic allelic imbalances and copy number alterations in human lung carcinoma. Int. J. Cancer *132*, 2020–2031.

Statnikov, A., Tsamardinos, I., Dosbayev, Y., and Aliferis, C.F. (2005). GEMS: a system for automated cancer diagnosis and biomarker discovery from microarray gene expression data. Int J Med Inform *74*, 491–503.

Sturm, N., Rossi, G., Lantuejoul, S., Papotti, M., Frachon, S., Claraz, C., Brichon, P.-Y., Brambilla, C., and Brambilla, E. (2002). Expression of thyroid transcription factor-1 in the spectrum of neuroendocrine cell lung proliferations with special interest in carcinoids. Human Pathology *33*, 175–182.

Subramanian, J., and Simon, R. (2010). Gene Expression-Based Prognostic Signatures in Lung Cancer: Ready for Clinical Use? JNCI Journal of the National Cancer Institute *102*, 464–474.

Sutherland, K.D., Proost, N., Brouns, I., Adriaensen, D., Song, J.-Y., and Berns, A. (2011). Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. Cancer Cell *19*, 754–764.

Sutherland, K.D., Song, J.-Y., Kwon, M.C., Proost, N., Zevenhoven, J., and Berns, A. (2014). Multiple cells-of-origin of mutant K-Ras-induced mouse lung adenocarcinoma. Proceedings of the National Academy of Sciences *111*, 4952–4957.

Suvà, M.L., Rheinbay, E., Gillespie, S.M., Patel, A.P., Wakimoto, H., Rabkin, S.D., Riggi, N., Chi, A.S., Cahill, D.P., Nahed, B.V., et al. (2014). Reconstructing and Reprogramming the Tumor-Propagating Potential of Glioblastoma Stem-like Cells. Cell *157*, 580–594.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics *22*, 1540–1542.

Swanton, C. (2012). Intratumor Heterogeneity: Evolution through Space and Time. Cancer Research *72*, 4875–4882.

Swarts, D.R.A., Ramaekers, F.C.S., and Speel, E.-J.M. (2012). Biochimica et Biophysica Acta. BBA - Reviews on Cancer *1826*, 255–271.

Szerlip, N.J., Pedraza, A., Chakravarty, D., Azim, M., McGuire, J., Fang, Y., Ozawa, T., Holland, E.C., Huse, J.T., Jhanwar, S., et al. (2012). Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. Proc. Natl. Acad. Sci. U.S.A. *109*, 3041–3046.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U.S.A. *98*, 10869–10874.

Tabb, D.L., Fernando, C.G., and Chambers, M.C. (2007). MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. J. Proteome Res. *6*, 654–661.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676.

Takano, T., Sada, K., and Yamamura, H. (2002). Role of Protein-Tyrosine Kinase Syk in Oxidative Stress Signaling in B Cells. Antioxidants & Redox Signaling *4*, 533–541.

Takeuchi, T., Tomida, S., Yatabe, Y., Kosaka, T., Osada, H., Yanagisawa, K., Mitsudomi, T., and Takahashi, T. (2006). Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. Journal of Clinical Oncology *24*, 1679–1688.

Tam, W.L., and Weinberg, R.A. (2013). The epigenetics of epithelial-mesenchymal plasticity in cancer. Nature Medicine 19, 1438–1449.

Team, R.C. (2012). R: A Language and Environment for Statistical Computing (Vienna, Austria).

Therneau, T.M., and Grambsch, P.M. (2000). Modeling survival data: extending the Cox model.

Thomas, L.A., Akins, M.R., and Biederer, T. (2008). Expression and adhesion profiles of SynCAM molecules indicate distinct neuronal functions. Journal of Comparative Neurology *510*, 47–67.

Tobar, N., Villar, V., and Santibanez, J.F. (2010). ROS-NFkappaB mediates TGF-beta1-induced expression of urokinase-type plasminogen activator, matrix metalloproteinase-9 and cell invasion. Mol Cell Biochem *340*, 195–202.

Tran, P.T., Fan, A.C., Bendapudi, P.K., Koh, S., Komatsubara, K., Chen, J., Horng, G., Bellovin, D.I., Giuriato, S., Wang, C.S., et al. (2008). Combined Inactivation of MYC and K-Ras Oncogenes Reverses Tumorigenesis in Lung Adenocarcinomas and Lymphomas. PLoS ONE *3*, e2125.

Travis, W.D. (2010). Advances in neuroendocrine lung tumors. Annals of Oncology *21*, vii65–vii71.

Travis, W.D. (2012). Update on small cell carcinoma and its differentiation from squamous cell carcinoma and other non-small cell carcinomas. Mod Pathol *25 Suppl 1*, S18–S30.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 1–16.

Tsurutani, J. (2005). Inhibition of the Phosphatidylinositol 3-Kinase/Akt/Mammalian Target of Rapamycin Pathway but not the MEK/ERK Pathway Attenuates Laminin-Mediated Small Cell Lung Cancer Cellular Survival and Resistance to Imatinib Mesylate or Chemotherapy. Cancer Research *65*, 8423–8432.

Tyson, D.R., Garbett, S.P., Frick, P.L., and Quaranta, V. (2012). Fractional proliferation: a method to deconvolve cell population dynamics from single-cell data. Nature Methods *9*, 923–928.

Udyavar, A.R., Hoeksema, M.D., Clark, J.E., Zou, Y., Tang, Z., Li, Z., Li, M., Chen, H., Statnikov, A., Shyr, Y., et al. (2013). Co-expression network analysis identifies SpleenTyrosine Kinase (SYK) as a candidate oncogenic driver in a subset of small-cell lung cancer. BMC Systems Biology 7, S1.

Velcheti, V., and Govindan, R. (2007). Hedgehog Signaling Pathway and Lung Cancer. Journal of Thoracic Oncology *2*, 7–10.

Verhaak, R.G.W., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al. (2010). Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell *17*, 98–110.

Villani, M., Barbieri, A., and Serra, R. (2011). A dynamical model of genetic networks for cell differentiation. PLoS ONE, 6 (3), e17703.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2012). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380–384.

Walker, G.E., Antoniono, R.J., Ross, H.J., Paisley, T.E., and Oh, Y. (2005). Neuroendocrine-like differentiation of non-small cell lung carcinoma cells: regulation by cAMP and the interaction of mac25/IGFBP-rP1 and 25.1. Oncogene *25*, 1943–1954.

Walls, M., Baxi, S.M., Mehta, P.P., Liu, K.K.C., Zhu, J., Estrella, H., Li, C., Zientek, M., Zong, Q., Smeal, T., et al. (2014). Targeting Small Cell Lung Cancer Harboring PIK3CA Mutation with a Selective Oral PI3K Inhibitor PF-4989216. Clinical Cancer Research *20*, 631–643.

Wang, A., Huang, K., Shen, Y., Xue, Z., Cai, C., Horvath, S., and Fan, G. (2011a). Functional Modules Distinguish Human Induced Pluripotent Stem Cells from Embryonic Stem Cells. Stem Cells and Development *20*, 1937–1950.

Wang, R.-S., Saadatpour, A., and Albert, R. (2012). Boolean modeling in systems biology: an overview of methodology and applications. Physical Biology *9*, 055001–055015.

Wang, Z.-X., Lu, B.-B., Yang, J.-S., Wang, K.-M., and De, W. (2011b). Adenovirus-mediated siRNA targeting c-Met inhibits proliferation and invasion of small-cell lung cancer (SCLC) cells. J. Surg. Res. *171*, 127–135.

Warburton, D., Wuenschell, C., Flores-Delgado, G., and Anderson, K. (1998). Commitment and differentiation of lung cell lineages. Biochem. Cell Biol *76*, 971–995.

Wardwell, N.R., and Massion, P.P. (2005). Novel Strategies for the Early Detection and Prevention of Lung Cancer. Seminars in Oncology *32*, 259–268.

Weinstein, I.B. (2002). Addiction to Oncogenes--the Achilles Heal of Cancer. Science *297*, 63–64.

Weiss, J., Sos, M.L., Seidel, D., Peifer, M., Zander, T., Heuckmann, J.M., Ullrich, R.T., Menon, R., Maier, S., Soltermann, A., et al. (2010). Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. Science Translational Medicine *2*, 62ra93–62ra93.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318–324.

Wilkerson, M.D., Yin, X., Hoadley, K.A., Liu, Y., Hayward, M.C., Cabanski, C.R., Muldrew, K., Miller, C.R., Randell, S.H., Socinski, M.A., et al. (2010). Lung Squamous Cell Carcinoma mRNA Expression Subtypes Are Reproducible, Clinically Important, and Correspond to Normal Cell Types. Clin. Cancer Res. *16*, 4864–4875.

Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

Wilkerson, M.D., Yin, X., Walter, V., Zhao, N., Cabanski, C.R., Hayward, M.C., Miller, C.R., Socinski, M.A., Parsons, A.M., Thorne, L.B., et al. (2012). Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. PLoS ONE *7*, e36530.

Winden, K.D., Oldham, M.C., Mirnics, K., Ebert, P.J., Swan, C.H., Levitt, P., Rubenstein, J.L., Horvath, S., and Geschwind, D.H. (2009). The organization of the transcriptional network in specific neuronal classes. Molecular Systems Biology *5*, 291.

WM, B., L, H., A, L., F, F., A, S., P, K., TB, B., T, P., W, H., EG, H., et al. (2006). Predictive and Prognostic Factors in Small Cell Lung Carcinoma (SCLC) – Analysis from Routine Clinical Practice. Anticancer Res. *26*, 4825–4832.

Woyach, J.A., Johnson, A.J., and Byrd, J.C. (2012). The B-cell receptor signaling pathway as a therapeutic target in CLL. Blood *120*, 1175–1184.

Wu, Z.-Q., Li, X.-Y., Hu, C.Y., Ford, M., Kleer, C.G., and Weiss, S.J. (2012). Canonical Wnt signaling regulates Slug activity and links epithelial–mesenchymal transition with epigenetic Breast Cancer 1, Early Onset (BRCA1) repression. Proc. Natl. Acad. Sci. U.S.A. *109*, 16654–16659.

Xu, C., Fillmore, C.M., Koyama, S., Wu, H., Zhao, Y., Chen, Z., Herter-Sprie, G.S., Akbay, E.A., Tchaicha, J.H., Altabef, A., et al. (2014). Loss of Lkb1 and Pten Leads to Lung Squamous Cell Carcinoma with Elevated PD-L1 Expression. Cancer cell *25*, 590–604.

Yamamoto, H., Shigematsu, H., Nomura, M., Lockwood, W.W., Sato, M., Okumura, N., Soh, J., Suzuki, M., Wistuba, I.I., Fong, K.M., et al. (2008). PIK3CA Mutations and Copy Number Gains in Human Lung Cancers. Cancer Research *68*, 6913–6921.

Yeoh, E.-J., Ross, M.E., Shurtleff, S.A., Williams, W.K., Patel, D., Mahfouz, R., Behm, F.G., Raimondi, S.C., Relling, M.V., Patel, A., et al. (2002). Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell *1*, 133–143.

Yuan, Y., Chen, H., Ma, G., Cao, X., and Liu, Z. (2012). Reelin Is Involved in Transforming Growth Factor-β1-Induced Cell Migration in Esophageal Carcinoma Cells. PLoS ONE *7*, e31802.

Yuan, Y., Mendez, R., Sahin, A., and Le Dai, J. (2001). Hypermethylation Leads to Silencing of the *SYK* Gene in Human Breast Cancer. Cancer Research *61*, 5558–5561.

Zarogoulidis, K., Zarogoulidis, P., K, D., E, B., Machairiotis, N., Tsakiridis, K., and Katsikogiannis, N. (2013). Treatment of non-small cell lung cancer (NSCLC). Journal of Thoracic Disease, 5, supp 4,S389-S396.

Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Research 33, W741–W748.

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J., Shaddox, K.F., Kim, S., et al. (2014). Proteogenomic characterization of human colon and rectal cancer. Nature 513, 382–387.

Zhang, J., Fujimoto, J., Zhang, J., Wedge, D.C., Song, X., Zhang, J., Seth, S., Chow, C.-W., Cao, Y., Gumbs, C., et al. (2014b). Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science 346, 256–259.

Zhang, J., Benavente, C.A., McEvoy, J., Flores-Otero, J., Ding, L., Chen, X., Ulyanov, A., Wu, G., Wilson, M., Wang, J., et al. (2012). A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature 481, 329–334.

Zhang, Y., and He, J. (2013). The development of targeted therapy in small cell lung cancer. J Thorac Dis 5, 538–548.

Zhao, Y., Butler, E.B., and Tan, M. (2013). Targeting cellular metabolism to improve cancer therapeutics. Cell Death Dis 4, e532–10.

Zhu, J., Wang, J., Shi, Z., Franklin, J.L., Deane, N.G., Coffey, R.J., Beauchamp, R.D., and Zhang, B. (2013). Deciphering Genomic Alterations in Colorectal Cancer through Transcriptional Subtype-Based Network Analysis. PLoS ONE 8, e79282.